A Connectionist Simulation: Towards a Model of Child Language Development

Syed Sibte Raza Abidi

Ph. D. Thesis

Submitted to the University of Surrey in partial fulfilment of requirements for the degree of Doctor of Philosophy

Department of Mathematical and Computing Sciences
University of Surrey
Guildford
Surrey

August 1994
Our research focuses on the connectionist simulation of child language development within the age group 9-24 months. We present a hybrid connectionist model - ACCLAIM (A Connectionist Child Language development and Imitation Model), comprising 'supervised' and 'unsupervised' learning connectionist networks, that take into account the diverse nature of inputs to and outputs from a child learning his or her first language. The model is used to simulate the child's development of concepts, acquisition of words, ostensive naming of concepts, understanding of conceptual and semantic relations and the learning of word-order. The simulation produces child-like one-word and two-word sentences. The simulation of aspects of child language development are 'language informed' in that the data used in the simulation was taken from extant child language corpora. Theoretical underpinnings of our simulation were based on Jean Piaget's notions of cognitive development. The efficacy of hybrid connectionist models is demonstrated through the operationalisation of real child language data. The simulations indicate that connectionist networks can simulate developmental behaviour, and both connectionist and developmental psychology communities can benefit from such a contribution.
Acknowledgements

I would like to acknowledge Dr. Khurshid Ahmad for his intellectual guidance. His suggestions were helpful in formulating a theoretically sound framework for pursuing my research. His attention and endurance need to be commended during the compilation of this thesis, and certainly his ideas has helped me in writing this thesis.

I would like to thank the Ministry of Science and Technology, Government of Pakistan for financing me during the earlier years of my studies. In addition, I would like to express my gratitude to the department for providing me financial assistance in the third year of my research, which enabled me to continue my work.

Special thanks goes to Paul Holmes-Higgin for sharing information, having useful discussions and assisting in technical matters, and also to Andrea Davies for proof-reading the drafts of my thesis. Caroline McInnes deserves a note of thanks for coping with all tasks and arrangements relevant to my research that were thrust on her, usually at a short notice. I would also like to thank other members of the AI Group, particularly John Wright, for sharing their utilities and providing me with a friendly working atmosphere. I would also like to thank the computing staff for their co-operation.

I would like to express my thanks to Prof. Ron Shail and Prof. Steve Schuman for their support and hospitality in providing me with a stimulating environment for research.

My wife Samina deserves special thanks for her care, attention and understanding, and especially for her company when I was working for long hours in the night at the university. My parents can not be ever thanked enough for all what they have done for me and guiding me to this stage where I am concluding my thesis.
Chapter 4

The Connectionist Implementation of a Child Language Development Model - ACCLAIM

4.1. Introduction 83
4.2. What Do We Learn From Previous Connectionist Simulations 84
4.3. A Psycholinguistic Model of Child Language Development: Basis of a Connectionist Model 88
4.4. Specification of Connectionist Modules Constituting ACCLAIM 91
   4.4.1. Concept Memory and Word Lexicon 93
   4.4.2. Naming Connection Network 95
   4.4.3. Conceptual Relation Network 97
   4.4.4. Semantic Relation Network 103
   4.4.5. Word-order Testing Network 106
4.5. The Architecture of ACCLAIM: A Synthesis of Connectionist Modules 108
4.6. Is a Hybrid of Connectionist Networks Really Needed to Simulate Child Language Development? 113
4.7. Conclusion 117

Chapter 5

A Connectionist Simulation of Child Language Development: One-word and Two-word Language

5.1. Introduction 119
5.2. Experiment 1: Development of a Concept Memory 123
   5.2.1. Experiment 1a: Simulation of the Development of Concept Memory 124
   5.2.2. Experiment 1b: Concept Generalisation 133
5.3. Experiment 2: Development of a Word Lexicon 136
5.4. Experiment 3: Learning Conceptual Relations and Associating Them With One-word Utterances 141
   5.3.1. A Novel Modification of a Backpropagation Network to Realise Bi-directional Associations 149
5.5. Experiment 4: Ostensive Naming-Development of 'Naming Connections' 153
   5.5.1. Experiment 4b: Retrieval of a Concept-Word Pair using Naming Connections 159
5.6. Experiment 5: Learning Semantic Relations 163
5.7. Experiment 6: Learning Word-order 169
5.8. Experiment 7: Production of Two-word Sentences 175
5.9. Conclusions 182

Chapter 6

Conclusions and Future Work 186
6.1 Conclusions 186
6.2. Future Work 190

Appendix A

Bibliography
Chapter 1

Introduction

Connectionism is a research discipline that aims to understand the nature of human intelligence by simulating aspects of human behaviour through a collection of idealised neurons. Connectionist networks or neural networks involve co-operative computations among a large number of highly interconnected neurons or more appropriately processing units. The processing units are provided a variety of stimuli and by communicating with each other they respond in a manner that mimics aspects of human behaviour, particularly aspects of human learning. There is a substantial volume of literature in connectionism that deals with learning through experience and learning through evolution. This thesis intends to explore the role of connectionism in simulating child language development.

Connectionism draws much of its inspiration from neurosciences in that the neuron is taken as the basic processing unit. Each such processing unit is characterised by an activation level, analogous to the state of polarisation of a neuron, an output value representing the firing rate of the neuron, and a set of input and output connections, representing a neuron's axons and dendrites from and to other units, respectively. These characteristics are expressed in a mathematical formalism such that a unit's activation level and output value are expressed as (real) numbers, and its connections with other units have an associated weight (synaptic strength) which determine the effect of the incoming input on the activation level of the unit. The use of a collection of idealised neurons is at the heart of studies in 'microcognition'- a term coined in the seventies. This term emphasises the role of simple individual processors or processing units, rather than the generic approach of symbol manipulation as is generally practised in AI.

Given the complexity of the human brain, both in terms of structure and function, it appears rather ambitious to computationally emulate the human brain in its entirety. Connectionism, nevertheless, uses some of the observations about the human nervous system to develop computer models that attempt to simulate a relationship between architecture and emerging behaviour. The connectionist spirit, therefore,
is to develop simple mathematical models of the brain. These models coupled with psychological theories, can be employed to simulate such aspects of human intelligence that specifically involve an interaction with the environment and result in the emergence or development of some kind of intelligent behaviour.

Connectionist networks mimic the neural structure of the brain *albeit* rather simplistically in that a connectionist network comprises a large number of computationally simple processing units which are highly interconnected through plastic connections. The plasticity in the architecture of a connectionist network is introduced with the help of varying connection weights that can change over time and with experience. The configuration of the processing units dynamically adapts to the environment as a consequence of learning. Put simply, learning in connectionist networks can be envisaged as the problem of finding a set of connection weights which allow the connectionist network to store experiential knowledge and to exploit it to simulate the desired behaviour. One can then argue that connectionist networks have a natural propensity for storing experiential knowledge which is acquired and retained through *training* or *learning* as opposed to explicit programming.

Interests in this new style of computation, that is, 'brain-style' computation has been characterised as the study of *Connectionism* or *Connectionist Networks* by some (Feldman and Ballard, 1982), and *Neural Networks* by others (Hopfield, 1982; Lippmann, 1987). Terms like *Parallel Distributed Processing (PDP) systems* (Rumelhart and McClelland, 1986), *Neural Computing* (Aleksander, 1989) and *Artificial Neural Systems* (Simpson, 1990) have been used together with *Neurocomputing* (Hecht-Nielsen, 1990), and *Subsymbolic Processing* (Smolensky, 1988).

Neurally-inspired theorising has an interesting past. Network models which were predecessors of contemporary connectionist networks were developed during the early 1940s. The foundation was laid by McCulloch and Pitts (1943) who proposed a simple model of 'neuron-like' computational units that could perform logical computations. An illustration of how such networks store information, essentially as associations between two units, was proposed by David Hebb in 1949. Later, von Neuman in 1956 extended McCulloch and Pitts network to make it more reliable by introducing some statistical notions.
In the sixties, Rosenblatt extended the scope of McCulloch and Pitts network by introducing the notion of layers of units in the network thus suggesting a new processing scheme. Rosenblatt (1962) referred to such networks as 'perceptrons'. In the sixties and seventies, research in this field declined to a large extent much due to the unavailability of the computing resources needed to exploit the true potential of connectionist networks and the simultaneous emergence of symbolic systems. Connectionism was revived again in the late eighties, particularly through the efforts of David Rumelhart and John McClelland. Other prominent contributors include Geoffrey Hinton, John Feldman, David Ballard, Paul Smolensky, and John Hopfield.

Connectionist literature does contain many data structures and algorithms that have a 'natural propensity to learn' (Alexsander, 1989). But what is learnt in many connectionist networks cannot be regarded as developmental learning, this is despite the fact that developmental learning can be simulated in connectionist architectures.

Our objective is to simulate child language development using carefully evaluated and well-interpreted connectionist networks. We believe that connectionism, developmental psychology and psycholinguistics will benefit from this simulation. It is gratifying and encouraging to see that the philosopher and linguist John Searle has noted the value of the computational implementation of a philosophical theory: 'Both philosophy and computer science benefit from this combination. On the one hand you can't design a good natural language processing program unless you have a good philosophy of natural language to start with; on the other hand, the attempt to implement a theory of natural language computationally provides a rigorous form of testing the theory. The beauty of working with computers is that the computer will not allow you to fudge, avoid issues or theorise ambiguously' (Searle, 1991: xiii).

We believe that our work will have a modestly similar impact on the developmental psychology, psycholinguistics and the connectionist communities. As remarked earlier, the connectionist community is keenly interested in simulating human learning. A very good exemplar of human learning is child language development. A development that depends on learning through experience and learning through evolution. Child language development has been studied extensively, but despite some excellent
work in this field, there is a tendency to avoid issues and to theorise ambiguously. And, it has to be said, that there are no objective means for testing the effectiveness of a theory in child language or indeed analysing existing corpora of carefully collected data in the field.

Certainly computer-based simulation, has little or no role in helping the child language community. There are many reasons for this, but the most important one appears to be the fact that, apart from connectionist architectures, much of the data structures and algorithms used in computer simulations have little learning potential.

1.1. Learning in Connectionist Networks
The ability to learn or adapt is considered to be an indication of intelligence. For example, Gould and Marley note that 'learning is often thought of as the alternative to instinct, which is the information passed genetically from one generation to the next. Most of us think the ability to learn is the hallmark of intelligence. The difference between learning and instinct is said to distinguish human beings from lower animals such as insects' (1987). In general, learning is defined as 'a change in behaviour in response to training, experience or environment' (Harston, 1990: 72). Connectionists like Jacobs and Schumann have argued that 'learning involves alteration of the microanatomical and molecular neural structure to the point where information can be retained and retrieved so as to be able to effect behaviour' (1992: 287).

In psychological parlance, learning is believed to take place through a variety of mechanisms, such as by example, observation, analogy, association, relational, rote, and by discovery (Reber, 1984: 395-8).

The simulation of human learning mechanisms using traditional AI learning models is well nigh impossible. Consequently, it is suggested that 'the traditional symbolic AI view suffers from a very unhuman-like brittleness.' (Gasser, 1990: 179). By contrast, connectionism acknowledges a relationship between form and function, i.e., a correlation between the architecture and the emerging behaviour: connectionist networks use a neurally-inspired architecture and are physiologically and psychologically more plausible than traditional AI models. Rumelhart and McClelland (1986) claim that connectionist network have the 'obvious physiological flavour' for simulating human cognition.
Connectionist networks learn in two modes - supervised and unsupervised. The supervised learning mode requires the presence of a supervisor to check the correctness of learning, whereas in the unsupervised mode, learning progresses regardless of any instant validation. In both the learning modes information to be learnt is presented in an iterative manner and learning is carried out according to a learning algorithm. Various learning algorithms exist in the connectionist literature to teach networks to associate particular responses with particular stimuli by way of modifying the weights on the connections (for typical examples see Grossberg, 1978; Hinton and Sejnowski, 1986; Rumelhart, Hinton and Williams, 1986; Kohonen, 1988). By means of such learning algorithms, a network initialised with random weights on its connections, can be made sensitive to the most frequently occurring aspects of various input patterns. The learning algorithm ensures that the connectionist network extracts out and learns the subtle correlations between the co-occurring aspects of the input.

The ability of connectionist networks to adaptively learn offers the possibility for simulating cognitive and motor processes as diverse as the development and production of both first and second language, concept development due to both verbal and perceptual stimuli, object recognition, motor-movement development, problem-solving and memory. From a psychological point of view, learning involved in the above processes takes place by example, observation and discovery, leading to the learning of new and the updating of previous knowledge.

1.2. Language Learning and Connectionism
Language learning provides an interesting framework to build sophisticated information systems of the future. Language is generally learnt in a 'naturalistic' setting; the setting is very noisy and the input to and output from a child appear to be indeterminate, that is, the input does not always obey a predetermined sequence and order. Language learning is evolutionary, dynamic and is delineated by various 'milestones' - one word language, two word language and the emergence of early syntax. There is a premium on the child correcting his or her own errors. Furthermore, language processing, which includes learning, comprehension and production, appears to be distributed over various areas of the child's brain, and for that matter over various areas of an adult's brain. Psycho- and socio-linguists have argued that there is a strong correlation between language learning and other tasks such as motor co-
ordination, concept development, categorisation, perception, biological growth and social influences.

Child language development appears to involve an interaction of multiple tasks and the satisfaction of multiple constraints. In this regard, connectionism offers mechanisms, such as adaptive learning, generalisation, self-organisation, and pattern-recognition that appear to have direct relevance towards a computational simulation of child language development.

Indeed, one of the first things a child needs to do is to learn, acquire and assimilate new information. In early childhood, the child starts to understand his or her environment and to learn the language in the environment for survival. In many ways the ability to learn a language and subsequently to use a symbol system as complex as human language, sets humans apart from other species.

The observations of how children learn language, particularly the biological and psychological factors which might influence a child’s language have been documented in the literature of a number of related disciplines: developmental psychology, psycholinguistics, clinical linguistics, neurolinguistics, sociolinguistics and so on. Four major points appear in the above mentioned literature: (a) the influence of maturity on the child’s language; (b) the child’s ability to generalise and specialise, i.e., to ‘categorise’; (c) the dynamic growth of language; and (d) the child’s interaction with the external world and interaction with other humans. It appears that all the above act as an indicator to the extent to which the child can articulate his or her needs related to the external world. These indicators (a-d) also help to evaluate the child’s capacities to learn and to use language. Furth gives a description of the nature of child language learning, which incorporates psychological-cognitive, sociological-environmental and developmental issues. He argues that children ‘learn language in the same way in which they adapt to customs and regulations within the family or society, such as the daily routine, the differences between what the child himself, his siblings, friends and adults can do, particularly also to responsiveness to signs of interpersonal communications other than speech, including of course expressions of feelings and emotions. In brief, once the child has reached a stage where he can make a symbol-oriented response, he is ready to respond meaningfully to the total situation of which spoken language is part and as a by-product he learns language’ (1969: 115).
Child theorists such as Lois Bloom (1973), Roger Brown (1973), Katherine Nelson (1973), M. Braine (1976), and Alison Gophik (1984) relate the development of a child's language to his or her ability to generalise, to recognise patterns, to form concepts and understand the notion of 'rule-governed' behaviour for comprehending, acquiring and producing language fragments. Aspects of these abilities are termed by researchers as 'cognitive tasks'. The existence of these cognitive tasks, or rather the indirect observation that they exist, and that they account for language development has motivated many researchers, including psychologists, neurolinguists and computer scientists to simulate language development in children on computer systems.

Amongst the few linguists with an active interest in connectionism and in child language development there is an opinion that, 'the intertwining of language development with other psychological factors becomes extremely intricate when a fully adequate description is attempted' (Schnelle, 1971:176). Accordingly, our strategy is to choose a formalism that is developmental in nature and based on semantic plausibility, that is, on the 'meaning' of children's utterances rather than on the structure of their utterances.

Language development can be contemplated as an interaction of various language related tasks, for instance the development of concepts, lexical growth, naming of concepts, learning conceptual relations and semantic relations and understanding an underlying word-order in adult language. The implied discussion of such tasks can be further distinguished by demarcating the environmental considerations from what can be regarded as the innate ability of the brain to learn language. Our argument is that some of these tasks can then be simulated by using supervised learning networks whilst other can be simulated by unsupervised learning networks.

It has to be said that, almost by brute force, one can simulate all the language related tasks using one connectionist learning algorithm. But this would be implausible in that language learning is an active interaction of different learning strategies. We propose a large scale 'hybrid connectionist model' - a connectionist model which synthesises a variety of connectionist networks (regarded as modules) and takes advantage of the features each network has to offer. In a hybrid connectionist model each task is
simulated by a separate connectionist network. Hybrid connectionist models can deal with the diversity of input stimuli, which may be perceptual, verbal, or functional and varying learning mechanisms which are mainly classified as supervised and unsupervised. We believe that only in a hybrid architecture one can address all aspects of language development in an integrated fashion, allowing the incorporation of all types of stimuli, learning mechanisms, knowledge types and varying environmental influences.

1.3. Rule-Based Models of Child Language Development

Language development among children has not only earned attention from psychologists and linguists, but also from computer scientists. Research in computer simulations of child language development can be traced back to the seventies in terms of the works of McMaster et al. (1976) and Block et al. (1975). The success of simulation-based studies have inspired a number of workers including Siskind (1990), Hill (1983), Langley (1982) and Selfridge (1982), to build rule-based simulation models of child language development or acquisition as it is referred to by some researchers. Siskind focuses on the acquisition of core meaning of words, Hill and Langley focus on the acquisition of syntax, and Selfridge on the utility of semantics in language development. The approach adopted by each of these researchers is based on the characterisation of early language, taking into account factors such as children's typical linguistic and non-linguistic input, world knowledge, conceptual development, along with postulated learning rules and strategies reported in the literature of the time. Implementation of the above simulation models, is based on traditional AI formalism, such as frames, conceptual dependency grammar or other propositional schemata.

Here we present a brief description of four prominent child language development models and provide a critique based on (a) what linguistic skills (and knowledge of the language) these models acquired, and (b) how the models were implemented, including formalism used, inputs and outputs from each model, and the learning paradigms used.

Siskind's Model (1990): Siskind's model aims at the acquisition of a lexicon and the core meaning of words. Siskind's model uses AI schemata to implement a modular structure, comprising a parser, linker, and an inference component. Input to the model consists of sequences of scenes that are described both
linguistically as a set of simple sentences, and 'pictorially' in terms of conceptual structure descriptions that represent a set of visual scenes. The representation formalism used is based on Jackendoff-style concept structures.

Language learning in Siskind's model is supervised and incorporates a training session, with a teacher describing the 'non-linguistic' component of the 'linguistic' input in a correct and constrained fashion. In the training session, the system uses semantic constraint satisfaction techniques and a set of heuristics to match the linguistic and non-linguistic inputs, in such a way as to infer lexical information from the inputs. The output produced as a result of learning are concept structures given as a lexica for explaining the non-linguistic input with linguistic input.

We can argue that, although Siskind's model incorporates some cognitive aspects of learning, representation, and knowledge in a modular structure to simulate an associative learning behaviour, yet there were no explicit attempts to establish any psychological plausibility to the overall simulation, nor was there any psychological basis for the choice of the representation formalism and learning mechanisms. Another shortcoming of Siskind's simulation is that it is not based on a specific corpus of psycholinguistic data, and is only motivated by speculations about how a child may acquire word meaning in the early stages of language development.

**Hill's Model (1983):** Hill's model simulates the early part of language development, learning both to understand and to generate language at the level of the two year old child. Hill's model simulates learning through the generation of a flat 'template grammar' which evolves into a recursive hierarchical context-free grammar. The model takes as input adult sentences and as output produces child-like sentences that are either repetitions of adult sentences or responses to the adult input sentence in accordance with the current state of the learnt grammar. Hill's simulation is based on child language data collected from a two-year old child named Claire.

Hill's model (Figure 1) is a schema theoretic model, using weighted semantic nets as a dynamic data structures to encode the child's grammar, conceptual knowledge, physical context and a word lexicon.
Prior to learning an initial word lexicon was given, which also had a direct mapping with a set of concepts comprising the conceptual knowledge. The learning paradigm adjusts the weights of the edges, during a training session that involves a repetitive presentation of the adult sentences. The learning relies on word-order and on encoding relations among the concepts. What is learnt, in each presentation of the sentence, depends on the language experience of the model and what has been learnt so far. Words and concepts are classified into categories, and the classification is based on 'word use'. The progress of the child is simulated by developing the grammar gradually. This is done by attending to the adult input and using rules of salience to focus on examples within the adult data to form word classes and to build a grammar.

![Diagram of Hill's model of child language development]

**Hypothesnar & Generation**
- word classes and grammar
- Accommodate structures through successive reorganisation

**Adult sentence**

**Lexicon**
- Concepts and world knowledge
- Present physical context

**Physical context of utterance**

**Child-like repetition or response**

Figure 1: Hill's model of child language development

**Selfridge's Model - CHILD (1982):** Selfridge's model (Figure 2) is based on the assumption that concepts exists before language, and therefore the simulation mainly focuses on the use of semantics in the acquisition of language. Psycholinguistic data collected from a child - Joshua was used for the simulation, however the age of Joshua is not specific. The model learns to understand commands and also to generate declarative sentences. Selfridge's model, has as its input adult sentences together with simulated visual input. Much like Hill, Selfridge also provides as input the meaning of input sentences and information about the situations and goals of communication. The output of Selfridge's model is the child's response, this response could either be a description of an action or it may be verbal.

Selfridge uses Schank's conceptual dependency frames to implement a representation based on meaning. According to this representation, the knowledge structures are built or learnt by attaching lexical knowledge to the concepts. Learning language in CHILD, involves the learning of two aspects: firstly,
learning how to fill slots and secondly, learning information about where slot fillers are placed in a sentence in relation to other words in the sentence. During the learning process, CHILD takes words from the input sentence, retrieves appropriate conceptual dependency frames, and matches the words against the conceptual dependency requirements for filling slots. The matching is based on word meaning and word categories. The learning mechanism incorporates feedback information in that CHILD approves or disapproves the sentences generated by it.

CHILD starts to learn language with no initial knowledge, however by using mechanisms for focusing attention and a set of learning rules and rules of inference, CHILD learns to understand simple commands and to produce simple declarative sentences.

Langley's Model - AMBER (1982): Langley's model (Figure 3) is an adaptive production system based on error recovery, accounting for the gradual learning of language over time, particularly the order in which morphemes are learnt. The simulation is not based on empirical psycholinguistic data, rather on the assumption that the learning mechanism is specific to language development. AMBER, takes as input adult sentences paired with a representation of the meaning and the topic of the sentence. The output of AMBER is a system generated sentence with a description of the main topic of the generated sentences.
AMBER employs a process grammar, encoded as a set of actions. Meaning is represented as a tree structure employing a small set of relations such as agent, action, object, size and colour. AMBER, both understands and learns to produce language through these meaning relations.

Learning in AMBER starts with one-word utterances and progresses to learn appropriate morphemes, suffixes and prefixes. The gradual nature of learning is achieved by attaching weights to the rules (where the weights indicate the relative strength or weakness of the rule), and each rule must be learnt many times before the rule is strong enough to mask a previously learnt rule. Learning also involves a repetitive presentation of input sentences, the more the level of difficulty of the morpheme use, the longer AMBER would take to learn it. For generating sentences, AMBER gets a proposition, predicts a sentence, then compares its generated sentence with the adult sentence it received. Based on the error, AMBER is adjusted to account for the discrepancy between the two. In this way, AMBER incorporates the notion of supervision, both in learning and sentence production.

Table 1 presents a concise review of the structural, functional and theoretical description of the above mentioned computer models of child language development.
The child language development models discussed above can be divided into two categories, based on how the elements of meaning and syntactic structures are processes and stored: (a) Models that handle meaning and syntax in a unified manner, and (b) models that show a rather clear separation between grammar and the conceptual/semantic knowledge domain. Langley’s and Selfridge’s models assume that the language domain and the conceptual domain are not separated, and that both a linguistic structure and its meaning are processed together. In contrast, Hill’s model processes grammar and the semantic knowledge in a separate manner. An obvious limitation in Siskind’s and Langley’s model is that they are not based on child language data, also in Selfridge’s simulation the age of the subject is not known.

Talking about what is innate and what is learnt by these models, Morikawa argues that ‘no model is found which has its initial setting firmly based on some theoretical assumptions or hypothesis made from children’s data, the most common innate mechanism among many existing models is a kind of cognitive propensity to recognise patterns and relations, to form classes of concepts and words, to reorganise stored information and so on’ (1988: 140). In almost all the above computer models these mechanisms are
incorporated, and it appears that the researchers have taken the view that the child brings to language development 'innate' perceptual and conceptual abilities which enable him or her to learn language.

The above models are mainly symbolic and the simulations mainly consider cognition and its various aspects such as representation, learning, environment knowledge and processing, at the higher psychological level, with no reference to their neurobiological origin, if there is any, and therefore these simulations can be argued to simulate the so-called macro structure of cognition.

Hill, claims that symbolic models, like the ones discussed above have advantages in 'representing schemata, inheritance hierarchies, and sequential control' (1992:770). This claim has not been substantiated either way in that the representation schemata (and the implicit inheritance hierarchies) in symbolic models have a number of restrictions, whereas connectionist networks, its proponents claim, can support representation schemata, inheritance hierarchies, and in a restricted manner also 'sequential control'. Representation schemata and inheritance hierarchies can be fairly easily built using connectionist architectures. There are a number of connectionist networks, based on categorisation algorithms, that have successfully simulated classification, both supervised and unsupervised (Levine, 1991: 195-201). Such classification schemata can be used to build taxonomic hierarchies and thus simulate inheritance.

A General Critique of Rule-Based Models: Igor Alexsander has proposed an interesting critique of rule based systems. He argues that the focus of the AI paradigm is based on the premise that an understanding of what the brain does represents a true understanding only if it can be explicitly expressed as a set of rules that, in turn can be run on a computer which subsequently performs artificially intelligent tasks. In rule-based systems the complexity and the number of rules required to capture even the simplest of behaviours appears to increase explosively at times. Also that, AI systems appear not to benefit from the functional use of experiential knowledge therefore rules for language, speech and scene understanding tasks are difficult to find. Problem solving by AI methodology is inherently serial, whilst

---

1 Based on the lecture notes of Dr. K. Ahmad.
a number of cognitive tasks require parallel processing that take into account multiple constraints. Finally, knowledge acquisition or learning is generally regarded as basically a pre-programmed task (1989: 1-33).

The above critique of AI reveals serious shortcomings in the AI paradigm, particularly the dominance of the notion that intelligent activity amongst humans can be, more or less, simulated by predicating the existence of a macrostructure of cognition, and subsequently the simulation of this macrostructure will result in intelligent behaviour. On the contrary, the connectionist standpoint is that intelligent behaviour can be simulated through computation in massively parallel networks of simple processors that store all their long-term knowledge in connection strengths. This knowledge is acquired through training rather than programming.

1.4. A Connectionist Child Language development and Imitation Model - ACCLAIM
This thesis is concerned with the development of an information processing system, a connectionist model, which simulates the development of children's language within the age group of 9-24 months. Our psycholinguistic framework for studying and simulating child language is originally due to Jean Piaget. Piaget has emphasised cognitive development to take place in stages and is the precursor to language development. Such a framework has been refined and elaborated by prominent child theorists mentioned in Section 1.2.

For language development, if one, like Jean Piaget, assumes that development takes place in stages, then the child's early sentences are an excellent indicator of his or her language at each stage. For this reason, literature in child psychology contains elaborate accounts of real-life children's language, explicitly categorised into various developmental stages. Within the connectionism domain, we propose an approach for integrating various connectionist networks to develop a 'hybrid connectionist model' and then exploiting the combined learning potential of the connectionist model to train it to learn the language of a 9-24 month old child. The training can be accomplished by using suitably transformed language and conceptualisation data described in child language studies. We draw data extensively from
the longitudinal studies of the child language development by Bloom (1973), Brown (1973), Nelson (1973 & 74), and Braine (1976). This data is used to train a variety of connectionist networks to recognise the typical input to a 9-24 month child and to produce responses, i.e., one and two word sentences, that are similar to children's early language. For this reason, we regard our connectionist simulation as 'language informed'.

According to our hypothesis, development of language embodies an implicit assumption on the part of the child that there is an orderly mapping between the sounds uttered by adults and 'objects', 'events' and 'actions' perceived by the child. Development of language is environmentally governed, where the child's learning of language fragments is predicated by their repetitive occurrence. Emergence of language is preceded by 'concepts', which are a result of ongoing 'conceptualisation' and 'categorisation' tasks. For this reason, we regard language development as one aspect of the child's overall 'cognitive' development.

The levels of linguistic descriptions mainly addressed in our simulations are semantics (semantic feature based concept development, conceptual relations and semantic relations), phonology (representing and learning words as phonemes) and some syntax (word order formulae signifying pivot grammar).

On the connectionist front we present ACCLAIM - A Connectionist Child Language development and Imitation Model (Figure 4) to simulate child language development within the age-group 9-24 months.
ACCLAIM has been implemented as a 'hybrid' connectionist architecture integrating a variety of connectionist networks, including Kohonen maps, backpropagation networks, additive Grossberg Networks, networks with Hebbian connections incorporating the spreading activation mechanism. ACCLAIM has been used to simulate the development of concepts amongst children together with the ostensive naming of these concepts: the concept memory and word lexicon have been simulated using Kohonen maps and are linked together through a Hebbian connection based naming connection network. Backpropagation networks have been used to implement a conceptual relation network (for one-word utterances) and a word-order testing network (for two-word sentences). Children's evolving semantic performance has been simulated by a semantic relation network using additive Grossberg network. Thus, aspects of what can be construed to be innate development have been simulated using unsupervised learning regimes, like Kohonen maps and Hebbian connections, and environmentally-determined features of language development have been simulated using supervised learning regimes, like backpropagation networks. ACCLAIM has been trained on 'realistic' child language data and has learnt to recognise and produce one-word and two-word sentences.

Our choice of input for the simulation is (1) an adult sentence and (2) a visual image of the world. The input adult sentence is simplified in the sense that just a two-word collocation containing the meaning of the sentence is taken into account. The two-word collocations are represented in terms of their phonological representations, however we omit any consideration of phonology learning. The visual input is based on the child's visual perception and encodes visual information, such as semantic features, about the entities in the child's immediate environment.

1.5. What Have We Achieved?
In the connectionist literature we notice that developmental psychologists have shown interest in connectionism. Similarly, the connectionist community have indulged in developmental psychology issues. However, the existing evidence suggests that researchers from both communities have not really stepped far enough from their parent domain to conclusively exploit the information in the other domain. For instance, the attempts of psychologists to employ connectionist networks to simulate a psychological
theory, is restricted to just the use of simple connectionist networks. Likewise, connectionists have used simple psychological theories and data to demonstrate that connectionist networks have potential for psychological simulations. We believe that connectionism in fact offers a more enhanced programme of inter-disciplinary co-operation: psychologists can experiment with more elaborate and powerful connectionist configurations, while connectionists can demonstrate the efficacy of connectionist networks for psychological simulations using more complex theories and realistic data.

This thesis contributes by taking a step forward in narrowing the gap between connectionism and developmental psychology. In our research, we proposed how to develop a large scale hybrid connectionist model that has the functionality to implement a well-grounded psychological theory. This was achieved by interpreting the underlying cognitive aspects of representation, learning, environment, processing and knowledge into a connectionist terminology. Within connectionism we operationalised real child language data and utilised the learning potential of connectionist networks to 'learn' the data in a more naturalistic setting such that the learnt connectionist model mimics aspects of the language of a child within the age group 9-24 months.

The contribution of this thesis lies in the fact that our attempt to simulate aspects of child language development within the connectionist paradigm is both unique and innovative in the sense that we have proposed a connectionist simulation approach that is based on psychological theories and child language data. Our simulations have provided an answer to the question that whether connectionist networks can simulate developmental behaviour?, and indeed we have demonstrated that connectionist networks can simulate developmental learning: knowledge is not programmed, rather starting with no a priori knowledge our connectionist networks learn knowledge over time and with experience. We have also addressed the central notion of learning through experience and evolution by incorporating both these scenarios in our hybrid connectionist model - ACCLAIM. Last by not least, we like to mention that ours is the first attempt to operationalise existing child language data and theories using connectionist networks and this initial attempt has been very successful as is evident by the results of our simulations.
In a simulation context, we have simulated a concept memory, using a Kohonen map, which explicates the child's conceptual knowledge and its development. As compared with earlier attempts of concept development our simulation scheme is much improved and psychological well-grounded. The distinguishing features of our simulation are the automatic categorisation of concepts emerging as a side-effect of learning, language is not a prerequisite for concept development (cognition hypothesis), the learning scheme is self-motivated (unsupervised) as opposed to rote learning in earlier simulations, the provision to learn new concepts, and the ability to generalise to learnt concepts when encountered with novel and erroneous input. We argue that our simulation of concept development demonstrates the efficacy of Kohonen maps which are postulated to have both neurobiological and psychological relevance for concept development and their storage.

A simulation of the child's developing lexical knowledge resulted in the origination of the child's word lexicon. The connectionist architecture used, Kohonen maps, simulating an unsupervised learning mechanism discriminated various words in terms of their phonetic constituents and as a side-effect of learning the automatic emergence of similarity neighbourhoods of similar sounding words was realised.

Children's language at the one-word stage have been simulated in terms of the learning of conceptual relations using a backpropagation network. We regard our simulation as systematising Bloom's notion of conceptual relations and their relationship with function words in a supervised learning environment. The learnt conceptual relation network was able to produce child-like one-word utterances that correspond with real child language data from Bloom (1973). In this simulation we also proposed a novel modification to the structure of a typical backpropagation network in order to demonstrate the existence of transitive relationships in backpropagation networks.

We simulated the ostensive naming of concepts by mapping the child's conceptual knowledge to his or her lexical knowledge, the process regarded in child language literature as the naming of concepts. This was achieved by a novel connectionist implementation in which two Kohonen maps (one simulating the concept memory and the other the word lexicon) were interconnected using Hebbian connections. Interestingly, this association between concepts and words was simulated in an unsupervised manner with
no external stimuli directing the learning algorithm to connect a concept with a particular word. The learnt **naming connection network** then provided opportunities to retrieve a concept upon hearing a word or alternatively to express a concept by retrieving the corresponding word.

Transition of child language from the one-word stage to the two-word stage - a step towards the production of child-like two-word sentences was simulated by learning a set of semantic relations between conceptual categories as proposed by Brown (1973). A **semantic relation network** based on the connectionist network - unsupervised learning additive Grossberg network was used to perform the simulation. Although a simpler, yet psychologically less significant simulation could be done by using a backpropagation network that learnt to associate two concept categories. However, to maintain psychological plausibility we performed a much complex simulation which incorporated an interaction between three connectionist networks (the earlier learnt concepts memory, word lexicon and naming connection network) to transform the input stimuli to a representation that was psychologically plausible for learning semantic relations. Our processing scheme was entirely unsupervised - the interacting networks on their own sequenced the flow of information and in doing so knowledge was transformed as it was passed from one connectionist network to another.

A simple version of syntax was simulated by way of learning a set of word-ordering formulae - a simple pivot grammar. A backpropagation network performing pattern association was used for this simulation, resulting in a **word-order testing network**. The two stages of word-order learning mentioned in child language literature, i.e. first learning 'groping patterns' and then the actual word-order was simulated. This simulation involved the transformation of the input stimuli to word-order information and for that purpose we again incorporated an interaction between the concept memory, word lexicon, naming connection network and semantic relation network. The word-order information learnt by the connectionist network can not only be used to test the word-order between two words but also to correct the word-order. The manner in which the input stimuli is transformed by an interaction amongst various connectionist networks is quite unique in the connectionist literature.
Finally, a simulation of the production of child-like two-word sentences was performed to demonstrate the developmental nature of all earlier simulations - starting with no knowledge we first developed the knowledge necessary to simulate the production of one-word utterances and then we progressed to the two-word stage and our connectionist model - ACCLAIM developed to produce two-word sentences. The two-word sentences produced by ACCLAIM correspond with real child language data from Bloom (1973). This simulation can be seen as an excellent example of how knowledge learnt and stored in various connectionist networks can be used in a unified manner - the spirit of a hybrid model.

We argue that our work is different from the earlier simulations of linguistic behaviour in three significant respects: first, we simulate a number of aspects of human language including lexical organisation and lexical access, conceptual memory, semantics, pivot grammar and word-order for studying evolving linguistic behaviour. Second, our focus is on the development or the evolution of linguistic behaviour amongst children. The notions of innate structures notwithstanding, language is learnt over a period of time and involves environmental input. This includes input from the physical environment, caretakers, siblings and others, together with language learnt by the child on his or her own initiative with or without supervision, either through the maturity of the nervous system or through some other natural gift. Third, we believe that the interdependence of language learnt via environmental input and self-motivated language learning, can surely influence the kinds of connectionist network architectures that either simulate 'supervised learning' or 'unsupervised learning'.

Finally, we claim that this thesis has elucidated the efficacy of connectionist simulation programs that exhibit some kind of learning that is particularly relevant to child language development. Both child language development theories and connectionism can benefit from this combination: on the one hand, one cannot design a good connectionist network program unless one starts with an exemplar learning phenomenon that is theoretically well-grounded; in the other hand, the attempt to simulate the observations about or the existing theories of child language development will provide a rigorous test of the theories.
1.6. Structure of the Thesis

This thesis is divided into five other chapters, focusing on the various aspects of a connectionist simulation of the development of language by a child. The progression of the discussion is as such:

Chapter two is theoretical in nature and is concerned with the description of a psycholinguistic framework eliciting child language development for the purposes of a connectionist simulation. In our case Piaget's notions of cognitive development forms the basis for an explanation of language development. In a Piagetian perspective, we would discuss learning mechanisms based on assimilation and accommodation, children's concepts, referential purpose of children early words, notions of conceptual and semantic relations and the existence of some syntax embedded in word-order knowledge. Explanations in this chapter are based on the accounts of prominent child theorists and psychologists.

Chapter three provides a connectionist interpretation to some of the basis aspects of cognition which include representation, learning, knowledge, environment and processing. The idea here is to relate connectionist terminology to psychological accounts of cognition with an emphasis on child language development. We would firstly discuss the implications of the above mentioned basic aspects of cognition in the child's development of language and then determine how our connectionist interpretation facilitates a realistic simulation of language development. For non-connectionists, we provide a brief overview of the candidate connectionist networks that we intend to use for our simulation (a detailed discussion of connectionist networks is given in Appendix A).

Chapter four discusses the design and specification of a connectionist child language development model - ACCLAIM, which is implemented as a hybrid (or modular) connectionist architecture. In this respect we would propose a methodology for the development of hybrid connectionist models and then demonstrate the efficacy of our approach by developing ACCLAIM.

Chapter five gives a detailed account of a connectionist simulation of child language development. The connectionist simulation described in this chapter brings together the discussion in previous chapters. The simulation is based on the psychological framework given in chapter two, employs the connectionist
terminology eliciting language development based on the discussion in chapter three and uses the connectionist model - ACCLAIM developed in chapter four. We would perform nine separate simulations of various aspects of child language development and quantify the results in terms of similar observations in psychological studies of child language development. We would demonstrate the production of child-like one-word utterances and two-word sentences.

Chapter six concludes the thesis. We would discuss our contribution to the research programme in connectionism. We would then answer the question as to what have we achieved in this thesis. Finally, we would conclude the thesis with suggestions for future research in the field of connectionism and cognition, with proposals for a connectionist workbench.
Chapter 2

Psycholinguistic Aspects of Child Language Development

2.1. Introduction
Language development research continues to develop apace and a variety of theories, supporting evidence and counter claims, are being put forward fairly regularly. In this chapter we describe a psycholinguistic framework for studying and simulating child language development, particularly within the 9 - 24 month age group. The framework, originally due to Jean Piaget, emphasises cognitive development as the precursor to language development, and has been refined and elaborated by Lois Bloom (1973), Katherine Nelson (1973 & 1974), Roger Brown (1973) and his student the late Richard Cromer (1974 & 1991).

This thesis is concerned with the 'emergence' of language and therefore we will focus on the first Piagetian stage and its six substages. Our aim is to simulate language development within a connectionist framework, based on available data about the earliest stages of children's development.

2.2. Notes on Piaget's Theory of Cognitive Development
In many ways, Piaget seeks to strike a balance between a child's inborn capacities (innateness) and the effect of the child's immediate environment. Our understanding of Piaget's theories are based on the notes of Flavell (1963), Furth (1969), Ault (1983) and Ingram (1989).
2.2.1. Stages of Cognitive Development

In Jean Piaget's model of cognitive development, a child progresses steadily from an infant to a young adult through a number of stages (shown in Table 2). Each stage is delineated by 'developmental milestones', characterising the development of certain properties of children's functioning: spatial, physical, conceptual, linguistic, social, and so on. These stages range from sensori-motor co-ordination at around the second birthday, to the ability of the child to think about the world in abstract terms at around the twelfth birthday. Language, according to Piaget and his followers, appears at around the first birthday.

<table>
<thead>
<tr>
<th>Stage</th>
<th>Approximate Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensori-motor</td>
<td>Birth - (1 1/2 - 2) years</td>
</tr>
<tr>
<td>Preoperational</td>
<td>(1 1/2 - 2) - (6 - 7) years</td>
</tr>
<tr>
<td>Concrete Operational</td>
<td>(6 - 7) - (11 - 12) years</td>
</tr>
<tr>
<td>Formal Operational</td>
<td>(11 - 12) - through adulthood</td>
</tr>
</tbody>
</table>

Table 2: Piaget's stages of cognitive development

Anisfeld (1984: 20), inspired by Piaget's notions, observes that children in the sensori-motor stage do not think symbolically. In the preoperational stage children appear to think symbolically, but their thinking is not yet analytic. By contrast, children in the concrete operational stage possess the ability to analyse wholes into component parts. Although, at this stage children can solve concrete problems, yet their ability to manipulate abstract ideas are rather limited. Finally, in the formal operational stage children acquire formal operations and the capacity for 'hypothetical thought'.

In Piaget's general account, the sensori-motor stage is further divided into six substages, during which the child gradually becomes able to define practical goals and proceed systematically to achieve them. Here, we will illustrate Piaget's six substages for sensori-motor development in terms of their role in the child's development of language. (see Table 3).

---

1 A stage in Piagetian terms 'accounts for those changes in human functioning which are caused by a genuine qualitative shift, or for the construction of new representational systems in an individual's knowledge system'. (Dromi 1987:12).
<table>
<thead>
<tr>
<th>Substages (Age)</th>
<th>Psycho-biological Development</th>
<th>Level of Language</th>
<th>Object Permanence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Substage 1 c. 0 - 1 months</td>
<td>Activating reflexes.</td>
<td></td>
<td>Objects have no independent existence.</td>
</tr>
<tr>
<td>Substage 2 c. 1 - 4 months</td>
<td>Primary circular reactions. Assimilation becomes separate from accommodation.</td>
<td></td>
<td>No interest in vanishing object; objects exist only as parts of actions.</td>
</tr>
<tr>
<td>Substage 3 c. 4 - 8 months</td>
<td>Secondary circular reactions. Preserves interesting sights; actions are repeated because of their consequences.</td>
<td>Babble</td>
<td>Anticipates where object may fall; reaches for partially hidden object; searches for object just seen; associates object with other's actions.</td>
</tr>
<tr>
<td>Substage 4 c. 8 - 12 months</td>
<td>Co-ordination of secondary circular reactions. Schemes are directed toward goals. Sensori-motor intelligence</td>
<td>Presymbolic uses of words. Words used as 'signs' and usage governed by personal subjective considerations.</td>
<td>Pursues hidden objects presumed to be at previous site.</td>
</tr>
<tr>
<td>Substage 5 c. 12 - 18 months</td>
<td>Tertiary circular reactions. New schemes evolve to achieve goals. Separation of representation from action.</td>
<td>Words used as personal symbols - words used to compare and comment, and usage inspired by an appreciation of social conventional nature of meaning.</td>
<td>Monitors all visible displacements, occasionally reverts to original site of object.</td>
</tr>
<tr>
<td>Substage 6 c. 18 - 24 months</td>
<td>New schemes devised through mental combination. Emergence of internal representation.</td>
<td>Words used as socialised symbols. Object permanence and vocabulary growth are synchronous. Ratio of correct usage to overextensions (functional similarity or family resemblances) minimises.</td>
<td>Takes invisible displacement into account; represents objects mentally. These mental images serve as one symbolic form providing the basis for language.</td>
</tr>
</tbody>
</table>

Table 3: Piaget's sub-stages of sensori-motor development. Based on Small's (1991:19-78) description of the characteristics of development during the sensori-motor phase, particularly her table 2.1 (pg. 21).

In this brief description of Piaget's sensori-motor period, we find evidence about the origins of the notion of object permanence as a precursor to language. Bryen (1981: 162-164) has suggested some cognitive factors for language development that can be derived from Piaget's description of the six stages of sensori-motor development:

First, in order to develop the 'inten' to communicate, the child must separate himself or herself from the environment and his or her actions. This separation from the
environment and actions, firstly enables the child to develop semantic categories, such as agent, action and object, and secondly predicates the combination of semantic categories, in many ways, to constitute sentences.

Second, the child must develop a 'permanent object concept' in order to understand that his or her world is more or less stable. The inability to have a permanent object concept would reflect in the words used to symbolise these concepts.

Third, through active experimentation or play with objects, the child firstly learns that objects can be acted upon and secondly that these actions can be different. Also, the child at the same time learns that objects have different properties and functions. As a result, objects acquire different meanings, which are subsequently mapped onto different words.

Fourth, the child learns the basis of symbolisation, i.e., Y can be represented by X. This act of symbolisation enables the child to separate the signifier (the word) from the significant (the object or event). (Anisfeld, 1984)

Fifth, an important achievement is the development of a child's 'mental representation', i.e., the ability to let one aspect of experience stand for, refer to or represent another. This ability to mentally represent the world is regarded to be the precursor to language, as early language appears to reflect the child's ability to use speech (words and simple sentences) to represent or refer to his/her understanding of the environment.

The central issue which is evident in Piaget's explanations of sensori-motor development (Table 3) is the notion of circular reactions - reactions that are based on feedback either from the environment or from the child itself. The child is in constant interaction with his or her environment, he or she receives some stimuli which instigate some action by the child and the child's existing knowledge then provides the necessary feedback to handle the stimuli. In this manner the child's knowledge increases with age (development that is time-varying) as he or she experience the surrounding environment. One can argue that it may be this circularity of reactions which, coupled with the child's biological endowment to respond to various stimuli leads to overall cognitive development.
Circular reactions are observed to be present in all stages of sensori-motor development, and indeed with age the influence of these reactions becomes even more pronounced, as one observes in the development of language: initially secondary circular reactions lead to babble, and a co-ordination of the same circular reactions is evident in the presymbolic use of words, however later due to tertiary circular reactions the child is now able to use words as personal symbols. This in turn leads to object permanency and the development of internal representation, both of which are relevant to language.

We argue that a computational model simulating child language development need to account for these time-varying circular reactions. Not only should the computational model be able to learn but the learning mechanism should be incremental in nature such that whatever is learnt is constrained by some feedback from the learning model itself, i.e., the learning model builds on its existing knowledge. In connectionist networks we find a learning mechanism that is based on feedback. Furthermore, the connectionist network as a data structure can be modified through experience. One can therefore argue that connectionist networks can simulate the kind of circular reactions that are so prominent in Piaget’s description of sensori-motor development.

2.2.2. ‘Assimilation’, ‘Accommodation’ and ‘Organisation’
Piaget’s theory of learning regards learning to be ‘developmental’ and to proceed through an ‘active interaction’ with the environment. According to this theory, a newly born child has no a priori knowledge of the environment or of the way in which he or she can act upon it. However, at birth children are equipped with perceptual sensory capacities and inborn reflexes, that is, an innate stimulus response relationship. The cognitive development of a child is made possible through an interaction between what Piaget calls ‘assimilation’ - a process by which perceptual stimuli are absorbed and interpreted in terms of what Piaget regards as existing knowledge, ‘accommodation’ - a co-occurring process whereby the ‘internal structures’ are adjusted to facilitate the assimilation of the new perceptual stimuli, and ‘organisation’ - a process that develops relations between internal structures. Further has further elaborated on assimilation and accommodation. He defines assimilation as being ‘the incorporating process of an operative action. A taking in of environmental data, not in a causal,
mechanistic sense, but as a function of an internal structure that by its own nature seeks activity through assimilation of potential material from the environment. The accompanying process of accommodation is defined as 'the outgoing process of an operative action oriented toward some particular reality state. Accommodation applies a general structure to a particular situation; as such, it always contains some element of newness. In a restricted sense, accommodation to a new situation leads to a differentiation of a previous structure and thus the emergence of new structures'. The dynamic interplay between assimilation and accommodation which, perhaps leads to cognitive and language development can be understood as follows: if a child is given a stimulus with which he or she is somewhat familiar, or has some experience of, then this stimulus is assimilated into the so-called existing knowledge. However, if the new information does not fit well into the existing internal structures of the child then the child attempts to accommodate this information in the internal structure by modification or extension. (Furth, 1969: 291-2).

Piagetian notions of assimilation and accommodation, synthesising biological growth and psychological development, have a computational interpretation, albeit rather a simplistic one. We argue that a computational interpretation of Piaget's notions of assimilation and accommodation need to incorporate data structures that can learn. By learning we mean that the data structures should have the tendency to modify or expand to incorporate new information and experiences by way of a continuous interaction with the environment. Traditional AI structures may suffice to represent knowledge, but they lack the ability to learn in a developmental, time-varying manner. On the contrary we find connectionist networks equipped with learning mechanisms that have some empathy with Piaget's notions of assimilation and accommodation (cf. Bechtal & Abrahamsen, 1991; McClelland, 1989).

2.2.3. 'Cognition' and Language Development - 'Cognition Hypothesis'
The Piagetian framework synthesises broader issues of biological and psychological developments, and although the framework provides a systematic basis for simulating child language development, one needs to focus further on the specifics. For instance, one has to answer to questions like what is it that children acquire: sounds related to concepts, objects, events and actions or the concepts are initiated by
'seeing', 'smelling' or 'tasting' objects, or do the actions inspire 'concepts'. Despite extensive inquiries, debates and experimental evidence over the centuries, one cannot find categorical answers to the questions posed above. Nevertheless, further explorations of developmental notions, proposed by Piaget and others, has led to the so-called cognition hypothesis.

Cromer has argued that 'at all stages of thought, the later as well as the earlier, it is cognition which affects language and not the reverse' (1991: 13). Cromer's cognition hypothesis thus predicates that the 'cognitive structures and operations make language acquisition possible' (1991: 57). It is perhaps possible to empathise with Cromer's hypothesis that language development chronologically follows cognitive development, that is, after the sensori-motor stage. It has been suggested that children formulate concepts much before the age of two. Piaget had argued that non-linguistic conceptual representation is a precursor to 'representational language', that is, language used to refer to past events and absent objects. Cromer appears to support this suggestion when he argues that 'children's concepts exist independently of and prior to even the simplest language used to refer to those concepts' (1991: 58), and he argues to the extent that cognitive development and possession of concepts are the necessary pre-requisites for language development or, as he at times calls it, 'language acquisition'.

Cromer's cognition hypothesis appears to be supported by a number of experiments on a range of linguistic activities during the language development stages. Cromer quotes experiments by Seigel (1978) regarding the hypothesis that concepts develop prior to language. Siegel's experiments required children to perform two tasks, a cognitive task and a language task. There were four possible outcomes of the experiments, that is, children passing both tasks, failing both tasks, passing the cognitive task but failing the linguistic task, and passing the linguistic task but failing the cognitive task. Of the four possible outcomes, Siegel reported the first three outcomes, that is, children either passed or failed both the tasks or in case they passed the cognitive task but failed the linguistic task. According to Cromer neither Siegel nor anybody else has reported that any child completed the linguistic task without passing the cognitive task, thus providing evidence to the hypothesis that cognitive development precedes the development of language.
2.3. Milestone Studies of Child Language Development

In this section we look briefly at some longitudinal studies in child language development with a view to use some of this data for simulation purposes.

Lois Bloom's semantic approach to language description

Bloom (1973) has argued that a child's language structure is a consequence of the child's efforts to communicate specific meanings. Bloom focuses on the context-related meaning of a child's early language, and has noted a number of short-comings in structuralistic explanations, based on transformational grammar in describing child's early sentences.

Bloom's longitudinal study comprises a study of three children from 9 months to 22 months. The observations range from first words (c 9-10 months) to syntax (c 22 months). Bloom transcribed children's speech output and annotated the transcription with a systematic running account of the non-linguistic context, including the child's own actions. Bloom found evidence that a child intends to express certain meanings with even his or her earliest sentences, meanings that go beyond simple naming, and that actually assert the existence of, or request the creation of, particular conceptual relations. Bloom's conclusions and findings were as follows: (1) A child's speech is very much tied to his or her situational context; (2) The classification of sentences is based more on the meaning of each sentence than on the details of its structure, as basic grammatical relations (subject, verb, object) are not present in children's early grammar; (3) The child's utterances, along with their context, provide the basis for assigning a 'semantic function' to a partial syntactic structure. This specification of semantic functions for sentences distinguishes Bloom's contribution; (4) Children showed evidence of knowing both substantive and functional aspects of the language. The children distinguished between the grammatical meaning of single words like mommy as subject (Mommy push), as object (Push mommy), as possession (Mommy's shoe), and as possessed (my mommy); (5) It is Bloom's contention that the reason children utter one word at a time is that they lack a linguistic code for representing information through semantic-syntactic relations between words. Children, in effect, develop conceptual representations of regularly occurring events and they tag on whatever words that conveniently code such conceptual notions.
Roger Brown and the cognitive pre-requisites for language development

Much like Piaget, Roger Brown (1973) divided the child language development period into five stages, labelled stage I - stage V, where the mean length of utterance (MLU) is the indicator of development. Brown proposed stage I to account for the period beginning from the emergence of the first multiword utterances and continuing until MLU reaches 2.0. Further stages are defined by increments of 0.5 to the MLU.

Brown performed a longitudinal study of three children during their sensori-motor stage, Adam, Eve and Sarah. His contribution is the characterisation of their two-word and three-word sentences. He named these sentences as 'telegraphic', because of their simplicity. Brown proposed that the first two stages of language development accounted for such sentences. The first stage involved the development of semantic roles and syntactic relations, and the second stage accounted for the development of grammatical morphemes and the modulation of meaning.

Brown proposed that embodied in children's utterances are a set of meanings that include nomination, recurrence, nonexistence, agent and action, action and object, agent and object, action or location, entity and locative, possessor and possession, and demonstrative and entity. Brown argued that this set describes the meaning expressed in children's utterances in stage I, and is the prelinguistic knowledge required to develop language.

Brown's study confirms some of the claims of Piaget and his followers. Brown's nomination and recurrence refer to the child's ability to recognise objects and actions, whereas nonexistence presumes the ability to anticipate objects and actions and the enduring nature of objects. Brown adds an interpretation to these developments: 'I think that the first sentences express the construction of reality which is the terminal achievement of the sensori-motor intelligence' (1973b: 200).

Two issues remain unresolved in Brown's study (Mandler, 1993: 281): first, whether this is the full range of possible meanings, and second a precise description of how the notions of objecthood, agency, actions
and location might be represented. These issues are still the focus of research by developmental psychologists, and answers are sought in computational paradigms as well, as for instance, connectionism.

Katherine Nelson's exposition of children's early words and their meaning

Katherine Nelson's (1973) study was longitudinal, studying the first 50 words acquired by a group of 18 children between the ages of 15 to 24 months. Nelson's approach was to identify the 'referential-expressive dimension' based on the way in which the words were used over time and in different contexts. Nelson found considerable uniformity in her subject's early vocabularies, and on that basis identified six categories of early words: general nominals, specific nominals, actions, modifiers, personal-social words and function words (listed in order of frequency). The strength of Nelson's study was that, within an explicitly longitudinal framework, she was able to demonstrate the differences between children at roughly equivalent levels of development.

An important contribution of Nelson was that she attempted with some success to show that the child's use of a word and the child's conceptual knowledge is not the same as the relationship that exists between an adult's use of a word and the concept represented by it. She concludes that, when children learn single words, their utterance reflects only partial correspondence between meanings attached to the word by an adult, thus explaining the over- and under-extensions in children's speech.

The role of function in concept development has been stressed by Nelson who distinguishes between the processes of concept formation and concept identification. The child, according to Nelson, is interested in what objects do, or what he or she could do with them, and he or she forms a concept which has a 'functional, action-based core'. However, identifying new examples of a concept was based on perceptual features - the semantic feature theory.

Nelson's speculated that children can be divided into two groups: the so-called object-oriented referential group that tends to acquire higher vocabulary than the self-oriented expressive group. The latter group appears to have fewer words in their vocabulary, yet a better grasp on grammar. Nelson argues that the
above differences notwithstanding, it appears that all children basically distinguish between objects and non-objects. This antonymy is the basis of a 'semantic tree' (Figure 5): the objects and non-objects are leaves on a tree. The object 'side' of the tree subdivides into animate and inanimate. The animate 'subcategory' is subdivided into people and animals. The inanimate subcategory into personal items, like toys, food and clothing, and impersonal items like vehicles and furniture.

![Figure 5: An exemplar of Nelson's semantic tree](image)

The referential/expressive or object/non-object distinction is, as other child language researchers including Nelson have pointed out, is not the only one, but one of the many dimensions of language development. Nevertheless we believe that such a dimension provides for a 'realistic' connectionist network simulation (cf. Section 5.2) of the semantic development of a child. The subcategorisation of the object and non-object categories is referred to as the creation of a partially ordered hierarchy: objects > animate > people > specific or non-object > object related > properties > relation.

From the point of view of a connectionist simulation we propose that Nelson's 'semantic structure' is highly relevant to classify 'object' and 'non-object' concepts at a considerable level of detail. Based on such a classification one can therefore, construct a semantic feature vector, for each object and non-object such that the components of the vector, individual semantic features, like people, animal, toys, action etc. may help us to decide what object/non-object the child is talking about. Later, in Section 2.4.1. we use Nelson's semantic structure to create semantic feature vectors for simulating concept development.
Although Nelson's referential-expressive dimension is used extensively for studying differences in the early language development of children, three major problems have recently been suggested by Lieven et. al. (1993: 288): first, there is a confusion in the literature as to whether the categories are functionally or formally defined. Second, there is no consistency in the use of the referential word categories. Third, expressiveness is defined, rather negatively, in terms of non-referentiality rather than in terms of its non-specific characteristics.

For our purpose, i.e., a connectionist simulation of child language development, the above studies inform us with key issues regarding child language. Bloom's semantic description of child language is useful in that it provides us with a set of conceptual relations that can be learnt by a connectionist network that can simulate the production of child-like one-word utterances. Likewise, Brown's semantic relations not only attribute meaning to children's earliest utterances but also enable us to build on Bloom's descriptions in demonstrating the transition of language from one-word to two-word sentences. Later, in Section 5.6 we simulate the learning of Brown's semantic relations in an unsupervised manner. As mentioned above, Nelson's semantic structures are also very relevant to us to represent children's concepts so that they can be learnt by a connectionist network (cf. Section 5.2.).

2.4. Children's Concepts

According to Piagetian scholars 'a concept is identical with an individual's internal structure or scheme and corresponds to the level of that structure. In its verbal manifestations, concept is a verbalised expression of a logical concept together with its verbalised comprehension; however, verbalisation is extrinsic to the logical concept as such' (Furth, 1969: 292). Now, if cognition precedes language development, particularly during the latter half of the sensori-motor stage, then it can be argued that it is these concepts about objects that children express in their initial one and two-word sentences.

Within a Piagetian framework, therefore, in order to understand the nature of children's concepts, and subsequently the development of concepts, it is necessary to understand how children organise their experiences in terms of concepts. This, in turn, implies the existence of concept representation schemes.
2.4.1. Concept Representation Scheme

The discussion of how concepts, words and meaning are related, or indeed are not or cannot be related, is a philosophical one in the broadest sense of the word. In this respect we follow Bierwisch (1970) who has pioneered the so-called 'semantic feature theory', and has considerable following amongst child language researchers. According to Bierwisch '[..] all semantic structures might finally be reduced to components representing the basic dispositions of the cognitive and perceptual structure of the human organism. [...] what is learned during the process of language acquisition, is not the semantic components, but rather their particular combinations in special concepts, and the assignment of phonemic forms and morphological properties of these concepts' (Nelson, 1974: 272).

Bierwisch's approach has been either adopted by or has a strong resonance in the works of child language researchers like Nelson (1973 & 1974), Clark (1978), Anglin (1983) amongst others. These authors have used the semantic feature theory for explaining how concepts are acquired and represented, and how a word is attributed meaning or how meaning is assigned a word. For instance, Nelson (1974: 272) has argued that children's concepts comprise a combination or 'bundle' of semantic features and can be defined by rules expressing the relations between the various semantic features derived from a child's experience with objects, people, and events. Semantic features may be perceptual (contours, movement, size, shape, colour, etc.), conceptual (existence, age, etc.). positive or negative, i.e., the presence or absence of a feature, or the features may be functional (state of objects, persons and events).

The semantic feature theory has certain implications for concept development, generalisation and categorisation. Accordingly, concept development involves the abstraction of single features, the creation of relationships amongst the features and an organisation of the abstracted semantic features such that jointly they represent a concept. This is achieved by attaching semantic features one at a time, from more general to more specific concepts. It may be noted here that in order to acquire new concepts, it is important that the semantic feature representations must be flexible and amenable to modification when novel experiences are encountered.
It should be noted that a complete set of semantic features has yet to be identified by researchers and the history of philosophy shows that this is indeed a task that seems not to be easily accomplished. Furthermore, it is not also fully understood how children learn which semantic features are critical to a concept. However, it appears that there is some evidence of the use of semantic features in the child language development literature. And, thus semantic features, represented as a vector in a multi-dimensional space, have a considerable computational appeal in that one can, in some approximate way, quantify meaning. We have made extensive use of semantic feature vectors for representing concepts, sounds and words. Connectionists including Hinton (1989), Hinton and Shallice (1989) and Dell (1986) have also used the notion of semantic features in their respective connectionist simulations. In fact, Hinton proposes that 'the simplest way to represent meaning of a word in a connectionist network is to use binary or real-valued semantic features and to dedicate a single "semantic" unit to each semantic feature. The meaning of the word is then a pattern of activity across the semantic units.' (1989: 16). However, we have found that in order to use semantic features it is important to further examine what constitutes a semantic feature, for example one may expand on the perceptual, conceptual, positive/negative features described by Nelson.

Representing concepts for a connectionist simulation of concept development

To represent concepts in a connectionist environment for a simulation of concept development (later discussed in section 5.2) we adopt the above-mentioned 'semantic feature' based formalism which describes the similarities and differences between various concepts and also helps in defining categories. Each concept in our connectionist representation scheme is represented by a 20-dimensional 'semantic feature vector' comprising two types of features: 'defining features' - determining a category structure, and 'individual features' - distinguishing individual concepts within a category. We discuss below how these defining and individual features are used to construct a semantic feature vector for representing a concept in our concept representation scheme.

Churchland and Sejnowski argue the relevance of the so-called defining features in categorising concepts as 'the more frequently a feature occurs in various input vectors, the more likely it is to be salient in categorising an input as belonging to a certain class' (1992: 102).
The defining features of concepts in our simulation are based on an 'object-oriented' taxonomy - a hierarchical semantic structure tree suggested by Nelson (cf. Section 2.4.1) which distinguishing features at each level. Recall that in Nelson's semantic structure (Figure 5) the object 'side' of the tree subdivides into animate and inanimate 'sub-categories'. The subcategorisation of the object and non-object categories is referred to as the creation of a partially ordered hierarchy:

- objects > animate > people > specific
- objects > animate > people > general
- objects > animate > animal > specific
- objects > animate > personal > toys
- objects > animate > personal > clothes
- objects > animate > personal > food
- non-object > object related > properties > relation

We believe that such an object/non-object distinction is extremely useful for a 'realistic' connectionist network simulation of the development of children's concepts. From Nelson's 'semantic structure' (Figure 5) one can construct a semantic feature vector, for each object and non-object such that the components of the vector, i.e. individual semantic features, like people, animal, toys, action etc. may help decide what object/non-object the child is talking about.

For representing various objects and non-objects a child might encounter, we have labelled Nelson's hierarchical structures in terms of binary digits (1 and 0). Two categories at the same level of the tree, for instance 'objects' and 'non-objects', are assigned the values 1 and 0, respectively. Similarly, for the category 'object', the sub-category 'animate' is assigned the value 1 and 'inanimate' is assigned the value 0. Once the tree is labelled (see Figure 6) one can determine the 'defining features' for a concept by using the value '1' to indicate that the object/non-object contains a particular feature and '0' to indicate otherwise. Therefore, according to this representation scheme, concepts belonging to the specific(1)-people(1)-animate(1)-object(1) category, for instance the concept 'dad' has a vector of defining features which is [1, 1, 1, 1]. Similarly, other conceptual categories are given their own labels, for instance the generic(0)-people(1)-animate(1)-object(1) category may be represented by the vector [1, 1, 1, 0].
Nelson's discussion is really at a meta-level in the sense that she talks about semantic categories, but there are no individual 'concepts' on this tree. The semantic structure is then useful for determining a set of 'defining features' that categorise objects into various concept categories. Bloom's (1973) assertion that children possess a variety of concepts, differing from one another in terms of salient features, implies that a category level abstraction alone may not suffice to represent children's concepts. Rather, individual concepts need to be analysed in more detail so that it becomes possible to further identify the individual concepts within concept categories. We argue that the features unique to a concept, i.e., the 'individual features' help discriminate one concept from other concepts having the same 'defining features'.

What seems relevant here is the specification of these individual features. Notwithstanding the juxtaposition of philosophical ideas about semantic features, we have collected a number of meaningful 'individual features' from various studies reported in child language literature. Commenting on the various aspects of children's language and behaviour, researchers have 'loosely' and implicitly mentioned some features that they regard have relevance to children's understanding of their environment. For instance, children are believed to distinguish various objects by observing aspects such as 'size', 'shape', 'colour' and even, at times, their 'function'. We have examined whether such aspects can be treated as individual (semantic) features that in turn can describe a concept. The individual features which we have collected adequately distinguish various concepts and hence serve the purpose of our connectionist simulation of the development of the concept memory.
In our connectionist representation scheme the individual features are based on a taxonomy of children's concepts suggested by Bloom (1973) and recently commented on by Anisfeld (1984). The taxonomy consists of seven different categories: objects, agents, events, states, locations, prepositions and 'function words'. Each category comprises a number of individual features that we believe may represent the concepts associated with the category. In Figure 7 we show a cross-section of the 'individual feature tree' for the category 'agents', mainly focusing on features related to 'Human beings'.

Now that we have derived individual feature trees for Bloom's (and Anisfeld's) taxonomy, it is possible for us to represent and incorporate within our connectionist representation scheme the various children's concepts reported by Bloom which are to be learnt during a simulation of the development of the concept memory (cf. Section 5.2). This is achieved by attaching concepts to the terminal nodes of the individual feature tree. The individual features of a particular concept, for instance 'dad', can then be obtained by translating the constituent individual features into binary digits:

Agents -> Human -> Human Beings -> Not self -> Familiar -> Does care -> Is Kin -> Male ->
Size (Large) -> Has name -> 'Dad'
Individual features for 'Dad' = [1, 1, 0, 0, 1, 1, 1, 1]

In devising a connectionist representation of concepts we have presented a synthesis of both Nelson's and Bloom's descriptions to devise a connectionist concept representation scheme that takes into account the 'defining features' that define super-ordinate categories, and the 'individual features' which uniquely

* Functions words are regarded as expressing personal intentions, commands and desires
identify Bloom's concept. In this way, the semantic feature vector encodes two types of information: super-ordinate category information (defining features) and concept-specific information (individual features). Of course, this is an open question in semantics and in philosophy. However, our reasons for attaching a feature vector to each concept comprising defining and individual features is purely pragmatic. Table 4 shows exemplar semantic feature vectors for the concepts 'dad', 'mum' and 'dog' created from a synthesis of Nelson's semantic structure and the individual features derived by us from child language literature. Such semantic feature vectors would be used by us later (Section 5.2) in a connectionist simulation of concept development.

<table>
<thead>
<tr>
<th>Concept Instance</th>
<th>Defining Features</th>
<th>Individual Features</th>
<th>Semantic Feature Vector</th>
</tr>
</thead>
<tbody>
<tr>
<td>dad</td>
<td>object - animate - people - specific</td>
<td>agents, human, human-beings, not self, familiar, does cares, is kin, male, large, has name</td>
<td>[1,1,1,1,1,0,1,1,1,1,1,1]</td>
</tr>
<tr>
<td>mum</td>
<td>object - animate - people - specific</td>
<td>agents, human, human-beings, not self, familiar, does cares, is kin, female, large, has name</td>
<td>[1,1,1,1,1,0,1,1,1,1,0,1]</td>
</tr>
<tr>
<td>dog</td>
<td>object - animate - animal - generic</td>
<td>agents, non-human, animal, is indoor, furry coat, unfamiliar, no distinct colour, has distinct, sound, medium, no name</td>
<td>[1,1,0,0,1,0,1,1,0,1,0,0]</td>
</tr>
</tbody>
</table>

Table 4: Exemplar semantic feature vectors for concepts - 'dad', 'mum', and 'dog' (defining features are given in bold type face)

2.4.2. Categorisation of Concepts

One of the main functions of human concepts is to categorise the world into objects, events, properties, and so on. More accurately, similar or close concepts are bound together by a structure known as a 'category', and the process of grouping similar concepts into categories is termed as 'categorisation'.

Children, some as young as 9 months, appear to 'categorise' or 'classify' concepts related to objects, behaviour or events as being equivalent to each other by utilising their perceptual abilities. The information acquired not only enables them to form object categories based on perceptual similarities. Furthermore, there is some evidence that children categorise the various sound patterns they hear, based on the underlying phonemic components of the words. If language development is assumed to depend on
the learning of new concepts and words, the consolidation/modified of old concepts and words, and the relationships among concepts and words, then this implies that categorisation is essential to language development.

A note on the criteria used by children in categorising objects, events and actions in terms of semantic features is relevant here. Nelson (1977: 128) stresses that the notion of decomposing concepts into primitive semantic features identifies the relations and similarities among various concepts, thus facilitating their 'categorisation' into categories. The semantic feature hypothesis suggests that objects in the world exhibit a large number of semantic features that enable them to be placed in many different categories, where every category has a principle or a rule that relates its members to each other. Therefore, in order to avoid cognitive 'overload' children must operate with some consistent principles of categorising the world in terms of perceptual and functional dimensions of objects and events.

Anisfeld (1984: 100) has suggested that at the end of the first year children perform an implicit sorting of objects into classes which becomes more sophisticated during the second year. Gopnik and Meltzoff (1987a, 1987b), and Sugarman (1983) have reported that there are striking changes in the manner in which children between the age of 15 and 21 months sort objects into categories. Gopnik and Meltzoff have coined the term 'exhaustive grouping' to describe this sophisticated behaviour, whereby 'children displace the objects to form separate spatially defined groups, each composed of objects from a different category' (1987a: 1092). Categories of concepts can also be seen as a means of extending children's knowledge about the environment: categories capture non obvious similarities among their member concepts, therefore by way of inductive inferences, children's knowledge about some member of the category can be extended to other members, and also to new concepts.

2.4.3. Development of Concepts

In Piagetian theory, an object concept is achieved when the child regards an object as a real, physical entity that exists and moves in the same space as the child. In psychology literature, concept
development among children is explained by a variety of learning strategies, such as learning by instruction, learning by deduction, learning by analogy, learning by induction which subsumes learning by examples, and learning by observation and discovery.

Concept development is a gradual 'process' and incorporates the application of formerly possessed knowledge to information experienced from the environment. Nelson (1977: 128), explaining children's ongoing conceptualisation process in terms of semantic features, suggesting that children perceive and analyse salient features of objects, people and events in their environment. The semantic features are put together into a representation structure, commonly termed a 'concept'.

Concept development is also viewed in terms of children's abilities to organise objects into categories. Markman (1989: 6) describes concept development in the following way: children, when learning a new concept, initially encounter a small sample of the extension of the concept, which in many cases may consist of a exemplar. For instance, a child might be shown a 'dog' and told that it is a 'dog'. One major assumption of this view is that children figure out the intension of a category from the sample of the extension of a category. Therefore, a child who both sees a 'dog' and hears its name, needs to analyse the dog's 'semantic features' which could be four legs, furry, brown colour, medium size, barks, and so on. A collection of such properties are deemed by the child as necessary and sufficient to describe the concept. As children make progress in determining the intension of a category, these semantic features allow them to make more accurate judgements about the extension of the category by evaluating potential members of the category. In summary, this view necessitates the presence of the following abilities to acquire a new concept: (a) analytic abilities to decompose objects into their constituent properties or 'semantic features'; (b) a hypothesis testing mechanism, for generating possible properties and evaluating them against new exemplars; and (c) a comparison mechanism whereby subsequent objects are compared with

\[1\] In Artificial Intelligence literature concept acquisition is described in terms of categorisation: according to Shapiro, 'an intelligent system must be able to form concepts, that is, classes of entities united by some principles. Such a principle might be a common use or goal, the same role in a structure forming a theory about something, or just similar perceptual characteristics' (1992: 249).

\[2\] The extension of a category is the set of objects that are members of the category, i.e., the set of objects that fulfil the criteria set forth in the intensional definition.

\[3\] The intension of a category is the set of attributes or features that define the category. It is sometimes viewed as the meaning of a 'category term'.
the intensional criteria of a category to determine category membership. This implies that the children may be aware of the degree of similarity of the new concept with other familiar concepts and hence can categorise the new concept into existing categories.

Finally, it appears that the overall strategy of concept development, comprising perception, analysis and categorisation, does not involve a teacher at all times, hence in connectionist learning terminology this strategy of concept development can be regarded as a case of ‘unsupervised learning’.

2.5. Children’s Early Words
What are children's first words, and indeed what do they mean? The search for an answer to this question has resulted in an ongoing investigation spanning several years with contributions from a large number of researchers. Candidate theories about the referential purpose of children's early words and their acquisitional profile have been presented by prominent child language researchers including Brown (1973), Bloom (1973), Nelson (1973 & 1985), Bruner (1975), Greenfield and Smith (1976) and Gopnik (1988). Child theorists argue that concepts and words are separate entities yet both are interrelated. For instance, Beilin has argued that ‘a concept in effect abstracts the constituent characteristics of that which it expresses whereas the words used in association with these concepts are verbal signs, and although they have their own characteristics and laws, they add nothing to the conceptual relations themselves. They merely designate conceptual articulation’. (1975: 343). Beilin makes reference to the notion of conceptual relations and later we examine the relevance of conceptual relations to children's early words.

2.5.1. Referential purpose
The appearance of words marks an important transition in children's communicative capability: previously, perhaps children could only communicate through crying, whining, gesturing and smiling. Children’s early words, i.e., the words uttered during the 9-18 month period, are communicative in nature and are regarded as 'one-word utterances', gradually combining to form two-word sentences.
Based on the various aspects of behaviour and context that children talk about in their initial utterances, Bloom divides children's first words into two categories or 'word-forms' - 'substantive words' and 'functional words'. Substantive word-forms rooted in the Piagetian object permanance hypothesis are object-specific and make reference to classes of objects and events based on their perceptual features. Functional word-forms make reference across these perceptually distinct classes of objects and events based on relational notions like "disappearance", 'recurrence', 'upness', 'cessation' and some others. Such a categorisation was earlier suggested by Sinclair (1970) and lately has been advocated by Small (1990).

Children might observe that although objects and events differ perceptually and functionally from one another, they nevertheless share certain 'behaviours', for instance objects and events 'exist', 'disappear', 'recur', 'are located' and so on. Although these behaviours may occur in relation to different objects and more so to different contexts, for the child they appear to have in common a particular feature - a unifying feature that could be represented by the meaning of 'functional word'. To Bloom it is this 'relational' nature of the meaning of functional words that accounts for their use in situations involving different objects, persons and events.

On the basis of each words' frequency, endurance, and most importantly referential purpose Bloom (1973: 85) has identified in Allison speech a small set of words which behave as functional words. Some of these functional words are 'more', 'there', 'up', 'uh oh', 'away', 'gone', 'no' and 'stop'. Not only did these words occur frequently and persistently in the period from 9-20 months. but each of these words were used in situations which shared features of behaviour and context.

It is interesting to note that in Allison's earlier speech spanning from 9 to 18 months functional words were noted to be dominant in terms of frequency of utterance, thereby indicating that Allison at that time was mostly talking about functional aspects of objects. However with increasing age (19 to 20 months)

*Bloom observes that Allison first used the functional word 'more' during a meal when she was offered a second portion of meal after she had already eaten the first portion. Later, Allison used the word 'more' to request the recurrence of food; to request her to be ticked again after been tickled earlier; also used to report another instance of an object in the presence of the first object.*
the frequency of the substantive words increased substantially compared with the frequency of the existing functional words. A comparison of the substantive and functional words (in Table 5) shows that whilst most of the functional words decrease in frequency of usage, the number of substantive words remains constant. However, words not in adult vocabulary appear to disappear altogether.

<table>
<thead>
<tr>
<th>Allison's Words</th>
<th>Frequency at 16 months</th>
<th>Frequency at 19 months</th>
<th>Frequency at 20 months</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Functional Words</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>there</td>
<td>30</td>
<td>21</td>
<td>3</td>
</tr>
<tr>
<td>up</td>
<td>27</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>more</td>
<td>24</td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td>down</td>
<td>22</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>no</td>
<td>21</td>
<td>23</td>
<td>20</td>
</tr>
<tr>
<td>gone</td>
<td>19</td>
<td>1</td>
<td>4 (away)</td>
</tr>
<tr>
<td>uh oh</td>
<td>7</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td><strong>Substantive words</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>baby</td>
<td>19</td>
<td>36</td>
<td>20</td>
</tr>
<tr>
<td>Mama</td>
<td>9</td>
<td>29</td>
<td>23</td>
</tr>
<tr>
<td>Dada</td>
<td>4</td>
<td>3</td>
<td>none</td>
</tr>
</tbody>
</table>

Table 5: Showing the frequency of functional and substantive words in Allison's speech spanning 9 to 20 months. The most frequent words observed at a certain age are highlighted

From the above discussion one can perhaps infer that the referential purpose of children's early words, mainly comprising functional words represent the 'organising activity' of the child forming basic action schemas with respect to objects in general, and that such words reflect the meaning of children's early sentences and are used in different situations that have a common context and behaviour. On the other hand, early substantive words represent the developing knowledge of categories of objects and are therefore names of object classes.

2.5.2. Children's One-word Utterances - Role of Conceptual Relations

The profusion of functional words in children's early speech suggests that children have acquired the meaning or relational nature of such words and use them as one-word sentences to communicate their internal 'intentions', 'desires' or even use them to comment about certain aspects of their environment.
By observing the earliest speech of various children, researchers noticed certain regularities in terms of
the presence of a small, yet consistent set of 'relational meanings' words (functional words). The claim
here is that in their earliest communicative endeavours children 'talk about objects (including persons):
the fact that objects exist, disappear and recur; they talk about actions: the fact that they can move and
act on objects; and they talk about locations: the fact that objects change locations relative to one another
in different contexts' (Bloom et al, 1985: 151). Once regularities in the relational meaning of children's
eyears were identified, researchers traced the conceptual antecedents of the semantics of early child
language to Piaget's description of sensori-motor intelligence in the first two years. Child theorists have
speculated that it is this awareness of object permanance which is mostly expressed by children in their
early speech, i.e., one-word utterances. Braine (1976: 92) have noted that at this stage children's object
related words refer to the disappearance, recurrence, plurality, location, request, negation, actions and
possession of objects. Bloom suggests that these are the kinds of conceptual relations children talk about
in their early speech and they persist throughout the development of language.

Bloom argues that the child's 'intention', manifested as function words could be predicted in terms of a
relatively small number of consistently recurring underlying 'conceptual relations': 'Children talked
about the existence, nonexistence, and recurrence of objects and events, the location of objects, the
possessive relation between persons and objects, and the relation between an action and the agent or goal
(locative or direct object) of the action. Thus the children were selective in what they chose to talk about,
and it was not the case that their utterances were merely the linguistic reflections of the many possible
interactions among persons, objects and events in the environment. Instead, it appeared that the
children's utterances manifested specific linguistic capacities with the ability to represent or code only a
limited number of conceptual distinctions' (1973: 24). From Bloom's accounts we have collected 25
contextual relations, 3 perceptual entities and 15 functional words. This information is now
systematically organised in Table 6.

* Similarly Gopnik have argued that 'a number of writers have noted that children's early words encode concepts that
correspond appearance and disappearance, relationships between actions and goals and spatial relationships, and that
these concepts appear to be related to the concepts that are developed at the end of the sensori-motor period,
between about 1;3 and 1;9' (1984: 496).
Table 6: Bloom’s conceptual relations with their corresponding function words

<table>
<thead>
<tr>
<th>Conceptual Relation</th>
<th>Perceptual Entity (ies)</th>
<th>Functional word</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disappearance</td>
<td>object, people</td>
<td>gone, away</td>
</tr>
<tr>
<td>Request, Recurrence</td>
<td>object, event</td>
<td>more</td>
</tr>
<tr>
<td>Existence</td>
<td>object, people</td>
<td>this, there</td>
</tr>
<tr>
<td>Existence</td>
<td>event</td>
<td>uh oh</td>
</tr>
<tr>
<td>Non-occurrence, Failure, Rejection</td>
<td>event</td>
<td>no</td>
</tr>
<tr>
<td>Non-existence</td>
<td>object, people</td>
<td>no</td>
</tr>
<tr>
<td>Greeting</td>
<td>people</td>
<td>Person’s name (Mama, Dada, etc.)</td>
</tr>
<tr>
<td>Cessation</td>
<td>event</td>
<td>stop</td>
</tr>
<tr>
<td>Pointing (to draw attention)</td>
<td>object, people</td>
<td>there</td>
</tr>
<tr>
<td>Actor</td>
<td>object, event</td>
<td>Person’s name (Mama, Dada, etc.)</td>
</tr>
<tr>
<td>State (large size)</td>
<td>object</td>
<td>big</td>
</tr>
<tr>
<td>State (small size)</td>
<td>object</td>
<td>small</td>
</tr>
<tr>
<td>State</td>
<td>people</td>
<td>dirty</td>
</tr>
<tr>
<td>Upness (Location)</td>
<td>object</td>
<td>up</td>
</tr>
<tr>
<td>Downness (Location)</td>
<td>object</td>
<td>down</td>
</tr>
<tr>
<td>Substantive</td>
<td>object</td>
<td>object’s name</td>
</tr>
<tr>
<td>Substantive</td>
<td>people</td>
<td>person’s name</td>
</tr>
</tbody>
</table>

Bloom’s predication of conceptual relations was supported by Schlesinger (1971), Bowerman (1973) and Brown (1973), although at times the same conceptual relations were named differently. It is interesting to note that such conceptual relations were not limited to children learning English; Brown in fact found evidence of conceptual relations in the data collected from Finnish, Russian and Chinese children.

We have used Bloom’s conceptual relations (the ad hoc list given in Table 6) together with the related perceptual entities and corresponding one-word utterances (functional words) to create 32 training (input and output) patterns for a connectionist simulation of the learning of conceptual relations and the production of one-word utterances (cf. Section 5.4). The input pattern comprises a conceptual relation and a perceptual entity, whereas the output pattern comprises the corresponding functional word. For instance an input pattern could be disappearance (conceptual relation) + object (perceptual entity) and the output pattern is the function word gone.
2.5.3. Acquisition of Words and the Naming Process

The development of vocabulary, or the so-called naming spurt\(^{10}\) (Gopnik & Meltzoff, 1992: 1092), and its dependence on the development of concepts, can again be related to the child's interaction with his/her environment. More specifically, it appears that the child develops an understanding of the 'semantic customs of his or her community' (ibid.). These customs, reflected in the perceptual, conceptual and functional features of objects, events, actions and so on, are in some ways learnt by the child, by forming an hypothesis about the categorical nature of objects, events, etc., and by testing his or her hypothesis by trying to name new objects, actions, etc. Cromer (1991: 221) has argued that the acquisition of words follows sensori-motor development, that is, the child develops the notion that he or she is an active person distinct from the objects that he or she comes in contact with. The distinction of self from the environment calls for communication, and to achieve this the child realises that every object, event and person has a 'name'. The learning of 'names' for objects, events and people then is viewed by researchers as a mapping task, where children are assumed to map linguistic knowledge on to their conceptual knowledge.

The naming of individual concepts is largely conducted through a process of ostension (Callanan, 1985; Ward et al., 1989). Ostension provides the child perceptual and conceptual cues to compare semantic features that might constitute a concept and a corresponding speech pattern related to a word. Ostension, in its basic form, involves the use of a "This is a --" construction while pointing to and labelling an object, person or event. The child in such situations is expected to focus on the object to which his or her attention is drawn by adults and also to simultaneously hear the adult's speech uttered at that moment. The child assumes the word he or she hears is the name of the object to which his attention is drawn, and that he or she is expected to subsequently associate the word with the concept of the object. The latter process is regarded as naming of the concept. 'Ostensive naming', can therefore be loosely regarded as either learning by instruction or learning from examples, thus providing an important basis for a connectionist simulation of the naming process.

\(^{10}\) Gopnik and Melzoff has argued that the 'naming spurt can also be thought of as a kind of categorisation behaviour' (1992: 1093). Also, Huttenlocher & Smiley (1987) and Bates et al (1979) argue that children as young as 9 months of age, comprehend names of objects, and spontaneously indicate objects in a category.
We believe that there are at least two circumstances when 'ostensive naming' can be simulated within a connectionist paradigm, particularly by algorithms and computational techniques used to conduct 'unsupervised techniques' (cf. Section 4.2). The first of the two circumstances relate to the assignment of a new word to a 'known concept' and the second to the assignment of a word to a 'novel' concept. In our connectionist simulation, naming therefore incorporates the creation of an association between a concept and word, in some cases the concept exists beforehand, whereas in other situation the child would learn a concept in addition to learning the word related to the concept.

2.6. Children's Two-word Sentences

In the previous section we have discussed how children's first utterances are attempts to give verbal expression to their conceptual knowledge about objects, people and events by uttering certain 'conceptual relations'. Children, in order to express their concepts in any language are suggested to possess 'semantic knowledge': 'the knowledge that makes it possible to relate sentences to world knowledge. This includes knowledge of individual word meaning, as well as how the meaning of a sentence is determined both by the meaning of the individual words and structure of the sentence' (Small 1990: 148).

Until now we have focused on the semantic based descriptions of child language because we find such a description consistent with Piaget's learning theory in general and cognition hypothesis in particular. A semantic based description of child language has certain merits which are also suitable for our purpose. We present here some advantages of the semantic based approach, as given by Harris (1990:42): (a) there is a general agreement that semantic descriptions open up more interesting and plausible explanations of early language development, Semantic based descriptions suggest links, first between early conceptual development (for example, Piaget's sensori-motor intelligence) and subsequent linguistic development, and second, between the meaning initially expressed in combinatorial speech and more complex

---

1. "The child is in possession of a concept but lacks the appropriate word to express it. According to Levine and Carey where 'the problem of word learning is one of mapping. The child's task involves mapping phonologically specified words which he or she encounters on to the concepts that constitute the meaning of these words for the child' (1982:645). It appears that in such situations the occurrence of the word relating to the concept would firstly initiate the process of learning the new word and later associating it with the existing concept.

2. "The child hears a novel word corresponding to a new concept. Brown hypothesises that, 'learning a new word typically plays a role in the acquisition of a corresponding lexical concept. On this view, the problem of word learning is not solely one of mapping. The child's task involves attaining the concept encoded by a phonologically specified word as well as discovering the correct mapping between the concept and the word'. (1972)"
sentences; (b) semantic descriptions do not invoke abstract grammatical concepts such as sentence subject, or a knowledge of lexical categories such as noun/verb. For this reason they are less likely to overemphasise a child's knowledge at any stage.

2.6.1. Emergence of Semantic Relations
It has been suggested that the transition from one-word sentences to two-word sentences is marked by children's ability to combine individual words in terms of the so-called semantic relations. In the two-word stage, children understand the functional and substantive aspects of an event and combine both the functional and substantive words to form two-word sentences. Early syntax, then can be described as a set of rules, the 'semantic relations', for combining function words with other substantive words. Brown (1973) has proposed a list of semantic relations which were inferred from children's two-word utterances. We have elaborated Brown's semantic relations by including Anisfeld's (1984) interpretation of the same together with some illustrative examples from Gardner (1978) given in Table 7.

Brown (1973) formalised such semantic relations with the formula $f(x)$, a fixed value, $f$, combined with a variable $x$ that can take many values. The meaning of the semantic relation $f(x)$ is determined by the meaning of the constant, $f$. For example, in the phrases 'more cookie', 'more juice' and 'more cheese', there is the constant functional word, 'more' combined with the variables $x$, 'cookie', 'juice', and 'cheese'. In all the above sentences the same semantic relation, recurrence, exists between the two words which is the actual or potential meaning of the word 'more'. The meaning of the resulting two-word sentence derives from the meaning of the functional word, and for that matter Bloom argues that the structural relationship between the two words in a sentence can be described as linear: the two words in the sentence are just joined together, their meanings are added, and no new meaning is to be inferred as a result of their combination.

---

13 For instance, when two-word sentences such as 'more cookie' occur in children's speech, the semantic relation between the two words 'more' and 'cookie' is inferred from the meaning of the functional word 'more' which is suggested to mean 'recurrence' of an object or event.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Agent + Action</td>
<td>Daddy eat</td>
<td>Action</td>
<td>People do things.</td>
</tr>
<tr>
<td>Agent + Object</td>
<td>Mommy truck (mummy push the truck)</td>
<td>Action</td>
<td>A person can perform actions on an object.</td>
</tr>
<tr>
<td>Action + Object</td>
<td>Give cookie</td>
<td>Action</td>
<td>Objects are acted upon.</td>
</tr>
<tr>
<td>Action + Location</td>
<td>Put table (put X on table)</td>
<td>Action</td>
<td>An action can occur in a specific place.</td>
</tr>
<tr>
<td>Demonstrative + Entity</td>
<td>That doggie</td>
<td>Demonstrative Naming</td>
<td>One of a set of objects can be specified.</td>
</tr>
<tr>
<td>Entity + Location</td>
<td>Daddy outside</td>
<td>Demonstrative Naming</td>
<td>An entity occupies a specific place</td>
</tr>
<tr>
<td>Possessor + Possession</td>
<td>My truck</td>
<td>Possession</td>
<td>People possess objects.</td>
</tr>
<tr>
<td>Attribution + Entity</td>
<td>Big ball</td>
<td>Attribution</td>
<td>Entities have characteristics.</td>
</tr>
<tr>
<td>Recurrence + Object</td>
<td>More milk</td>
<td>Recurrence</td>
<td>An object can be made to reappear.</td>
</tr>
<tr>
<td>Naming + Entity</td>
<td>Hi mommy, There truck</td>
<td>Demonstrative Naming</td>
<td>There exists a world of objects, whose members bear names.</td>
</tr>
<tr>
<td>Non-Existence + Object</td>
<td>No milk (non-existent or disappeared)</td>
<td>Negation</td>
<td>An object can disappear from a situation.</td>
</tr>
</tbody>
</table>

Table 7: Brown's semantic relations with exemplar children's two-word sentences and their meaning

Learning syntax then involves an observation of the recurring conceptual relations of existence, location, recurrence, disappearance and so on, and then their expression in speech. Thus, to utter the two-word sentence 'mummy door' the child first learns that persons like 'mummy' act and that objects like 'door' are acted upon - and not merely that 'mummy' occurs and then 'door' occurs. Therefore, it is the semantic relation actor-object that helps in the articulation of a two-word sentence. In our connectionist simulation, the development of language from the one-word stage to the two-word stage, or the evolution of syntax, is simulated as the learning of Brown's semantic relations. This may involve the creation of
appropriate associations among the various conceptual relations. The learnt semantic relations are then used to produce children's two-word sentence by combining words according to Brown's list of semantic relations.

2.6.2. Evolution of Syntax - Word-order

Braine (1976) argues that a close look at children's corpora reveals that there are some combinations which do not have positional consistency, i.e. similar word combinations with apparently free word-order. Braine gives two explanations for this kind of behaviour. First, the child might be trying to express a certain kind of meaning before he has learned rules specifying how that kind of meaning should be expressed. And second, the child might have learnt two rules, one for each order, thus, the child may possess rules \(\text{more} + X\) and \(X + \text{more}\). Braine's first explanation attributes free word-order to lack of learning, whereas the second explanation states free word-order as a result of learning. Braine speculates that this apparently free word-order is a result of the child 'groping' to express a meaning before he or she has acquired a sufficient set of rules to express it, and the emergent pattern is called a 'groping pattern'. Groping patterns have been reported to exist for a short period of time, and later a dominant pattern emerges. For instance, one of Braine's subjects', Andrew, produced utterances involving the use of the word 'allgone' in both combinations \(\text{allgone} + X\) and \(X + \text{allgone}\) (e.g. \text{allgone juice} and \text{plane allgone}). Later the word combination became ordered such that the order \(\text{allgone} + X\) disappeared (1976: 10).

Braine argues that the sentences produced by these formulae are consistent with adult language, hence the word-order of children's sentences must be a variant of the word-order of the language to which they are exposed. In Braine's corpora the most frequent positional patterns, that meet the statistical criterion include \(\text{see} + X\), \(\text{it} + X\), and \(\text{there} + X\) as patterns that draw attention to or identify something; \(\text{want} + X\) is a request form; \(\text{bye-bye} + X\) means that \(X\) has gone. The productive potential of such word-bound

\[\text{Schlesinger (1971) proposes that in order to account for novel adult utterances, children attend to the semantic relations of adult's utterances and abstract structural rules. For instance, when children hear adult sentences which contain two-word collocations, such as 'Mommy's shoe', 'Daddy's coat', 'Daddy's hat', they maybe are aware that Mommy and Daddy are people, whereas shoe, hat and coat are objects. Building on his or her knowledge the child would learn the sequencing of the words expressing the semantic relation possessor (+human) - possession (+concrete object).}\]
formulae reflect the fact that the child may insert a variety of words in the 'X' position to form two-word sentences. The formulae underlying children's sentences are, therefore, rules that map meaning into a surface structure (Bloom 1976:11).

We conclude the discussion of the prevalent word-order in child language by referring to Braine's (1976) suggestion that, we can infer from the productivity and semantic consistency of children's sentences as to where the words expressing the underlying component concepts are to be positioned in the utterance. Furthermore, the word-order formulae seem to be constrained by the pattern of adult speech heard by the child and thus provide information on word-order to some extent.

2.7. Towards a Psycholinguistic Framework For Language Development

The above rather extended discussion of the work on child language development should be regarded as a precursor to a systematic language informed simulation of child language development. Systematic in the sense that as language development is not instantaneous and is time-varied, a successful simulation would require a theoretical framework that in itself has stages and milestones. Hence, we have attempted to use a Piagetian description of child language development. This description has a neuro-biological plausibility, and regards language development to follow other (cognitive and motor) developments. The tenets of our psycholinguistic framework for simulating child language development using connectionist networks can be summarised as follows (shown in Figure 8):

a) The development of language depends strongly upon the development of cognition - the cognition hypothesis (Cromer, 1991). According to this hypothesis cognitive development precedes the development of language, and the child's conceptual knowledge about his or her environment is therefore a necessary base for the development of language. For language development the implication of this assumption is that 'preverbal' concepts are strongly related to later language development.

b) Cognitive development of a child is made possible through an interaction between two processes - assimilation and accommodation.
c) Cognitive development in a child takes place in 'stages'. There is a continuity in the child's knowledge, such that knowledge at one stage of development is utilised in later stages. Growth in knowledge is gradual and continuous.

d) The child is considered to be an active information processor of its environment, forming concepts about objects, people and events. To form an understanding about the environment, the child has to attend to, perceive, and store in memory observations made about his or her prevalent environment.

e) The child possesses a set of semantic features, and it appears that children tend to encode concepts through the use of a set of semantic features.

f) The child is involved in an on-going conceptualisation process, throughout his or her development. Conceptualisation is partly due to perceiving the environment in terms of semantic features, and partly due to biological maturity.

g) The child's lexical growth, i.e., acquisition of words is predicated by the child's ability to analyse phonemic information and store words in terms of their phonemic constituents.

h) Relationships exist between the child's concepts and words, such that, words are verbal manifestations of the child's conceptual knowledge. Both, the words and concepts are organised into distinct categories.

i) Child's initial sentences reflect their awareness about conceptual relations, such as recurrence, disappearance, location and others of the kind.

j) The transition from one-word sentences to two-word sentences is marked by children's ability to plan to manipulate the meaning of individual words in terms of semantic relations. Children use sentences to express semantic relations by structuring underlying concepts represented as words.

k) The presence of word-order in children's two-word sentences attributes a child the intention to communicate certain semantic relations. (Brown 1973: 41)
Figure 8: Our proposed framework for child language development based on the assumptions drawn in this chapter. Each assumption is numbered from a-k and we show these labels alongside the aspects they correspond to in the above figure.
Chapter 3

Connectionism: A Framework for 'Cognitive Modelling'

3.1. Introduction
The purpose of this chapter is to make an initial attempt to justify the use of connectionist architectures to simulate child language development. Our attempt is based partly on psycholinguistic descriptions of how language development proceeds, that is, the 'cognition hypothesis' and the Piagetian explanandum, and partly on a connectionist interpretation of language development. A connectionist interpretation of language development, in this chapter, would not only relate connectionism to Piagetian notions, that of assimilation, accommodation, and organisation (cf. Chapter 2), but we would also relate connectionism to psycholinguistic explanations of child language development.

We note that the development of a child's language can be expressed in terms of how a child 'articulates' what he or she 'knows', and, whatever the child 'knows' is a consequence of experience and biological disposition. What appears central to children's development of language is the notion of 'learning', which in a Piagetian sense is, 'learning is the acquisition of knowledge due to some particular information provided by the environment. Learning is inconceivable without a theoretically prior interior structure of equilibrium which provides the capacity to learn and the structuring of the learning process; in the wide sense, it includes both' (Furth, 1969: 294). The elaboration of the term learning, 'learning' involves a number of other terms - some are Piagetian in origin, like 'assimilation and accommodation', and others that are used by a host of related disciplines, include 'representation', 'knowledge', 'environment' and 'processing'. Note that these are cognitive terms discussed by various child language experts as follows:

First, the child is expected to learn, almost everything from his or her own immediate environment comprising language, social customs, survival techniques etc., through 'assimilation', 'accommodation' and 'organisation': covering experience and biological growth (Anisfeld 1984: 16-18).
Second, developmental psychologists suggest that there are three types of knowledge: 'sensori-motor' knowledge, e.g. knowledge of how to walk, 'imaginal' knowledge, for instance, ability to recall a face, and 'conceptual' knowledge or the abstract descriptive propositions 'concepts', that in some way characterise the member of specific categories (Small, 1990: 178 - 211).

Third, in order to store and subsequently retrieve what has been learnt, the child might use some abstract symbolic form; and it could be argued that for each of the three different kinds of knowledge, the child might use a different representation scheme.

Fourth, the child must be able to interrelate environmental inputs (and self-generated inputs), to a range of outputs the child can generate, and be able to select which of the outputs he or she will prefer; in other words, the child should have access to and control over mechanism to process input data (Bloom & Lashley, 1978: 267)

Fifth, Piagetian notion of development and more recent finding stress that early environmental experiences do tend to have an influence on intellectual development, and consequently upon language development, hence, one must take into account aspects of environment (Kagan, 1981).

In this chapter, we propose that the so-called notions of cognition - 'learning', 'knowledge', 'representation', 'processing', and 'environment' are essential for simulating child language development.

As an initial attempt to justify connectionism as an apt computational framework for simulating language development, we would provide here a connectionist interpretation to these above mentioned notions. Perhaps the success of our simulation might indicate that our connectionist interpretation has a psychological motivation.

Our connectionist interpretation of terms like 'representation', 'learning' and so on is based on the seminal paper of James McClelland (1988), in which he provides an introductory exposition of connectionism's influence on modelling aspects of cognition. Much of the interest in McClelland's exposition lies in the fact that he provides a connectionist interpretation of the 'micro-structure of cognition', such that both the neurobiological and psychological aspects of a cognitive activity are integrated in a single framework. In this chapter, we also contribute by extending McClelland's introductory exposition to a more detailed end, where the extensions are made in two dimensions:

(i) McClelland has mostly concentrated on the traditional 'back-propagation learning network'.

Connectionist literature, at present, discusses a variety of other connectionist architectures. We believe
that aspects of cognition should also be interpreted in terms of these other connectionist architectures. For this purpose, we would provide an interpretation of cognition that can be simulated by connectionist networks such as Kohonen maps, Grossberg networks and others. Furthermore, previous interpretations based on back-propagation learning networks would be extended to a much richer explanation.

(ii) McClelland’s interpretations are generic in nature and cover a variety of cognitive activities, without going into any specific details or providing a developmental profile of these activities. However, much of cognitive behaviour is usually understood in a developmental sense. We, however, would extensively provide a connectionist interpretation of a developing cognitive activity - child language development. We show by way of a connectionist simulation of child language development, a developmental profile that combines both theoretical and empirical investigations in this field.

3.2. Representation

*Representation* is defined in psychological parlance as a thing that symbolises or represents another thing, and to 'represent' means to make present something not present. A Piagetian elaboration of the term representation, involves two entities: 'a representing entity, a signifier and a represented entity, a signified' (Anisfeld, 1984: 7-14). Development of representation, whereby an experience can be stored and retrieved, is attributed by Piaget to the developments during the sensori-motor stage, and is suggested to lead the way to the subsequent development of language.

Connectionism, involved as it is in the simulation of learning and other cognitive 'faculties' in humans, redefines 'representation' particularly relating the term to the architecture of 'biological' and 'artificial' connectionist networks. McClelland has suggested that 'representations in connectionist networks are patterns of activations over the units in the network [...] Connectionist representations are truly active in the sense that they give rise to further processing activity directly, without any need for a central controller'. McClelland has elaborated further by arguing that *representation* delineates the endpoints in a continuum: one pole is *localist representation*, 'signifying a one-to-one correspondence between the information and the units which implement it', and the other is *distributed representation*, 'a pattern of
activation over an ensemble of units denotes an entity, i.e. a many-to-one relationship, and each unit is involved in representing many different entities' (1988:109-110).

Connectionists like Hinton (1986) and Elman (1989) maintain that *evolution* is built into a (distributed) connectionist representation. Connectionists argue that, connectionist network's plasticity, time-varying behaviour and its ability to respond to the environment facilitates learning. Learning leads to the development of 'internal' representations that are 'autonomous' and 'globally interpretable' within the network. Such representations are called internal because, (a) they are created within the 'plastic' structure of the connectionist network due to the learning mechanism, and (b) their creation is constrained by the fact that the connectionist network has no prior knowledge about the entity that is to be represented, rather the connectionist network extracts featural information from the environmental stimuli to develop the representation.

One can, perhaps, draw some parallels between the learning behaviour of the network and that of a child learning a language. The child appears to analyse and respond to the various inputs from his or her environment, including linguistic and perceptual input. There is, indeed, some plasticity that can be discerned in the behaviour of the child: error correction, belief revision, categorisation and so on.

Connectionist representations can be 'naturally' graded or in other words 'continuous', that is, the representations not necessarily comprise of binary 1 and 0 values. 'Naturally graded' representations increase the possible space of representation, thereby facilitating the network to generalise new representations to previously learnt representations, therefore allowing flexible processing, and adaptability to an unfamiliar input.

If we accept the semantic feature hypothesis, than it is possible to argue that connectionist representations can predicate 'categorisation' in a manner that is similar to the type of categorisation performed by a child. It can be argued that, 'automatic categorisation' by a connectionist network denotes 'intelligent' processing in that the connectionist network is not provided with a definition of the semantic features and the possible relationships among them. However, while developing the representations, the connectionist
network itself deduces the similarity among the various concepts and automatically categorises them. In a feature space, representations of conceptually close entities are mapped in proximity to each other, and can be deemed to belong to the same category. The boundaries of these emergent categories are 'flexible' and this is for two reasons, (i) they originate from experiences with the environment, which may be noisy at times, and (ii) the members of different categories share some features, allowing the categories to overlap, thereby incorporating the notion of 'family resemblances'.

William Ramsey (1992), a philosopher with interest in connectionism, suggests that if one is to understand connectionism's bearing, if any, on the philosophical views of representation, then what is needed is a taxonomy that classifies models in terms of (a) what is being represented, i.e., is it concepts, propositions, features or some other information, and (b) how this is represented, i.e., through individual unit, activation patterns or connection weights. In this respect, Ramsey has suggested four different types of connectionist representations, illustrated in Figure 8.

![Figure 8: Types of connectionist representations (Ramsey, 1992: 261)](image)

We summarise below Ramsey's (1992: 259 - 272) explanation of these four variations of connectionist representation.

**Distributed Representations (Type 1):** Such representations can be envisaged as the typical 'distributed representation', where single units represent microfeatures or low-level properties, and the overall activation pattern over a number of units represents a concept. Units may either be assigned microfeatures prior to learning, or alternatively the microfeatures are assigned by a learning mechanism.
Either way, a concept is deemed to be activated when its defining microfeatures are excited by an activation pattern.

Internal Representations (Type 2): This type of representation is very much like type (1) representation with the only difference that individual units do not have any semantic significance and only activation patterns are regarded to have a representational capacity. Type 2 representations are regarded as 'internal representations', typically generated during learning across the 'internal units' of a connectionist network.

Atomic Propositional Representations (Type 3): Such representations give a new dimension to connectionist representation by representing propositions. These representations are much like type 2 representations, however the ensemble of activation values on the internal units represent propositions, instead of concepts. Here again the individual units do not have any semantic interpretation.

Molecular Propositional Representations (Type 4): Such representation have originated as a result of attempts to implement some form of compositional semantics. The approach relies upon slight variations in activations patterns in order to capture propositional structure. In such representations, semantic units are represented by activation patterns, which can then be combined to form 'molecular' or propositional representations. The syntactic role a particular concept plays in the representation of a proposition is determined not by its relation to other concept units but, rather, by the way it is represented by its own activation values. The emergent representations are highly context sensitive and slight variations in activation values can encode syntactic or conceptual roles played by concepts in different propositions.

In terms of Ramsey’s characterisation of connectionist representations it appears that connectionist have usually favoured type (1) representation, for instance Dell (1988) has used 10 individual features to represent a concept, similarly in Hinton’s (1986) work each concept is represented as a collection of individual features, the so-called semantic features. We also in the same tradition have used a 20-dimensional semantic feature representation to represent a concept in our simulation of the development of the concept memory (cf. Section 5.2). Other connectionist for instance Miikkulainen and Dyer (1987 and 1988) have used employed the type (2)representation in their simulations.
3.3. Learning

'Learning' is an extensively studied subject, particularly in psychology and more recently in AI and Connection Science. The influence of psychology on connectionist learning has always been very direct: Hebbian learning, that is, to reinforce the connection weights between simultaneously active units, was inspired by Pavlovian learning models. More recently, some connectionists and philosophers of science, including McClelland (1988 & 1989), Bechtal & Abrahamsen (1991) and Arbib et al. (1987), have reinterpreted Piagetian notions of learning in the connectionist paradigm. Specifically, this reinterpretation focuses on Piagetian notions of accommodation and assimilation.

A Piagetian definition of human learning includes references to, environment, prior interior structure, capacity to learn and a learning process (cf. Section 3.1). And, according to McClelland, connectionist networks have the capacity to learn, as they employ a learning process that causes changes to the prior interior structure of the connectionist network, as a result of interaction with the environment (1989: 109).

Human learning takes place in a variety of ways, which may either involve the presence or not of a teacher or otherwise. In similar terms, connectionist learning mechanisms are broadly classified as supervised and unsupervised. Supervised learning, as its name suggests, requires an external teacher to monitor the correctness of the network's response to an input pattern. Unsupervised learning, on the contrary, does not require an external teacher, rather relies on built-in rules for self-modification that change the weights of the connectionist network in response to the input stimuli.

McClelland's essay on the implications of connectionism for 'cognition and development' includes the description of 'the learning principle' governing cognitive development: 'adjust the parameters of the mind in proportion to the extent to which their adjustment can produce a reduction in the discrepancy between expected and observed events' (1989:20). McClelland further notes that this learning principle 'captures the 'residue of Piaget's accommodation process', in that accommodation involves an adjustment

---

14 'Learning' is 'the process of acquiring knowledge or the actual possession of such'. More accurately, again according to Reber, learning is defined as 'a relatively permanent change in response potentiality which occurs as a result of reinforced practice' (1985:395).
of mental structures in response to discrepancies between an 'expected' and an 'observed' event. For McClelland, the novelty of this learning principle is the fact that it can be implemented using a connectionist network. In connectionist terminology, then, the above Piagetian notions of learning could be reinterpreted as follows (Table 8):

<table>
<thead>
<tr>
<th>Piagetian Constructs</th>
<th>Analogous Connectionist Notions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameters of the mind</td>
<td>Connections among units. Both entities are amenable to alteration due to experience.</td>
</tr>
<tr>
<td>Expected event</td>
<td>Desired pattern of activation over the network's output units.</td>
</tr>
<tr>
<td>Observed event</td>
<td>Actual pattern of activation produced over the network's output units.</td>
</tr>
<tr>
<td>Adjustment of the parameters</td>
<td>Connectionist learning processes that involve adjustment of connections.</td>
</tr>
<tr>
<td>Discrepancy reduction</td>
<td>'Error minimisation' process during connectionist learning, reducing error between expected and observed pattern of activation.</td>
</tr>
</tbody>
</table>

Table 8: Reinterpretation of Piagetian notions of learning using connectionist terminology

Indeed Bechtal, a philosopher of science, and Abrahamsen, a developmental psychologist, have similarly noted how connectionism may help in reinterpreting certain Piagetian constructs: 'connectionism is a modern mechanism for achieving stage-like states by means of the somewhat mysterious processes of assimilation and accommodation' (1991: 271). In this thesis we discuss how Piagetian notions of assimilation and accommodation can be implemented into connectionist networks simulating one particular aspect of cognition - child language development.

Assimilation, the Piagetians claim, enables the child to notice the similarities between a new event and some other familiar event and then respond to the new event in a manner that is characteristic to a similarly familiar event. The implication here is that the child 'generalises' the response associated with a familiar event to a similarly new event. Bechtal and Abrahamsen have argued that in connectionist terms assimilation can be reinterpreted in terms of 'the tendency of an interactive network to settle into the most appropriate of its stable states when input is presented to it; in Piaget's language, this is the schema to which the experience has been assimilated' (1991:271). Assimilation, then, can be interpreted as the ability of a connectionist network to generalise a new experience to some previously learnt experience by settling into a stable state which is associated with an old experience similar to the new experience. In some connectionist network architectures, new experiences are analysed in terms of their featural information and generalised to the most similar experience. Assimilation, in connectionist networks,
may thus be viewed as a new object, event, person or linguistic entity being recognised on the basis of the similarity of its representation to the previously learnt representation of an object, event, person or linguistic entity.

Accommodation refers to the child's ability to modify his or her internal knowledge structure to attend and to understand new sensory information. Again, Bechtal & Abrahamsen have suggested that in connectionist terms the Piagetian notion of accommodation can be reinterpreted as 'the changes in activations as well as weights that occur in order to assimilate the experience' (1991: 271).

Accommodation, then, can be interpreted as the adjustments in the connectivity pattern of the network in order to incorporate new information. Since knowledge in a connectionist network is contained in its connectivity pattern therefore connection weight changes lead to the accommodation of new information by reducing the error discrepancy between the desired and observed response of the network. A connectionist interpretation of accommodation requires that changes in the connectivity pattern of a connectionist network to incorporate new information may not result in the loss of previously stored information. In fact, learnt information should assist in the accommodation of new information, this then suggests a possible interaction between accommodation and assimilation such that firstly the connectionist network assimilates some new knowledge (i.e., recognises it in terms of learnt knowledge) and then accommodates the new knowledge (i.e., changes the existing connection weights) to learn the new information. Accommodation in connectionist networks may thus be viewed as the modification in the connection strengths among the units in order to incorporate a new experience, whilst also retaining the previously learnt knowledge.

We believe that, an advantage of the connectionist interpretation of Piagetian notions of cognition is that it can be implemented into connectionist networks and can be observed by simulating a variety of cognition oriented scenarios, for instance the simulation of concept development, language development, language production and so on. Indeed, Bechtal and Abrahamsen have claimed that not only can one reinterpret Piagetian constructs, but it might be also possible to augment and replace some of those constructs!
The learning attribute of connectionist networks not only distinguishes connectionist networks from traditional AI systems in which both knowledge and intelligence are statistically implanted or programmed by a programmer. In connectionist networks no knowledge is initially programmed, rather connectionist networks learn their own knowledge; 'connectionist intelligence', if there could be such a term, is then inherent in the fact that given a variety of information, connectionist networks on their own decide what to learn, how to learn and how to use the learnt knowledge both to learn new knowledge and to process the learnt knowledge. One can then argue that the learning capability of connectionist networks imparts a semblance of 'intelligence' to them. The argument goes that if human learning is regarded as a characteristic of 'intelligence', then we need to ascertain that whether connectionist networks also possess similar intelligent abilities as are attributed to humans. Intelligence is defined as by a developmental psychologist - Gregory (1987, 375-6) as follows:

(i) The ability to learn to adjust oneself to the environment, i.e., to adapt oneself adequately to relatively new situations in life.
(ii) The capacity to acquire knowledge, possess it and use it.
(iii) A biological mechanism by which the effects of a complexity of stimuli are brought together and given a somewhat unified effect in behaviour.
(iv) The capacity to learn to profit by experience.

From the above definition of intelligence one aspect that qualifies someone as being 'intelligent' is the ability to learn, adapt or reconfigure with respect to the environment. Indeed, the ability to learn is the hallmark of connectionist networks: one can therefore attribute a certain degree of intelligence to connectionist networks as no knowledge is programmed into a connectionist network, rather what is 'innate' are just the principles to learn or acquire knowledge from experience. When exposed to a learning environment a connectionist network utilises its learning principles to develop an appropriate behaviour conforming with the experiences encountered. Again, the behaviour acquired is not transient, but instead is retained in the connectionist network's knowledge base, i.e., the connection weights and is explicated through the network's response when experiencing a similar environment. Such attributes may certainly render some psychological plausibility and developmental realism in connectionist simulations of child language development.
3.4. Knowledge

Knowledge, in a Piagetian framework is defined in terms of the interaction between human knowledge acquisition and the environment, whereby the human behaviour is structured as a result of the interaction. In connectionism, McClelland stresses the role of knowledge in human behaviour, suggesting that 'crucial to the idea of cognition is the notion that information processing is guided by knowledge. [...] in connectionist networks, the knowledge is stored in the connections among the processing units.' (1988: 110-11). This view of connectionist knowledge representation is in contrast to the knowledge representation scheme of traditional AI systems, where knowledge is represented by symbols, propositions, and rules. Connectionism, offers an alternative to this declarative way of knowledge representation, that is, connectionist networks encode knowledge as connection weights among the processing units. Such a knowledge representation scheme is highly distributed, yet interconnected and interactive. On a philosophical note connectionism mainly due to the nature of its representation is regarded as a subsymbolic paradigm (Smolensky, 1988), a variant from the traditional AI symbolic paradigm.

The distinguishing aspect of knowledge stored in a connectionist network as connection weights is that it is not 'symbolic', and can be regarded as implicit and as inaccessible to examination. Knowledge stored in such a manner is explicated by the flow of activation through the connection weights, which in turn instantiates a pattern of activation over the processing units. In this way knowledge residing in the connection weights is made explicit by the 'dynamic processing' of the connectionist network algorithm. This then further implies that, the connectionist weights not only encode knowledge, but they also store information about how to control the flow of activation. There are different mechanisms to incorporate knowledge in connectionist networks. For instance, knowledge can be learnt through error correction techniques in a supervised environment (Rumelhart & McClelland, 1986), whereas in an unsupervised environment knowledge is acquired through self-organisation (Kohonen, 1984).

Now consider Piaget's 'schemas' with a view to understand how knowledge related issues are dealt with by the connectionist community. Rumelhart et al. (1986b) developed a connectionist network that incorporated schemas as the means to represent knowledge. 'Microfeatures' defining a schema were
represented as processing units, and the constraints among the units were encoded as weighted connections. Knowledge processing was based on constraint satisfaction techniques. Rumelhart et al. claimed to 'show that certain features that are problematic with conventional representations of schemata are better dealt with in the PDP [parallel distributed processing] language' (1986b: 21).

From a developmental perspective, an interesting aspect of connectionist networks is that, the increment of knowledge does not involve the addition of units to the existing structure, rather a modification of connection weights suffices to embody new knowledge in the existing structure. On the contrary in symbolic systems it appears that, to accommodate new knowledge there is a need to add an extra knowledge 'token' or 'unit' to the existing knowledge structure. In neurobiological terms, AI schemata appear rather implausible in that neurobiologists have observed that no new neurons or cell assemblies are added to the brain to incorporate new knowledge; rather what appears to happen in the brain is the strengthening or modification of existing connections among the neurons in response to environmental inputs. Therefore, it can be argued that, connectionism provides a more plausible mechanism for both representing knowledge and accommodating new knowledge, and these issues seem central to the simulation of the development of language amongst children.

3.5. Environment
Environment stands for the total physical and social surrounding of an individual organism. According to child language researchers working within a Piagetian framework, the child interacts with his or her environment by perceiving the wide variety of linguistic and non-linguistic stimuli present in his or her environment, and then assimilating and accommodating the relevant stimuli to his or her knowledge structure. It can be argued that a description of the environment is essential for any connectionist model simulating the development of language.

McClelland, describes the environment in which connectionist networks operate as 'an ensemble of possible patterns that might be presented to the network. These patterns are usually considered as separate events, however it is possible that a pattern may consist of a sequence of events' (1988: 112).
It is important to distinguish between what is meant by environment in a non-learning connectionist network and a learning connectionist network. For the non-learning connectionist networks, like interactive activation and competition networks, environment specifies the range and nature of the input patterns that can be successfully recognised by the network. Since non-learning connectionist networks do not evolve, and have a lesser degree of adaptability to unknown input patterns, thus it is important to restrict their overall environment to a known number of input patterns. For a connectionist network with learning potential, the environment determines what type of stimuli can be learnt by the network: the environment is divided into 'stimulus space' regions, where each region contains stimuli that have a common nature.

In a connectionist network the environment influences learning in two ways: firstly, the environment affects the extent of learning performed by the network, and secondly it determines the developmental nature of learning. It follows then, that an important part of simulating the development of any cognitive skill is to map out the experiences, expressed as input patterns that form the 'connectionist environment' for learning. The connectionist environment for learning is then characterised by two factors: (a) how the training patterns are presented to the network, and (b) how much the training patterns correspond to real-life data.

**How the training patterns are presented to the network:** The presentation of training patterns depends on whether the patterns are presented in a uniform or random order; whether they are presented in epochs of learning in a repetitive manner; and whether they are presented as one set of stimuli so that any regularities among the various patterns may be utilised. We performed some simulations to evaluate the affect of the above considerations and observed that for the same set of input patterns the learning of a connectionist network varied if the patterns were presented in a different manner each time. However, learning achieved by random, repetitive presentation of input patterns was found to be best, and we would use this scheme for simulating child language development.

If we correlate this connectionist learning environment with that the human learning environment then we can argue that our connectionist environment has psychological plausibility to a certain extent, both in
terms of repetitiveness and randomness. For instance, it is suggested by researchers that the verbal stimuli to a child is characterised by repetitiveness: ' [...] it was common in speech to children to maintain the same message across several sentences, varying minor features which do not alter meaning' (Ervin-Tripp, 1973: 281). Note also that the verbal stimuli presented to a child does not have a determined order. Rather, various sentences are spoken to a child appear in quite a random manner, mainly depending on the prevalent situation. We therefore propose that, for simulating cognition, the connectionist learning environment must incorporate a random presentation of the learning patterns to the connectionist networks.

**How much the training patterns correspond to real-life data:** The second factor regarding the connectionist environment for learning concerns the extent to which the learning data used in the simulation corresponds to real-life data. Our connectionist environment is made psychologically realistic by selecting training data from studies of child language development, especially the study by Bloom (1973) and attempts to incorporate both the verbal and perceptual experiences of a child. The child's verbal stimuli are adult sentences uttered in conversations with the child and are encoded as a two-word collocation of adult words. For the perceptual stimuli, we have adopted a distributed representation, that is, patterns of semantic features. In this regard we believe that our interpretation of connectionist learning environment appears to correspond with the child's environment for learning language.

**3.6. Processing**

Processing is an aspect of connectionist networks that is interrelated to many aspects of connectionist networks, such as learning, knowledge, representation and environment. McClelland argues that 'processing in connectionist models occurs through the evolution of patterns of activation over time. This process is governed by assumptions about the exact way in which the activations of units are updated, as a function of their inputs.' (1988: 110). The so-called evolution of patterns of activation during...
connectionist processing can be attributed to the updating of the activation level of units, resulting in new activation levels.

'Information processing' in connectionist networks is at some variance with information processing in symbolic systems. Symbolic systems process information interpreted as symbols and contained in a data structure, where the data structure may at times refer to a psychological construct, such as, schemata, images, and so on. Connectionist networks, instead, do not interpret information processing as a set of rules with associated symbols. Rather, within connectionism information processing is interpreted as the flow of activation among the units, resulting in the generation of an activity pattern over the units.

An important aspect of connectionist processing is the absence of a central controller that supervises the over all flow of information in a connectionist network. As stated earlier, since both the processing knowledge and world knowledge are stored in the connections among the processing units, there is no distinction between a representation and a processor in connectionist networks. These factors thus eliminate the need for a separate processor to monitor the overall processing of a connectionist network.

Child language researchers stress that children utilise information processing strategies in the acquisition of new and the retrieval of old knowledge. 'Acquisition' concerns the perception of semantic features, realisation of similarities between various concepts, and reorganisation of the old knowledge to accommodate the new knowledge. Acquisition employs the same mechanisms as involved in connectionist learning, that is, abstracting properties from a set of instances and 'accommodating' the learnt knowledge in the connection weights. 'Retrieval' concerns the matching of elements in the existing knowledge structure. During learning the connectionist networks become sensitive to regularities that are highly complex and difficult to conceptualise within any single rule. Retrieval of this stored knowledge is achieved by the so-called processing of connectionist networks. Retrieval comprises of a pattern of activation generated over a set of processing units in response to a retrieval cue. If one has a localist representation then the unit with the highest activation level is deemed to correspond to the retrieved knowledge.
The role of context in human information processing is very important. If by context we mean interrelated elements that have an influence on the behaviour of any other element, than similarly the context for a given unit in a connectionist network is the whole network. This is because in a connectionist network every unit is potentially affected by the global activation level of all other units, and as connectionist processing involves an interaction among a large number of units the role of context is made evident both in learning and retrieval. For a particular type of connectionist network, such as Kohonen map, context is a set of neighbouring units. During retrieval, these neighbouring units exercise their influence by incrementing the activation level of the appropriate unit. Similarly, during learning, this contextual influence is evident in the adjustment of connection weights among neighbouring units. We believe that these notions of connectionist processing would be relevant in our simulation of child language development.

3.7. Connectionist Learning: Algorithms and Structures

In this section we intend to provide a flavour of the key notion in connectionism - learning: learning through experience and learning through evolution. Textbooks on connectionist networks begin with statements like 'learning would involve relatively enduring changes in a system of given architecture that results from its interaction with the environment. The most obvious form of learning is adjustment in the weights of connections' (Bechtal & Abrahamsen, 1991: 270).

Connectionists classify the variety of connectionist networks into two broad categories - supervised learning networks and unsupervised learning networks. A connectionist network based on supervised learning algorithms is a network that is provided with an input pattern along with a desired output pattern. The learning law for such networks typically computes an error, that is the difference between the desired output of the network to its actual output. The computed error is then used to modify the interconnections between the units. Best exemplars of supervised learning are perceptrons and backpropagation networks. However, a connectionist network based on unsupervised learning algorithm is a network that is presented only with a series of input patterns and is given no information or feedback at all about its performance or desired output. Such training procedures are generally used for
categorisation or statistical modelling applications because the network's response cannot be predicted by the designer of the network. Best exemplars of unsupervised learning connectionist networks are Kohonen maps, competitive networks and Hebbian learning networks.

Levine's (1991: 196) distinction of connectionist networks based on supervised and unsupervised learning algorithms is worth noting here: In networks based on supervised learning algorithms 'certain output nodes are trained to respond to certain "exemplar" patterns, and the changes in connection weights due to learning cause those same nodes to respond to more general classes of patterns'. Whereas, in a connectionist network based on unsupervised learning algorithm 'input patterns are presented in some sequence and the network discovers through self-organisation a "natural" categorisation of the sensory world'.

A brief introduction to a variety of supervised and unsupervised connectionist learning algorithms is in order here vis-à-vis child language development simulation. A detailed description of the connectionist networks discussed below, together with exemplar simulations explicating their learning behaviour is given in Appendix A.

3.7.1. Unsupervised Learning Connectionist Networks

Learning in an unsupervised manner is characterised by the fact that such learning does not rely on the feedback of an external 'teacher' verifying the goodness of learning. Rather, unsupervised connectionist networks learn on their own without any explicit supervision: a type of learning which is seemingly more akin to some aspects of exploratory and spontaneous learning observed in a developing child.

**Kohonen Maps**

Trevo Kohonen (1984) has introduced a connectionist unsupervised learning architecture - Kohonen maps, based on the theory of **self-organising feature maps**. Kohonen regarded the self-organising feature

16 An artificial intelligence program that is able to modify itself by adapting to its environment without help from an outsider. It is able to profit from experience. (Mercadal 1990:257)
mapping theory as both a method of organising complex knowledge and a model of learning in biological systems.

Structurally, a Kohonen map consists of two distinct layers of processing units: an input layer and an output layer also called the competitive layer. The input layer is used to present an n-dimensional input vector to the Kohonen map. The output layer, envisaged as a 'two dimensional map', maps the n-dimensional input vector to a lower (two) dimensional representation. Units in the output layer are best understood as competitive units as during learning they compete with each other to represent the learnt knowledge. Both the input and output layers are connected by weighted connections, such that each output layer unit is connected to all input layer units.

Learning in a Kohonen map is based on a process of 'self organisation' which changes the weight of the connections between the input and output layer. This results in not only the learning of the input patterns, where each input pattern is represented by a unique output unit, but additionally it also realises a topological mapping of the input patterns. The self-organising process segregates the n-dimensional input space into distinct regions or 'topological maps', where each region contains similar input patterns. In the connectionist literature this is regarded as the 'automatic categorisation' of the input data. In this way, learning in Kohonen maps provide a topology preserving mapping of a high-dimensional input space to a low-dimensional output space, thus enabling one to understand the complexity of a high-dimension input space by viewing it on a two-dimensional graphical map.

In this respect, one can argue that Kohonen maps not only learn but also categorise the input data in an unsupervised manner through an 'innate' capacity for feature detection. These attributes, along with its neurological plausibility, make Kohonen maps a suitable connectionist network for modelling psychological tasks that require unsupervised learning and categorisation, for instance the development of concepts (cf. Section 5.2) together with the acquisition of their lexical labels, i.e. words (cf. Section 5.3).
Hebbian Learning Rule

Hebb proposed the earliest and one of the simplest learning rule for creating a connection between two processing units in a connectionist network, based on the biological notion that in the nervous system learning occurs by strengthening the connections between two neurons, provided they are active at the same time. The central idea was that if two neurons are simultaneously active then over time activity in one neuron will cause some activity in the other neuron. This can happen when the two neurons share an inter-cell junction, i.e. a synapse, therefore allowing one neuron to map its activity onto the other, leading to the development of an association between the two neurons. In neuropsychological terms Hebb's rule posits that 'when two adjacent neurons are repeatedly active then contingent metabolic changes lead to a lowered synaptic resistance between the two cells. This in turn increases the probability that activity in one cell will cause activity in the other' (Quinlan, 1991: 4).

The connections originating from Hebbian learning rule are termed as 'Hebbian Connections', which essentially are useful in simulating an association between two processing units. The strength of the Hebbian connection is then dependent on the activation level of the connecting processing units, such that two highly active units would be connected by a strong Hebbian connection. In a connectionist network based on Hebbian learning rule, the connection strength (the weight) between two processing units can be modified by increasing or decreasing in proportion to the activation level of the two units. Hebbian connections are extremely useful to manifest relationships between different entities represented as units, for instance we use them to implement a relationship between learnt concepts and words, i.e., for ostensive naming purposes (cf. Section 5.5).

Additive Grossberg Networks

The additive Grossberg network (AGN) is a single-layer connectionist network in which units are connected by two types of connections, a positive feedback connection with themselves and negative lateral connections with other units in the network. Units in an AGN are termed as 'feature sensitive', i.e., representing the various features in the input patterns in a localist representation. Learning is based on the Hebbian learning algorithm, and involves a 'competition' amongst the feature sensitive units: the units compare the features of the input pattern with their internal parameters and the unit with the best
match is deemed as the 'winner'. In this way different units learn different aspects from their input, and the learning can be regarded as a very simple form of abstraction.

AGN's can be used as a composite memory: whenever an input pattern is presented, positive and negative weight adjustments are made to connections so as to indicate which features did occur and which did not occur. In fact, strong connections based on Hebbian learning are made between the units representing the present features. This information about feature pairs, can be conveniently represented as the weights of the units, stored in a matrix.

Grossberg, the originator of these type of networks argues that the AGN's learning mechanism is seemingly similar to the learning mechanism present in the human brain. One can argue that AGN's can be helpful for memory oriented tasks, particularly when an association between two features of a pattern is to be learnt in an unsupervised manner. This is the kind of association which seems akin to the semantic relations found in children's language, in fact we would use AGN's to simulate a memory storing semantic relations between two features - conceptual categories (cf. Section 5.6).

**Spreading Activation Algorithm**

The spreading activation process is basically an content addressable memory retrieval mechanism proposed by Rumelhart and McClelland (1986). This algorithm is very popular amongst linguists, neurobiologists, neuropathologists and cognitive psychologists. The popularity is due to the reasons that the notion of spreading activation, leading to association of concepts and ideas, is very intuitive on the one hand and on the other hand has theoretical grounding in the associationist literature.

Retrieving information using spreading activation is constrained by the mutual interaction between two units, i.e., the preset or learnt connection weight between two units. During the spreading activation process the processing units in a connectionist network receive input from connected units and pass an output signal to other units. The input and output signals are 'activation levels'\(^{17}\) of the processing units.

---

\(^{17}\) There is associated with each unit a momentary level of energy or activation known as the "Activation Level" of the unit. The activation level is a real number, and for unit \(i\) at time \(t\) it is represented as \(a_i(t)\).
In this manner, the spreading activation process provides a useful mechanism to pass the activation of one unit to other connected units: the activation of an 'parent' unit, when spread through the spreading activation mechanism activates other units, where the activation of the recipient units is relative to the strength of association between the two units.

For cognitive modelling, spreading activation provides a useful mechanism to identify the relationships between two units either implemented in the same connectionist network or else in separate connectionist networks. We find the spreading activation mechanism useful to identify the relationships, implemented as Hebbian connections, between concepts and words (cf. Section 5.5.1).

3.7.2. Supervised Learning Connectionist Networks

Backpropagation Networks

The early pioneers of connectionism, for instance, Rosenblatt (1962) and its later proponents, like Rumelhart and McClelland, have emphasised how a connectionist network can be trained in a supervised learning environment. Rumelhart and McClelland (1986), developed a class of connectionist networks - Backpropagation (BP) networks that exemplify such supervised learning. Loosely speaking, the so-called backpropagation learning algorithm allows the establishment of arbitrary, non-linear relationships between input and output patterns.

BP networks cited in the literature are essentially variants of a 'generic architecture' which comprises a multi-layered topology, with an input layer, an output layer, and an intermediate layer, the 'hidden layer': the input layer is the network's interface with the outside world as it receives the (input) patterns to be learnt, the hidden layer applies some mathematical functions to the input patterns and passes the transformed pattern to the output layer. These layers are interconnected by weighted connections, such that the input layer is connected to the hidden layer, which in turn is connected to the output layer.
Learning in BP networks requires an input pattern and the corresponding output pattern, i.e., the desired learnt response to the input pattern. The BP learning algorithm based on an 'error minimisation' process aims to achieve a mapping of the input pattern to the output pattern, such that the resultant mapping is the 'best fit' according to some measure of error to the entire set of input-output patterns, i.e. the training data. During learning the inter-layer connection weights are modified to achieve the desired mapping. Learning in BP networks aims to achieve a balance between the ability of the network to respond correctly to the learnt knowledge (i.e. memorisation) and the ability to give reasonable responses to an input that is similar, but not identical to that learnt by the network (i.e. generalisation).

BP networks can then be characterised by their potential to handle complex pattern-matching problems, i.e., learning a pre-defined set of input-output pairs, recognition of complex patterns which may be both learnt or novel (generalisation effects), performing non-trivial mapping functions and creating internal representations that capture the idiosyncrasies of the learnt knowledge. We intend to use BP networks for learning conceptual relations (cf. Section 5.4) and word-order (cf. Section 5.7) as the learning of these aspects seems to be carried out in a supervised manner and it basically involves the mapping of information, for instance learning conceptual relations involves the mapping of conceptual relations to words.

3.7.3. Overview of Candidate Connectionist Networks
An overview of the candidate connectionist networks clearly suggests that, each connectionist network provides specific features that are desired in modelling child language development (Table 9).
<table>
<thead>
<tr>
<th>Connectionist Network</th>
<th>Prominent Features Relevant to 'Cognitive' Modelling</th>
</tr>
</thead>
</table>
| **Supervised learning** | · Development of 'internal representation'  
| Backpropagation       | · Pattern matching  
|                       | · Simulate staged development  
|                       | · Simulate assimilation and accommodation  
|                       | · Learning relationships between input-output patterns  
|                       | · Generalisation  |
| **Unsupervised Learning** | · Categorisation  
| Kohonen maps          | · Feature detection  
|                       | · Generalisation  
|                       | · Simulate staged development  
|                       | · Simulate assimilation and accommodation  
|                       | · Topology preserving mapping from high to low dimension  |
| Grossberg Networks    | · Simulate competitive learning to associate two items  
|                       | · Implement a matrix memory for simulating long and short term memory  |
| Hebbian Connections   | · One-one relationship between two units  |

Table 9: Prominent characteristics of various connectionist networks

Based on its characteristics, a single connectionist network can efficiently simulate, in a plausible manner, a certain aspect of language development. However, the functional, psychological and neurological constraints for language development are so enormous that it is not possible to simulate all aspects of language with just one type of connectionist networks, for instance some aspects require supervised learning while other may require unsupervised learning. This suggests that in order to perform a realistic simulation of language development, one needs both supervised and unsupervised connectionist networks. Furthermore, within these broad categories, connectionist networks should be selected by evaluating the psychological and neurological requirements for a particular task and the structure and processing scheme of the connectionist network.


Our discussion marks a shift from the general tendency of connectionists, and even psychologists to some extent, of discussing cognition at a generic level. We have initiated a discussion at a rather specific level,
discussing the potential of connectionism to interpret a specific cognitive activity, that is, development of language in children. Indeed a connectionist simulation requires an in-depth analysis of the aspects of child language development within the realms of connectionism. And, it is this detailed analysis of connectionism which would provide further evidences regarding the plausibility of connectionism to both represent aspects of human cognition and to simulate their development.

We are aware of the ongoing debate concerning the relation of connectionism to cognitive theory and modelling. This research has elicited contributions from prominent connectionists including, Schneider (1987), McClelland (1988), Pinker & Prince (1988); Lachter & Bever (1988); Fodor & Pylyshyn (1988); Smolensky (1988); Bechtal & Abrahamsen (1991), Seidenberg (1993) and many others. Ardent connectionists have frequently proposed connectionism as an apt framework for cognitive modelling.

Consider Rumelhart and McClelland’s recommendations in this context: 'They hold out the hope of offering computationally sufficient and psychologically accurate mechanistic accounts of the phenomena of human cognition which have eluded successful explication in conventional computational formalisms; and they have radically altered the way we think about the time-course of processing, the nature of representation, and the mechanisms of learning' (1986:10).

Connectionism can certainly not be limited to just simulation purposes, rather it is best to view it as an ensemble of tools and theories that can be used to achieve different ends. One approach is to regard connectionism as providing a set of general theoretical principles that can be applied in a variety of domains. This can be seen as the philosophical dimension of connectionism when construed in this way, connectionism contributes to the development of theories; the emerging theories encompassing a variety of domains are explanatory in nature and not merely descriptive. This can be attributed to the fact that connectionist theories can be realised and simulated. Simulation, then, can be regarded as another use of connectionism. Connectionist simulations may include simulations for forecasting, scheduling, planning and control tasks, along with the use of connectionism for cognitive modelling. Here, connectionism functions as a statistical tool analysing a complex set of data. For cognitive modelling, connectionists recommend connectionism as a set of general principles, primitives, structures and approach that appears to provide a computational framework that is psychologically and neurologically plausible. In this
context Seidenberg argues that, ‘connectionist modelling should be seen as a framework that supports a variety of theoretical activities. It can, of course, be used as a tool for exploring descriptive theories. Although this is an important function, connectionism is by no means limited to this role. Rather, connectionist models serve a number of other useful functions. They generally expand the vocabulary of knowledge representation, processes, and learning mechanisms; encourage rigorous and explicit theories and provide a stronger link between theory and data’ (1993: 234)

Bechtal and Abrahamsen (1991: 269-278) have extended the argument regarding the plausibility of connectionism networks by arguing that not only connectionism can be seen as a vehicle for operationalising cognitive developmental theories, but that the developmental nature of connectionism could have an implication on various areas of cognitive theory, in particular cognitive development and linguistic development. The possible implications of connectionism are due to their (a) new interpretations of traditional constructs; and (b) more relevant for us the possibility to explore an area of concern by means of connectionist modelling using developmental data. To Bechtal & Abrahamsen the aptness of connectionism for modelling developmental cognitive aspects is validated by its ability to implement developmental aspects such as (i) the various mechanisms of development - assimilation and accommodation, (ii) the evidence of stage-like changes during development, (iii) the incorporation of context effects, i.e., the role of an environment, and (iv) the ability to account for developmental disabilities.

To us some of the features of connectionism that are relevant to simulating child language development are as follows:

- Connectionist networks exhibit intelligent behaviour without storing, or otherwise operating on structured symbolic expressions.
- Connectionist networks represent concepts about any entity not as an explicit data structure, rather, concepts are represented as patterns of activation, distributed throughout the network of processing units.
- Connectionist networks provide an alternate way of evaluating hypotheses. The units in the network stand for possible hypotheses about some thing; the activation level of the unit stands roughly for the strengths associated with the different possible hypotheses, and the constraints between hypotheses can be encoded in the networks as positive or negative connections.
• Connectionist networks provide representations that can be made interpretable by a variety of structurally divergent networks.
• Connectionist networks have been formalised using mathematical techniques.
• Connectionist networks can incorporate operations such as, best fit search and constraint satisfaction.
• Connectionist networks have exhibited non-trivial learning; they are able to self-organise, given only examples as inputs. For novel inputs connectionist networks degrades gracefully. Also, connectionist networks are more fault tolerant and error-correcting.
• Connectionist networks have a processing scheme that is an alternate approach to symbolic rules.
• Connectionist networks, automatically extract distinct features from the input while learning, and categorise the input data into distinct categories.
• Connectionist networks provide the mapping of a high-dimensional input space to a low-dimension space.
• Connectionist networks suggest an implementation of multiple flexible constraints.
• Connectionist networks can generalise to novel, incomplete or ambiguous input.
• Connectionist networks provide a dynamic updating of stored knowledge.

Now, prima facie, connectionism appears to be an apt framework for simulating aspects of cognition, for our purpose child language development. The connectionist framework available to us for modelling purposes encompasses a variety of representation schemes spanning from the localist to distributed; learning both in a supervised and unsupervised manner captures the incremental nature human development and is realised by the change of connection weights; knowledge is non-propositional and is encoded in connection weights; environmental influences during learning are incorporated by the manner in which training experiences are encountered by the connectionist network; processing is distributed in nature and aims to learn information or to explicate the learnt information.

In this chapter, we provided a connectionist interpretation to the basic aspects of cognition which are relevant to the simulation of child language development. We also presented a brief description of the type of connectionist networks available to us to perform a simulation of child language development. In the next chapter, we proceed to develop a 'hybrid' connectionist model for simulation child language development.
Chapter 4

The Connectionist Implementation of a Child Language Development Model - ACCLAIM

4.1. Introduction

Learning is a much debated topic in artificial intelligence, neurobiology and linguistics. Assuming that language is unique to human beings, the so-called natural language, then it can be argued that the development of language amongst children can provide us with pointers to a number of 'open questions' in the above-mentioned subject areas. For instance, the notion related to the so-called semantic primitives used to elicit and organise child language output has parallels with studies in artificial intelligence and linguistics; the role of the environment in determining motor and cognitive development has been studied extensively both in psychological and neurobiological studies and the studies in child language development are at once an exemplar of and an inspiration to, workers in psychology and neurobiology.

We believe that connectionist networks, or the so-called neural networks, provide a basis for studying child language development in that these networks emphasise learning, either from observations or from being 'told', and that the design of these networks simplifies questions related to the representation of linguistic and 'world' knowledge through the use of a network of 'simple' nodes and links. Our claim is that the role of simulation programs that exhibit some kind of learning, are particularly relevant to child language development.

Indeed, both connectionists and psychologists have envisaged connectionism as a vehicle to operationalise existing theories of human cognition and linguistic behaviour. There is some evidence for this attempt at operationalising theories in the literature, particularly in relation to the simulation of a number of cognitive activities. Notwithstanding the success of such earlier connectionist simulations, it is evident that researchers had aimed to simulate a particular aspect of cognition. Also the connectionist
architecture used to perform such simulations could be regarded as rather simple in psychological and neurobiological terms. Over the years connectionism has certainly matured, in a theoretical sense new connectionist networks and novel learning mechanisms have been formulated, also the philosophical implications of connectionism are seemingly now more well-grounded. We believe that now when the efficacy of connectionism is widely accepted, the scope of cognitive modelling using connectionist network need to be expanded.

4.2. What Do We Learn From Previous Connectionist Simulations

Many connectionist researchers have attempted to simulate linguistic behaviour. The literature on the connectionist simulation of language, more accurately, aspects of adult language, clearly indicates that many connectionist simulations are carried out using a single connectionist network and learning principle, albeit the scope of the investigation based on connectionist simulations seem rather limited; questions asked and answers sought were almost always dedicated to one particular aspect of human cognition. Prominent among the many connectionist simulations is the pioneering work of Rumelhart and McClelland's (1986) past tense learning model. There are a number of other researchers who have actually worked in various other natural language processing areas which includes morphology, phonology, semantics, lexical processing and associative memory (Table 10).

<table>
<thead>
<tr>
<th>Area of Research</th>
<th>Connectionist Simulation</th>
<th>Researcher(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lexical access.</td>
<td>Cottrell (1985)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kawamoto &amp; Anderson (1984)</td>
</tr>
<tr>
<td>Phonology Learning</td>
<td>Acquisition of phonology</td>
<td>Plumkett &amp; Marchman (1989).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Watson (1987)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Weber (1989)</td>
</tr>
<tr>
<td>Language Production</td>
<td>Sentence production based on lexical processing</td>
<td>Dell (1986 &amp; 1991)</td>
</tr>
<tr>
<td>Syntax</td>
<td>Net-linguistic 'Earley' Parser</td>
<td>Schnelle &amp; Doust (1992)</td>
</tr>
</tbody>
</table>

Table 10: Some prominent connectionist simulations of linguistic behaviour
Note that much of the work reported above focuses on a single aspect of adult linguistic behaviour, and not the learning of language. Most simulations are based on a multi-layered backpropagation network, a controlled feedback loop, implementing a supervised learning algorithm (cf. Section 3.7.2.) that enables the network to 'learn' or 'mimic' a particular aspect of adult language. The success of the simulation was the ability of the connectionist network to learn associations between a set of input and output patterns, and to match the output of the network with existing psycholinguistic data. The discussion of how the input and output was encoded is not very clear in the above mentioned simulations, however, one could surmise that the input was encoded beforehand, i.e. words, sounds or morphological input was encoded as binary or continuous patterns which the connectionist network then learnt. If we further examine the processing of the connectionist network used in each of the above connectionist simulations, it becomes clear that,

a) The connectionist network did not require the manipulation of simultaneous multiple constraints.

b) The connectionist network did not communicate with any other connectionist network to receive external information. This implies that the connectionist network relied on whatever knowledge it possessed to produce the desired linguistic behaviour. Note that a connectionist network can at best possess one type of knowledge whereas in reality a linguistic behaviour may require more than one type of knowledge.

c) The connectionist network could use only one form of representation, that is, either localist or distributed.

d) The connectionist network employed just one type of learning which was mostly supervised in nature, therefore we cannot draw any conclusions about the 'unsupervised learning' aspects of adult linguistic behaviour, and particularly the involvement of 'unsupervised learning' in the development of language amongst children.

e) The strategy employed by the above researchers is seemingly valid for simulating 'low-level' cognitive activities, however if one needs to simulate a 'high level' cognitive activity which involves an interplay of a variety of cognitive aspects the prevailing simulation scheme would certainly prove inadequate.
A claim common in the AI literature is that, given the right choice of representation schemata, for example semantic nets or frames, and a reasonable reasoning algorithm, such as backward chaining or forward chaining one can simulate aspects of 'cognition': learning, memory, knowledge, environment, etc. which we discussed in the last chapter. McClelland (1989) demonstrated in his seminal paper that connectionist architectures can also represent or more precisely can be used to simulate aspects of cognitive behaviour. McClelland, and his colleagues including Rumelhart, have pioneered the 'micro-cognitive' approach which they regard as psychologically and neurologically more plausible. One key aspect of their approach is the 'multiple simultaneous constraints' which are deemed to be important for all types of cognitive activities. In language processing, multiple simultaneous constraints manifest themselves as the interaction between the morphological, syntactical, phonological, lexical and semantic constraints, the choice of words, sentences and sounds. Therefore, if one has to simulate aspects related to language then it is essential to have the notion of multiple simultaneous constraints as explicable as possible and perhaps exploit the knowledge of multiple simultaneous constraints in order to build large scale language processing connectionist models. Again we note that none of the above simulations actually make a claim of exploiting multiple simultaneous constraints.

Psychologists, and in particular developmental psychologists, have consistently argued that human development, which may include the development of language, sensori-motor control, visual recognition, and object permanence are achieved through different learning mechanisms. The learning mechanisms discussed in this literature include error correction, classical conditioning, self-organisation and pattern classification. Therefore, to perform a realistic simulation of language processing one needs to (a) incorporate a variety of learning mechanisms; (b) manipulate a variety of inputs - perceptual, verbal and functional, etc.; and (c) include both localist and distributed representation schemes. We note that all these aspects are lacking in earlier connectionist simulations of linguistic behaviour.

It therefore becomes clear that due to the overall complexity of language development one would require a number of multi-layered connectionist networks: one to learn to process lexical input and output, yet another to learn phonology and more networks to learn concepts, semantic relations and word-order. It is apparent, from previous connectionist simulations, that although the structure and processing scheme of a
single connectionist network may produce desirable results yet there are other significant limitations for simulating a high-level cognitive activity, for instance child language development. We argue that such limitations can be overcome by developing large scale 'hybrid connectionist models' based on a 'modularity' principle which advocates the following tenets:

- Decompose a 'complex cognitive task' into simpler 'sub-tasks'.
- Simulate each sub-task through a single connectionist network.
- Develop a large scale 'hybrid connectionist model' Figure 9 by synthesising several individual, yet interactive connectionist networks.

According to our modular approach the individual connectionist networks process information in their own manner, yet are integrated in a 'hybrid connectionist model' comprising different learning algorithms and networks. In a hybrid connectionist model the individual connectionist networks retain their structural distinctness and can be viewed as independent 'modules' of a model. In order to implement a hybrid connectionist model, we have developed a seven phase strategy given below:

Figure 9: An exemplar 'hybrid' connectionist model comprising a number of connectionist networks integrated together.
(a) identify the sub-tasks constituting a complex cognitive activity. Each connectionist network simulating a sub-task is to be regarded as an independent module of a connectionist model.

(b) design appropriate connectionist networks that can simulate the sub-tasks. The design metrics are the number of layers, number of units in each layer, connectivity pattern of the layers and the activation function of the units.

(c) develop a knowledge representation scheme that can be shared by other connectionist networks, i.e., the knowledge stored in one network is accessible to other networks in the model.

(d) establish a communication mechanism among the connectionist networks so that information is accessible throughout the connectionist model.

(e) train each connectionist network with its respective stimuli separately, also there should be provisions to train more than one network simultaneously.

(f) represent explicitly the knowledge learnt by each connectionist network, such that it has some significance to an external observer.

(g) formulate a processing scheme that may synchronise the overall operation of the connectionist model. The processing scheme should retain concurrency, enhance the processing strengths of various networks and at the same time avoid unnecessary cross-talk (influence) between the connectionist networks.

4.3. A Psycholinguistic Model of Child Language Development: Basis of a Connectionist Model

Recall our discussion of the work of major child language researchers including Brown, Bloom, Cromer and Nelson, all working within a Piagetian framework and focusing on the famous 9-24 months landmark in the development of language among infants.

Based on the suggestions of child theorists consider the following model of child language development:

(i) The child receives two kinds of input from its environment: one is perceptual in nature and enables the child to categorise entities and events and the other is the caretaker input, mainly two-word collocates.

(ii) The 'innate' ability of the brain then helps the child to 'abstract' the features of the environment as a set of concepts and string of 'phonetic features' constituting words. (iii) In a supervised learning environment, the child 'builds' up a 'conceptual relation' memory of entities and events, and 'learns' to produce one-word sentences. (iv) Furthermore, in an unsupervised manner the child 'learns' to associate concepts and words. (v) The child generalises further and creates the so-called conceptual categories.
leading to the development of semantic relations among conceptual categories. (vi) Finally, the child through a process of trial and error, builds up collocates that conform to the word-order in his or her language.

The above-mentioned 'processes' or 'sub-tasks' can be further distinguished by demarcating the environmental considerations from what can be regarded as the 'innate' ability of the brain to learn language. Put crudely, we are arguing that some of the above-mentioned processes can be simulated by individual connectionist networks or 'connectionist structures'. Some of these processes can be simulated by the so-called connectionist 'supervised learning' algorithms whilst others can be simulated by using 'unsupervised learning' algorithms. Our argument here is that since the child is suggested to employ a variety of learning mechanisms during language development a plausible approach to simulate language development would be to include in the connectionist model all available connectionist learning algorithms that have some parallels with the child's overall learning strategy. This would lead to a 'hybrid connectionist architecture' that would involve a number of connectionist networks, each incorporating a learning mechanism that corresponds, in a restricted sense, to some psychological account of the child's learning of language. We believe that in adopting the above strategy we might be able to develop a more psychologically plausible connectionist architecture for modelling child language development, and that in this way we would also be exploiting the wide resources offered by connectionism.

The implied discussion of the above mentioned sub-tasks predicates the existence of a variety of connectionist modules18 each simulating a particular sub-task. Table 10 lists the various connectionist structures that are needed to implement the above-mentioned model of child language development, together with a specification of each connectionist structure's typical input, the effect of learning and the typical output produced by it.

---
18 We regard a connectionist module comprising a number of connectionist structures, whereas a hybrid connectionist model is implemented by synthesising a number of connectionist modules.
### Table 10: The various connectionist structures implementing the above-mentioned model of child language development

<table>
<thead>
<tr>
<th>Connectionist Structure</th>
<th>Typical Input</th>
<th>Learning Enables</th>
<th>Typical Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept Memory</td>
<td>Semantic feature vectors</td>
<td>To store the concepts available to a child</td>
<td>Child's concepts and their categorisation</td>
</tr>
<tr>
<td>Word Lexicon</td>
<td>Phonemic representations of words</td>
<td>To store the words available to a child</td>
<td>Child's vocabulary, words and their categorisation</td>
</tr>
<tr>
<td>Naming Connection Network</td>
<td>Semantic feature vectors &amp; Phonemic representations</td>
<td>To associate concepts with words.</td>
<td>Child's ability to name concepts.</td>
</tr>
<tr>
<td>Conceptual Relation Network</td>
<td>Visual and auditory stimuli</td>
<td>To produce one-word utterances</td>
<td>Learnt conceptual relations</td>
</tr>
<tr>
<td>Semantic Relation Network</td>
<td>Concepts identified in terms of their conceptual categories</td>
<td>To learn and represent the semantic relations among conceptual categories</td>
<td>Learnt semantic relations</td>
</tr>
<tr>
<td>Word-order Testing Network</td>
<td>Auditory stimuli (Two-word adult collocations)</td>
<td>To learn the word-order of adult sentences, and use it to produce two-word sentences</td>
<td>Word-ordering rules</td>
</tr>
</tbody>
</table>

It appears that the simulation of how language is learnt involves a number of inputs, a variety of processes, and different outputs. Taking a neurolinguistic view of the above model of language development, one can argue that the inputs are received and 'processed' in the different areas of the brain and there is some consensus that, say visual stimuli involve excitation of certain kinds of specific neurons and the reception of a response to auditory stimuli is managed by yet another set of neurons. One can, perhaps, extend these differences to imply that the 'representation' of these stimuli may need different schemata. For instance, the articulation of a response to a stimulus, like what is in this picture? may involve the name of an object or other related word or a mere gesture. From a neurolinguistic perspective the variety of stimuli - visual, auditory or conceptual, are received and processed by different areas of the brain. For instance, visual perception is processed in the 'concept memory', whereas auditory input is processed in terms of simpler sounds - phonemes, such that the phonemes are stored in specific areas of the brain which can be regarded as the 'word lexicon'. Similarly, there are areas of the brain dedicated to lexical processing, for instance a 'naming connection network'.
Table 11 is an outline of a specification for building an information processing system, a system that can simulate child language development. Like any other information processing system, this system should have clearly delineated inputs, processes and outputs.

<table>
<thead>
<tr>
<th>Input</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Perceptual input comprising semantic features describing concepts</td>
</tr>
<tr>
<td>2. Two-word collocations corresponding to adult’s speech.</td>
</tr>
<tr>
<td>Processing (Development)</td>
</tr>
<tr>
<td>1. Concept development and storage in memory</td>
</tr>
<tr>
<td>2. Word development and storage in memory.</td>
</tr>
<tr>
<td>3. Categorisation of concepts and words.</td>
</tr>
<tr>
<td>4. Ostensive naming - Learning associations between concepts and words.</td>
</tr>
<tr>
<td>4. Learning a set of conceptual relations and expressing them as one-word utterances by associating them with corresponding function words.</td>
</tr>
<tr>
<td>6. Enhancing the knowledge of conceptual relations which leads to the development of semantic relations among conceptual categories. This marks the transition from one-word utterances to two-word sentences.</td>
</tr>
<tr>
<td>7. Learning the inherent word-order in adult language</td>
</tr>
<tr>
<td>Output</td>
</tr>
<tr>
<td>1. Production of one-word utterances.</td>
</tr>
<tr>
<td>2. Production of two-word sentences.</td>
</tr>
</tbody>
</table>

Table 11: An information processing view of child language development, listing the types of the input, development (processing) and output achieved during language development for a connectionist model of child language development.

We therefore argue that in order to simulate child language development in a realistic manner one has to take into account both psycholinguistic and neurolinguistic aspects. These aspects, in turn, require the implementation of diverse connectionist networks that can account for the variety of inputs, processes and outputs involved in language development.

4.4. Specification of Connectionist Modules Constituting ACCLAIM

The design specification of ACCLAIM - a hybrid connectionist model involves (a) the selection of plausible connectionist networks and learning algorithms for various 'sub-tasks', (b) the topology of each connectionist network, i.e., a specification of the constituent layers and the connectivity pattern between units or layers, and (c) how the individual connectionist modules are synthesised.
For the selection of plausible connectionist networks, we argue that the choice of connectionist networks should take into consideration whether it is possible to create in the chosen connectionist network an 'abstraction' that at least reflects some properties of the cognitive and linguistic representations, behaviours and relations involved in child language development. For a complete picture of neurological and psychological plausibility, it remains of interest to speculate which connectionist network is best suited to simulate which cognitive activity. We believe that speculations on the plausibility of connectionist networks have future implications for formulating a connectionist workbench.

Agreement on the principles of how to synthesise the constituent connectionist modules is far from universal. In fact, this reflects the incredible variety of theoretical perspectives on language development. Through this diversity of opinions, guided by the psycholinguistic framework discussed in chapter two and the model of child language development discussed in Section 4.3, we propose a possible synthesis of the constituent connectionist structures (Figure 10) which can be envisaged as the 'hybrid' connectionist architecture of ACCLAIM.

We discuss below the specification of the independent connectionist structures constituting ACCLAIM.
4.4.1. Concept Memory and Word Lexicon

In connectionist parlance, children's concept memory, or the 'semantic store', where researchers claim that the acquired conceptual knowledge is 'stored', can be characterised by (a) the concept representation scheme, (b) the organisation of stored concepts, (c) the means for learning new concepts and (d) the mechanisms to retrieve stored concepts. Nelson (1973) argues that children's concepts, comprising objects, people, places or events that children comes in contact with, can be represented and categorised in terms of a number of 'semantic features'. The learning of new concepts, to a certain extent, can be regarded as an unsupervised process, whereby children appear to detect the salient 'semantic features' of an concept without any guidance. Storage of concepts is based on categorising them on the basis of perceived semantic features. On the basis of these assumptions, it appears that for the concept memory we have the following specifications:

| Input: An n-dimensional semantic feature vector, where n is the number of semantic features representing a concept. |
| Processes: Unsupervised learning, feature detection, addition of concepts without disturbing existing organisation, incorporation of Piaget's notions of assimilation and accommodation, the 'automatic' categorisation of concepts without a specification of explicit category labels. |
| Output: Storage of input semantic feature vectors, retrieval of stored concepts, representing each concept in a localist manner, i.e., associating it with one unit, the ability to generalise for novel concepts, reconstruct the semantic feature representation for incomplete patterns, exhibit family resemblances across conceptual categories. |

For the word lexicon, or the 'phonological store', children have been suggested to possess sophisticated phonological skills at the onset of language development, that enables them to analyse the constituent phonemes in the speech around them. Child theorists have argued that 'young children may well be capable of perceiving fine phonetic detail, but they may represent this detail as a set of only loosely organised phonetic features' (Charles-Luce, 1990:206). It seems appropriate then the words leant by the child and stored in the word-lexicon should be represented in terms of their constituent phonemic features. For simulation purposes, we have therefore represented each word as a 'phonetic feature vector'. Representation of words in terms of phonemes also provide the basis for categorising similar sounding words, such that words having similar phonemic representations are placed in similar phoneme specific categories. Therefore, whatever connectionist network is used to simulate the storage and
retrieval of words must be able to handle a highly structured set of phonological representations. The
specification for the 'phonological store' - word lexicon is as follows:

| **Input:** An n-dimensional phonemic feature vector, where n is the number of phonemes constituting the word. |
| **Processes:** Unsupervised learning, phoneme (feature) detection, addition of words without disturbing existing organisation, 'automatic' categorisation of words, incorporation of Piaget's notions of assimilation and accommodation. |
| **Output:** Storage of a number of different phonemic feature vectors, retrieval of stored words; representing each word in a localist manner, i.e., associating it with one unit, and the ability to generalise for new words. |

An examination of the characteristics of various connectionist networks (Table 9) reveals that the class of unsupervised learning connectionist network - Kohonen maps satisfies the above constraints and therefore renders a degree of psychological plausibility to the implementation of the concept memory and the word lexicon. The implementation of the word lexicon as Kohonen maps has also been proposed by Hulme et al (1991), who argues that instead of hand crafting a phonological store it is best to use the self-organising capabilities of Kohonen maps to store words. In ACCLAIM the connectionist concept memory and the word lexicon are therefore implemented as two independent Kohonen maps, that have the following parameters given in Table 12.

<table>
<thead>
<tr>
<th><strong>Input layer</strong></th>
<th><strong>Concept Memory</strong></th>
<th><strong>Word Lexicon</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>20 units</td>
<td>Accounting for a 20-dimensional semantic feature vector representing a concept</td>
<td>5 units</td>
</tr>
<tr>
<td>121 units</td>
<td>The concept memory has the capacity to store 121 concepts, however about 42 concepts would be stored.</td>
<td>121 units</td>
</tr>
<tr>
<td>11 rows &amp; 11 columns (121, 20)</td>
<td></td>
<td>11 rows &amp; 11 columns (121, 5)</td>
</tr>
</tbody>
</table>

Table 12: The parameters of the Kohonen maps simulating the child's concept memory and word lexicon

The psychological/neurological plausibility of Kohonen maps extends to exhibit the following features that are relevant in memory oriented tasks: (a) During information retrieval several items are activated at the same time, thus making visible all items related to the input pattern; (b) prototypical effects can be realised; (c) since information is stored in a localist manner new information does not disturb the existing

---

19 The neurological constraint for the concept memory and word lexicon is mainly the localisation of concepts/words through the formation of the kind of topographic maps suggested to be present in the human brain.
structure of the Kohonen map; (d) due to the categorisation of the input patterns, an incomplete or novel patterns can be recognised in terms of the learnt patterns; (e) family resemblances amongst concepts can be determined.

The backpropagation network can also be used to simulate a concept memory and a word lexicon. However, there are two problems related to the use of backpropagation networks. First, the network emphasises supervised learning - one may just about be justified in using backpropagation networks for word lexicon in that the caretakers do regulate learning of words to a certain extent, but concept memory is another story. Second, learning in BP networks modifies the pattern of inter-connectivity among units in such a manner that the final pattern (or weight vector) is not isomorphic with that resulting from any one experience. Although the pattern of inter-connectivity is the internal representation of the stimulus, no particular part of the network can be identified with any particular feature of the stimulus. This is because processing of each feature of the stimulus is distributed across the entire network of units. Thus the BP network eventually constructs a representation of the stimulus in which knowledge of the various features of the stimulus is distributed across a number of units.

Other connectionist networks such as Grossberg networks and competitive networks though provide an unsupervised learning environment, but fail to automatically categorise stored concepts or words into categories.

4.4.2. Naming Connection Network
Kohonen maps, apart from their psychological appropriateness for simulating the development, storage and retrieval of concepts and words, extend their utility by providing the functionality to associate their output units with units in any other connectionist network. This is a useful property for two reasons: first, an interface between a Kohonen map and another connectionist network enables an information exchange mechanism between the two connectionist networks. Second, this property establishes an n-dimensional mapping between a Kohonen map and another connectionist network. This mapping can be both, many-to-many, that is, all output units of the Kohonen map are associated with all units of the other
connectionist network, or many-to-one, that is, all output units of the Kohonen map are associated with just one unit of the other connectionist network, and vice versa. We intend to exploit these characteristics of Kohonen maps in the implementation of the naming connection network.

In child language literature (Nelson et al., 1978; Callahan, 1985) naming of concepts is regarded as the mapping of children's linguistic knowledge on to their conceptual knowledge. In connectionist terms, concept naming can be regarded as the 'unsupervised' development of an association between a lexical label (word) with its corresponding concept. Note that both the concept memory and word lexicon are implemented as two Kohonen maps, thus facilitating the mapping of concepts stored in the concept memory onto corresponding words stored in the word lexicon, and vice versa. Within our model of child language development the naming connection network implements such a mapping which is achieved by learning associative 'naming connections' between concepts and corresponding words. It therefore appears that the connectionist network most suited for implementing the naming connection network should (a) incorporate unsupervised learning and (b) should be able to simulate the associative bi-directional connections between two units.

The associative characteristics of Hebbian connections incorporating the Hebbian learning rule (cf. Section 3.7.1) appears relevant for the simulation of naming connection network, such that each unit in the concept memory is connected to all units in the word lexicon and vice versa, thereby establishing a many-many relationship between concept units and word units. This means that from each concept unit 121 connections would emanate connecting it to all the word units. The word unit connected with the Hebbian connection having the highest weight would correspond to the lexical label of the concept. The Hebbian learning algorithm ensures that the strongest connection is established among the most highly active units in both the maps, and slightly less strong connections are made between their neighbouring units. Such Hebbian connections learnt between concept and word units could be envisaged as the 'naming connections'. Figure 11 shows how the Hebbian connections can be established between the units in the concept memory and word lexicon.
Figure 11: Showing how Hebbian connections can be established to associate two connectionist structures. Each unit in the concept memory is connected by a Hebbian connection with all units in the word lexicon.

Since the naming connection network is to be implemented as an interface between the concept memory and the word lexicon, its input layer replicates the output layer of the concept memory, and its output layer replicates the output layer of the word lexicon, the resultant parameters of the naming connection network are:

- Input Layer: 121 units
- Output Layer: 121 units
- Total Hebbian connections: 14641 (121 * 121)
- Connectivity Matrix: [121, 121]

Once the naming connections have been learnt, activations from concept/word units can be transmitted to the word lexicon/concept memory by employing the spreading activation. Since the activation spread is constrained by the strength of the connection between two units, units having a strong connection with a highly active unit would acquire a high activation level. For instance, if activations are spread from the concept memory, the word unit which has a strong connection with the most active concept unit would acquire the highest activation level, and in connectionist terms the lexical label (word) of the concept would be deemed to be retrieved.

4.4.3. Conceptual Relation Network

As we understand from Bloom's accounts of children's language (cf. Section 2.5), the child is suggested to have conceptual relations about objects, people and events. These conceptual relations include, disappearance, existence, rejection, request and others of the kind (cf. Table 6). Children's one-word utterances comprise these conceptual relations expressed mainly through 'function words' and to a lesser extent through 'substantive words'. Put simply, it appears that children's one-word utterances are a
manifestation of a kind of *associative* learning; the child learns to associate or map conceptual relations on to corresponding words. The connectionist learning of such a mapping ensures that whenever a conceptual relation is intended to be expressed the corresponding word is selected and produced by the learnt connectionist network. The overall working of the connectionist network, initiating with the perception of an entity, identification of an intended conceptual relation, and finally the production of an appropriate word, is intended to mimic the child's production of one-word utterances. The conceptual relation network then needs to incorporate as input the conceptual relations suggested by Bloom (1973) and listed in Table 6, learning then involves the mapping of these conceptual relations to corresponding words uttered by children also noted in Table 6.

We believe that by way of uttering one-word sentences the child is in fact invoking a dialogue, though limited in content, with fellow human beings. More so, the child's communicative action then provides opportunities for adults to check and correct the child if his or her response to a certain situation is inappropriate. This then implies that there is a notion of supervision or corrective feedback from adults. It therefore appears that the environment (in connectionist terms) for learning conceptual relations can be characterised as supervised, where the child's responses are subject to introspection and are accordingly corrected so that the response may be more appropriate in the future.

Our choice of a connectionist network to simulate the learning of conceptual relations is guided by the above constraints. Accordingly, we use a connectionist network that belongs to the class of supervised connectionist network - Backpropagation networks (BP), that employ the so-called error-correcting backpropagation algorithm (cf. Section 3.7.2) which loosely speaking allows the supervised learning of arbitrary, non-linear relationships or mapping between input and output patterns. BP networks have been extensively used by researchers to simulate a number of cognitive and perceptive skills that involve a mapping of input patterns to output patterns: the learning of past tense by children (Rumelhart and McClelland 1986; Plunkett and Marchman 1989), the association of written language to spoken sounds (Sejnowski and Rosenberg 1986) and many other skills.
The psychological significance of BP networks have been extensively discussed in Appendix A, and amongst its other attributes we know that BP networks have a natural propensity to generalise in novel situations. Whilst generalising the BP network produces the output corresponding to a learnt input pattern that is closest to the novel input pattern. We argue that a realistic simulation of children's production of one-word utterances should be able to handle novel situations as in reality the child frequently encounters novel situations, nonetheless he or she seems to generalise rather well and is able to produce an appropriate response that is based on past experiences. This generalisation ability of BP networks further qualifies them as appropriate to simulate the learning of conceptual relations which are then to be used to produce one-word utterances.

The input and output layers of the conceptual relation network represent the conceptual relations and children's one-word utterances ('functional' words), respectively. We have adopted a 'localist' representation scheme to represent this information, such that one unit uniquely represents a conceptual relation, perceivable entity or word. We give below the organisation of the input and output layers of the BP network implementing the conceptual relation network.

**Organisation of the input layer:** The input layer of the backpropagation network, comprising 25 units, implements, in a localist manner, both the possible conceptual relations and the perceivable entities\(^{20}\). Since, two different kinds of information is represented by the input layer it is best to envisage the input layer as implementing two assemblies of units. The first assembly (say the conceptual relation assembly) comprises 22 units, one unit each for the 22 conceptual relations suggested by Bloom (cf. Table 6). The second assembly (say the perceivable entity assembly) comprises 3 units to represent the three perceptual entities - objects, persons and events. Figure 12 shows the organisation of the conceptual relations and perceptual entities in the input layer.

\(^{20}\) Note that the entities about which the child talks about are aspects of his/her environment which can be broadly categorised as objects, persons and events.
During the learning session various input training patterns, each comprising a conceptual relation component and a perceptual entity component, are presented to the input layer: the conceptual relation unit (in the conceptual relation assembly) and the perceptual entity unit (in the perceptual entity assembly) that correspond to the components of the input pattern are provided with a high activation level of value 1.0, whilst other units remain at a low activation level of value 0.

**Organisation of the output layer:** The output layer of the backpropagation network, comprising 18 units, encodes (a) the child's one-word utterances (words) in response to a conceptual relations and (b) the perceptual entity towards which the response is directed. Again, the output layer is divided into two assemblies of units: the one-word assembly, comprising 15 units corresponding to the prominent words observed by Bloom in children's early language (cf. Table 6); and the perceptual entity assembly similar to that implemented in the input layer. Figure 13 shows the organisation of the output layer.
The output units reflect the response of the connectionist network to the input pattern presented at the input layer. This is determined by noting the activation level of all units in the output layer: the output unit with the highest activation level in the one-word utterance assembly corresponds to the one-word utterance produced by the child in response to the input pattern and similarity the most active unit in the perceptual entity assembly represents the perceptual entity.

In the connectionist literature there is no criterion for selecting the appropriate number of hidden units for a mapping problem. One is therefore required to perform a number of experiments with varying hidden units to determine a network configuration suited to the problem. We therefore performed test learning sessions to determine the optimum number of hidden units that would realise the best mapping of the input patterns onto the output patterns and would also best generalise in novel situations. In each experiment we choose a BP network with a different hidden layer specification and subjected it to learn the required input-output patterns. The following criterion were examined in each experiment in order to determine the optimum number of hidden units:

**Quality of learning:** In each experiment the quality of learning was determined by presenting to the learnt network a number of conceptual relation and perceptual entity pairs and noting which function word was retrieved. Even though the word retrieved may be the correct word, yet to ensure a better quality of
learning we also took into account the error between the output produced and the desired output. If the error was higher than a certain acceptable level it meant that the quality of learning is not satisfactory and hence the number of hidden units need to be incremented.

**Speed of Learning:** The speed of learning was measured in terms of the number of iterations required by the network to learn the set of input-output patterns, and was desired to be reasonably fast without exerting too much processing demands on the simulating machine. We found that the lesser the number of hidden units the greater the number of iterations, and for that matter we increased or decreased the number of hidden units to suit our requirements.

**Generalisation Effects:** An important aspect of the learning task was to acquire the ability to generalise. The number of hidden units has a direct correlation with the generalisation effects offered by a learnt BP network. After learning, the generalisation capabilities of the network were tested by presenting it with novel input patterns and noting how well it was able to generalise. It should be noted that minimising the number of hidden units in a BP connectionist network encourages the network to search for regularities in the input stimuli which in turn leads to good generalisation effects, whereas with an increased number of hidden units the generalisation potential of the connectionist network is rather deteriorated. In fact, there was a very narrow band in between where the network exhibited the best generalisation effects.

One can now imagine that choosing the optimum number of hidden units requires to strike a balance between the above aspects, it seems like an optimisation problem where all the above constraints should be satisfied. After testing several networks implementing hidden units ranging from 3 to 10, we found that the BP network with 5 hidden units produced the best learning quality, reasonably fast learning speed and acceptable generalisation effects.

The parameters of the conceptual relation network are as follows:

- **Input layer:** 25 units
- **Hidden layer:** 5 units
- **Output layer:** 18 units
4.4.4. Semantic Relation Network

If one adheres to the views of child psychologists such as Bloom (1973) and Brown (1973) then development of 'semantic relations' appears to initiate with the presence of two concepts and the subsequent association, or more precisely 'semantic relations' between the categories of these concepts. Establishing semantic relations among concept categories appears to be a 'competitive process', where a set of candidate concept categories compete and ultimately the two winning concept categories are associated, such that all concepts of one category are associated with all concepts in the other category. Creation of such semantic relations between two concept categories is achieved in an unsupervised learning manner, based on the information received from the environment, particularly the adult language. Figure 14 gives the semantic relations among twelve conceptual categories suggested by Brown (1973).

![Figure 14: Brown's semantic relations that are to be learnt by the semantic relation network](image)

Learning of the semantic relations shown in Figure 14 seems to require a 'matrix' memory that can store association between two concept categories. Additionally, the strength of associations between concept categories need to increase with experience, that is, with the repeated presentation of the same stimulus, a stronger association is established between the two 'semantically related' concept categories.

---

21 It is argued that the child in order to avoid cognitive overload, economises the number of semantic relations, such that semantic relations are established among concept categories, rather than individual concepts. In fact, this arrangement facilitates the information processing capabilities of the child, as the child is now able to generalise a novel concept to members of the closest concept category.

22 It has been suggested that, with increasing age, or what some may call 'developing cognition', the child makes a move from the generic to the specific, that is, he or she subdivides the existing concepts in a category to form new and more specific categories. For instance, at an early age the child may group all concepts regarding animate beings, such as people, parents and animals, however at a later stage these concepts would be divided into possibly three independent categories - people, parents and animals. At this point, the existing semantic relations need to be modified in a dynamic manner, or maybe new semantic relations need to be learnt.
The choice of a connectionist network to simulate the learning of semantic relations in a psychologically and neurologically plausible manner, can be guided by the characteristics of the connectionist networks given in Table 9. It appears that, the Additive Grossberg Networks (AGN) provide a matrix memory organisation and a learning environment that is both unsupervised and competitive, making the network suitable for simulating the learning of semantic relations. In connectionist terms, the so-called semantic relations can be simulated by learning a 'connection' between two concept categories, where the concept categories are represented in a localist manner by individual units. To provide a more psychologically explicit simulation we have introduced a modification to the structure of the basic AGN (cf. Appendix A), such that the single layer architecture of the AGN is extended so as to allow it to behave as a multi-layer network, nonetheless inter-layer connections are still modified by the usual AGN learning rules. We give below our reasons for this modification.

A structural alteration in the basic structure of AGN was necessary to overcome a 'category identification' problem, encountered in two-word sentence production. The category identification problem can be understood by considering the following situation: if we have two semantic relations such as *agent* + *object* and *agent* + *action*, than the agent unit would have a semantic relation, that is, a connection with both the *object* unit and the *action* unit with equal weights. Now, if we provide an input to the *agent* unit, the spread of activation from the *agent* unit to other units in the learnt semantic relation network would result in the equal activation of both the *object* and *action* units. However, in order to determine the identity of the probable second concept category, we would anticipate one unit to have higher activation level as compared to other units. This indeterminacy of the second concept category is regarded by us as the 'category identification problem'. To resolve this category identification problem, we have implemented another AGN, which takes as input some perceptual stimuli and on the basis of this perceptual stimuli, determines the second concept category. For instance, continuing with the above example, if the perceptual stimuli corresponds to an *object*, than the perceptual stimuli would enable the *object unit* in the output layer to acquire an activation level higher than the *action unit*, thereby resolving the category identification problem. In this way, the semantic relation network is able to identify the
semantic relation as *agent + object*, and this is achieved without any external aid, rather by incorporating the perceptual stimuli to the modified AGN.

To counter the 'category identification problem' the semantic relation network is implemented as a three layered AGN (shown in Figure 15). The first layer is the input layer and accepts the first concept category, which for simulation purposes determines the 'intention' of the two-word sentence. The second layer is termed as the 'intermediate layer' and is divided into two assemblies: the first assembly replicates the units in the input layer, and during processing activates all units that are associated with the concept category given as input; the second assembly in the intermediate layer represents the perceptual input received by the semantic relation network to resolve the category identification problem. The third layer is the output layer, replicating the input layer it determines the second concept category, and this is made possible by spreading the activation from the intermediate layer. Processing is carried out in the modified AGN by spreading activations of the units, first from the input layer to the intermediate layer and next from the intermediate layer to the output layer. The most active unit in the input and output layers represent the semantic relation between two concept categories.

It may be noted that by adopting a 'layered approach', and implementing an 'intermediate layer' we actually reduced the structural complexity of an otherwise complicated connectionist network. We stress this because, in a typical connectionist network at least two layers are needed, one input and the other output. For two connectionist networks, we would have needed four layers, which would have meant...
more processing units, however our design eliminates this constraint by implementing an intermediate layer which serves as the output layer for the semantic relation network's input layer, and subsequently concatenates with the perceptual stimuli units to act as the input layer for the semantic relation network's output layer. More attractively, our modified AGN still retains the learning mechanisms for the typical AGN.

We consider the perceptual stimuli to represent the following: object, agent, action, attribute and location, inferred from Bloom's (1973) data. We understand that the perceptual stimuli received by children may contain information extending beyond these five categories, as child researchers would suggest, however for our simulation purpose we restrict the perceptual stimuli to these five categories. Since ACCLAIM has the potential to conveniently incorporate changes in its architecture it is possible to add another category, whenever the need arises.

The parameters of the semantic relation network are as follows:

- **Input layer**: 12 units (based on Brown's suggested twelve concept categories)
- **Intermediate Layer**: 12 (concept categories) + 5 (perceptual stimuli) units
- **Output Layer**: 12 units

### 4.4.5. Word-order Testing Network

At the two-word stage the prevalent task underlying word-order learning and testing is twofold: first to observe and learn the inherent word-order in adult sentences and, secondly, to arrange two concepts, expressed as two words, according to the learnt word-order to form a child-like two-word sentence. From a connectionist perspective, the learning of word-order can be simulated as a *pattern association* problem, i.e. associating the input pattern, comprising a two-word collocation, to itself. Based on the learnt word-order, word-order testing can be viewed as the comparison between the desired and actual output produced by a connectionist network.

BP networks have been recommended as pattern associators by Rumelhart and McClelland. Also, the notions of evaluating the BP network's response with the desired response are regarded as a psychological feature of BP learning (cf. Appendix A). In a BP network by evaluating the error produced
by the output layer of a BP network, one can provide a quantitative measure of the correctness of the hypothesis for a given word-order. For instance, if the word-order of an input pattern is previously learnt, the back-propagation network produces a low error value, whereas if the word-order of the input pattern is novel to the connectionist network it produces a high error value.

We have therefore chosen the BP network to implement the word-order testing network, as it is well suited for pattern association, hypothesis testing and for exhibiting good generalisation effects. Other connectionist networks do provide mechanisms for pattern association, however their generalisation and hypothesis testing capabilities are not as profound as that of BP networks. The parameters of the BP network implementing the word-order testing network are as follows:

| Input Layer: 12 units | Hidden Layer: 4 units | Output Layer: 12 units |

The structure of the word-order testing network (see Figure 16) is as follows: since the two words in the sentence are treated as members of a conceptual category, we have divided both the input and output layers into two assemblies of units, termed as conceptual category 1 and conceptual category 2. During learning at the input layer a two-word collocation represented as two conceptual categories is given. The ordering of these conceptual categories is similar to that of adult word-order. Since, the network is required to learn the input pattern, that is, associate the pattern to itself, the input pattern is again given at the output layer. Learning takes place over a period of iterations during which several two-word collocations are given to the network.

To test the word-order hypothesis for any two words during production, we have introduced a slight variation in the basic architecture of the traditional BP network (shown in Figure 16). We have extended the output layer to two different assemblies of output units, each representing the two-words (word 1 and word 2) in both the possible orders, that is, word 1 followed by word 2 or else word 2 followed by word 1. We present a two-word collocation at the input layer, and determine the possible word-order by observing the error generated at both these assemblies of output units. The output assembly having the least error value is taken to represent the correct word-order of the two words constituting the sentence.
4.5. The Architecture of ACCLAIM: A Synthesis of Connectionist Modules

Based on earlier proposals advocating the ‘modularity’ approach for modelling language development we now present the hybrid architecture of ACCLAIM. Earlier we identified a number of sub-tasks or processes that were deemed central to the development of language and in later sections we determined psychologically and neurologically plausible connectionist networks for the implementation of those processes. To recapitulate, Table 13 shows the tenets of a psycholinguistic model of language development, name of the connectionist structure implementing each psycholinguistic aspect of child language development and the connectionist network used.

<table>
<thead>
<tr>
<th>Connectionist structure</th>
<th>Psycholinguistic aspects</th>
<th>Connectionist network</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept memory</td>
<td>Learning and storage of concepts</td>
<td>Kohonen map</td>
</tr>
<tr>
<td>Word lexicon</td>
<td>Learning and storage of words</td>
<td>Kohonen map</td>
</tr>
<tr>
<td>Naming connection network</td>
<td>Concept naming</td>
<td>Hebbian connections</td>
</tr>
<tr>
<td>Conceptual relation network</td>
<td>Acquisition of conceptual relations</td>
<td>Backpropagation network</td>
</tr>
<tr>
<td>Semantic relation network</td>
<td>Acquisition of semantic relations</td>
<td>Additive Grossberg network</td>
</tr>
<tr>
<td>Word-order testing network</td>
<td>Learning syntax - word-order learning</td>
<td>Backpropagation</td>
</tr>
</tbody>
</table>

Table 13: An illustration of the tenets of a psycholinguistic model of child language development, showing the connectionist structures implementing each aspect of language development and the connectionist network used to implement them.

Each of ACCLAIM’s connectionist structures can be envisaged as an individual entity, embodying a different kind of knowledge. These connectionist structures can then be configured based on our psycholinguistic model to realise a variety of connectionist modules, where each module simulates an aspect of child language development. For instance, the concept naming module, simulating concept naming, comprises three connectionist structures - the concept memory, word lexicon and naming
connection network. It should be noted that within a connectionist module the connectionist networks retain their identity and interact with each other, in fact the complete processing within a module is constrained by the information stored in its constituent connectionist structures. In ACCLAIM, four different connectionist modules (shown in Figure 17) relevant to child language development are implemented by integrating various connectionist structures.

One of the advantages of the modular design of ACCLAIM is that knowledge learnt by a connectionist structure is utilised by more than one module, for instance the concepts stored in the concept memory are used by three modules - the one-word module, concept naming module and the semantic relation module. At a deeper level, each module again can be envisaged as an independent connectionist model, capable of simulating a psycholinguistic process on its own. For instance, a simulation of the child’s development of semantic relations can be performed by just employing the semantic relation module. Now that the major psycholinguistic processes involved in language development are identified by child theorists, ACCLAIM then provides the simulation framework whereby individual processes can be simulated in isolation from other co-occurring processes. It is possible that among all processes involved in child language development one may like to concentrate on a certain aspect of child language development,
say for instance children's one-word utterances. The modular approach of ACCLAIM makes it possible to work with one module at a time; simulate the process with a variety of data without taking into account other connectionist modules. More attractively, at a later stage the results of the simulation incorporating just one module can be used in simulations involving other modules.

The hybrid architecture of ACCLAIM originates from a configuration of the above mentioned connectionist modules (shown in Figure 18). A systematic interaction amongst all the modules not only simulates aspects of the development of language both at the one-word and two-word stage, but also enables to produce child-like one-word utterances and two-word sentences.

In earlier chapters we presented traditional AI models of language development and also connectionist models of linguistic behaviour. It was noted that these earlier models were rather limited in the sense that they lacked (a) the manipulation of multiple simultaneous constraints among various language aspects; (b) potential to implement a variety of learning mechanisms; (c) ability to handle more than one type of input; (d) a specification of the environment which is instrumental to learning; (e) the notion of development in terms of time and (f) indeed a large-scale distributed processing environment. We argue
that the hybrid architecture of ACCLAIM addresses the constraints that seemed lacking in earlier simulations of language behaviour.

Our approach of designing a modular, hybrid architecture for simulating language development is in some agreement with child language researchers, that there is an interaction at a neurological level when language is processed and when language is learnt. So, our plea of distributed processing using a number of communicating connectionist networks is based in part on neurological evidence and in part on our intuition of how language is processed and learnt. Therefore, to begin with a perceptual input and two-word collocations in adult speech and ultimately be able to produce two-word sentences one needs a number of connectionist networks that communicate with others, either all of them or a selected few. This communication then introduces the notion of multiple simultaneous constraints, in that, the networks would process the output of other networks and their performance would be constrained by the inputting network and the output receiving network.

It may be noted that the notion of distributed processing can be envisaged at two distinct levels: the 'network level' where knowledge is distributed across separate connectionist networks, and the 'unit level' where within each connectionist network knowledge is distributed across a number of units. Thus, our scheme of distributed processing both at the network and unit level may appeal to both group of researchers, the ones who advocate distributed processing at the unit level or the 'neuron' level, and the other group of workers who argue for a 'functional' localisation of the brain, suggesting distributed processing among those regions of the brain.

The notion of development with age is central to any aspect of human development. We observe that over a period of time the child marks a shift from one-word language to two-word language and so on. The developmental nature of ACCLAIM is emphasised by the fact that development can be seen as a function of time (where time is measured in terms of number of iterations); with age or with increasing number of iterations certain behaviours emerge in a connectionist network. For instance, during the learning of concepts the concept memory initially just learns the concepts presented to it. However, as the concept memory 'grows' time (i.e. the increasing number of iterations), it is able to discriminate
amongst the various concepts learnt by it and is able to categorise them. Therefore, the categorisation of concepts is a behaviour which appears later than the usual learning of concepts. The behaviours emerging after a certain time could either be a refinement of existing behaviours or else could be new behaviours emerging as a consequence of increased exposure to the environment and improved learning abilities. We argued earlier that the learning mechanisms implemented by connectionist networks can be viewed as correlates of Piaget's assimilation and accommodation processes; then again we have development which is a function of time, and whatever is learnt builds on earlier knowledge.

In ACCLAIM the activity of production of one-word utterances and two-word sentences can be viewed as 'data-driven'. This is so because there is no central controller monitoring the execution of the individual connectionist networks. A connectionist network receiving some input either from the environment or from another connectionist network, undergoes its assigned processing sequence and its output response is passed to other connected connectionist networks. The connectionist structures receiving input from other connectionist structures initiate their processing as soon 'data' is available to them. This scheme however, assumes a representation scheme interpretable by a variety of connectionist networks. Also, it may be noted that during this hierarchical flow of information in ACCLAIM there is no need for 'data validation', the connectionist structures act on whatever data they receive. In case, the data received by a connectionist network is inappropriate to some extent then the generalisation capability of connectionist networks is exploited to produced meaningful results.

The architecture of ACCLAIM and the resultant processing capabilities achieved should be an indicator as to how functionally and structurally divergent connectionist networks when synthesised together in a meaningful manner, i.e. based on a psycholinguistic model, can simulate a high-level cognitive activity such as child language development and production. Furthermore, these individual connectionist networks are trained separately, using 'language-informed' data.
4.6. Is a Hybrid of Connectionist Networks Really Needed to Simulate Child Language Development?

We argued earlier that in ACCLAIM the choice of using a particular connectionist network for some specific aspect of children's language is mainly guided by the degree of psychological plausibility offered by the chosen connectionist network, and to a lesser extent we look for some biological plausibility, if any. We know that the various integral aspects in the child's language are all learnt or developed in varying conditions, some are learnt without the aid of an adult guidance whilst others are learnt by being instructed by adults. The former type of learning corresponds to unsupervised learning in connectionist literature and the latter can be regarded as supervised learning. This classification of connectionist networks is further demarcated by their behaviour and properties which again emerge as a consequence of learning.

We now ask the question, is a hybrid connectionist architecture, comprising a variety of connectionist networks needed just to achieve psychological plausibility? or are there any behvioural and functional advantages that one might benefit from by using such an architecture?. We believe that the best way to answer the above question is to implement an experimental connectionist model that comprises the same connectionist structures as ACCLAIM, but each structure is implemented by using just one type of connectionist architecture and learning algorithm. We have chosen the widely used backpropagation (BP) network which incorporates a supervised learning algorithm to implement this experimental connectionist model. Next, we carried out individual simulations involving the learning of information specific to each connectionist structure using the same data we intend to use in our actual simulations using ACCLAIM. Table 14 gives the statistics of our experimental connectionist model for simulating child language development, the speed of learning in given in terms of the number of iterations required by the BP network to learn the given information.
We now examine the performance of each connectionist structure based on the BP algorithm, in particular we would be looking for the presence of certain behaviours that are suggested to emerge as a consequence of learning. We will enumerate what can be simulated using a BP network, and also point out what aspects of child language development BP network cannot simulate.

**Concept Memory**

The BP network was able to learn the semantic feature patterns corresponding to various concepts, and was able to retrieve a learnt concept by presenting its semantic feature pattern.

The problems faced with using the BP network for simulating development of concept memory were as follows:

I. The BP network did not automatically categorised the learnt concepts, hence it was not possible to determine the degree of similarity amongst various concepts. In fact, in earlier simulations of concept learning connectionists had to explicitly categorise concepts by associating a category label with each concept during learning. However, this is not the case in the child's development of concepts where the child itself determines the category of each concept learnt.

II. The BP network did not provide any provision to add new concepts to an existing learnt concept memory. Adding a new concept to the learnt concept memory meant learning all over again the

<table>
<thead>
<tr>
<th>ACCLAIM's connectionist structures</th>
<th>Implementation in ACCLAIM</th>
<th>Learning Type</th>
<th>Topology of BP Network</th>
<th>Learning Iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept memory</td>
<td>Kohonen map</td>
<td>Unsupervised</td>
<td>Input layer = 20</td>
<td>8346</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hidden layer = 8</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Output layer = 20</td>
<td></td>
</tr>
<tr>
<td>Word lexicon</td>
<td>Kohonen map</td>
<td>Unsupervised</td>
<td>Input layer = 5</td>
<td>5897</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hidden layer = 3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Output layer = 5</td>
<td></td>
</tr>
<tr>
<td>Naming connection network</td>
<td>Hebbian connection network</td>
<td>Unsupervised</td>
<td>Input layer = 20</td>
<td>10108</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hidden layer = 6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Output layer = 5</td>
<td></td>
</tr>
<tr>
<td>Conceptual relation network</td>
<td>Backpropagation network</td>
<td>Supervised</td>
<td>Input layer = 25</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hidden layer = 5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Output layer = 18</td>
<td></td>
</tr>
<tr>
<td>Semantic relation network1</td>
<td>Additive Grossberg network</td>
<td>Unsupervised</td>
<td>Input layer = 12</td>
<td>3698</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hidden layer = 3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Output layer = 12</td>
<td></td>
</tr>
<tr>
<td>Semantic relation network2</td>
<td>Additive Grossberg network</td>
<td>Unsupervised</td>
<td>Input layer = 4</td>
<td>2446</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hidden layer = 3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Output layer = 12</td>
<td></td>
</tr>
<tr>
<td>Word-order testing network</td>
<td>Backpropagation network</td>
<td>Supervised</td>
<td>Input layer = 12</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hidden layer = 4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Output layer = 12</td>
<td></td>
</tr>
</tbody>
</table>

Table 14: Statistics of our experimental connectionist model implementing ACCLAIM's connectionist structures using just BP networks
existing concepts together with the new concept. Again this is not the case in children's development of concepts.

III. The concepts learnt were represented in a distributed manner, as opposed to a desired localist representation that would have enabled the labelling of individual units to specific concepts.

IV. The learning was based on pattern association, whereas a competitive learning environment where similar concepts competed to be associated to a unit would have been more psychologically plausible.

The above short-coming of BP networks maybe due to the fact that the learning of concepts was carried out in a supervised learning environment as opposed to a teacher free (unsupervised) environment.

**Word Lexicon**

The BP network was able to learn the phonetic contents comprising a word and retrieved a learnt word when presented with a pattern corresponding to its constituent phonemes.

Below we enumerate aspects of the development of word lexicon that could not be simulated by BP networks.

V. Inability to automatically categorise words on the basis of their sounds - absence of similarity neighbourhoods.

VI. In ability to add new words to an existing learnt word lexicon.

VII. Distributed representation of words thus no individual units were assigned to specific words. In case one attempts to use individual units to represent individual words then addition of a new word requires changes in the architecture of the BP network.

VIII. Learning involved pattern association.

Again, the learning mechanism based on supervised learning for the development of a word lexicon may have some bearing on the above short-comings

**Naming connection network**

The BP network was able to learn to associate the semantic feature vector of a concept with its phonetic representation, i.e. word. The naming behaviour simulated was able to retrieve a learnt word when presented with the corresponding concepts' semantic feature pattern.
A few limitations were observed in the simulation which again could be due to the fact that learning was supervised in nature. However, in reality one assumes that there is no teacher present to direct the child to associate an 'abstract' and 'internal' concept to a word. The limitations were as follows:

IX. No possibility to learn new concept-word associations without re-learning existing knowledge.

X. The association learnt were uni-directional, i.e., concepts were associated with words and not vice versa. We believe it would have been more psychologically plausible if the associations between concepts and words were bi-directional. This would have ensured that not only a word could have been retrieved when presented with the corresponding concept but also a concept could be retrieved by presenting the corresponding word.

XI. The retrieval mechanism involved just the retrieval of a word associated with the given concept. The selection of the retrieval word did not involve a competition among a number of candidate words with the most appropriate word winning the competition. Psycholinguistic accounts of lexical recall suggest the presence of a competition amongst candidate words.

Semantic relation network 1 and 2

The BP network was able to associate one conceptual category with another, such that when one conceptual category was presented as input in response the conceptual category associated with it was highly activated. Again, the supervised learning aspect of BP networks diminished the psychological plausibility of the simulation as it implied that the child is being told to associate 'abstract' categories.

Other limitation observed were as follows:

XII. The learnt associations were uni-directional.

XIII. The mechanism to determine the associated conceptual category was not based on competition, thus it was not possible to know the degree to which other categories were associated with the input category.

The conceptual relation network and the word-order testing network are both implemented as BP networks in ACCLAIM therefore here we do not discuss the positive and negative aspects of their implementation.

From the above discussion it becomes evident that the BP network used in each simulation was able to learn and retrieve knowledge related to various aspects of a child's developing language. However, if one aims for psychological plausibility then certainly the above mentioned shortcoming of BP networks restrict their usage as appropriate connectionist networks for simulating the development of the above
mentioned aspects of children's language. If one ignores plausibility and just concentrates on functional aspects of the simulation even then BP networks seems deficient in certain respects, for instance, their inability to conveniently provide a localist representation, and in the simulation of learning of concepts the need for providing category labels during learning to manifest categorisation.

From the above discussion, the answer to the original question that is a hybrid model comprising a variety of connectionist networks is necessary for language development? is certainly yes: a hybrid architecture such as ACCLAIM for simulating child language development is not just an implementational convenience but rather it originates in order to provide both psychological and functional plausibility. By way of a hybrid connectionist architecture we are able to implement the best features offered by various connectionist networks to produce a more powerful and plausible simulation.

4.7. Conclusion
This chapter underlines our contributions to the field of connectionist modelling of child language development. By way of a connectionist model we have been able to operationalise 'static', and even at times disjoint theories and data sets about child language development. Earlier, we identified central aspects of language development from the psycholinguistic literature regarding child language development and attempted to relate them with one another so that assumptions about one aspect of language development can be used when considering some other aspect. By relating these psycholinguistic aspects in a linear manner, i.e., first concepts are learnt, then words, followed by the naming of concepts, we were able to formulate a model of child language development. The efficacy of such a model of child language development was further substantiated by the implementation of the tenets of the child language development model using individual connectionist structures and modules that offered both psychological and neurobiological plausibility. The hybrid connectionist model - ACCLAIM then forms the basis to simulate the development of child language and the production of one-word utterances and two-word sentences. We emphasise here that prior to the development of ACCLAIM, we find no evidence of a psychologically plausible connectionist model in the connectionist literature that incorporates a variety of connectionist networks to jointly simulate various aspects of child
language development leading to the production of child-like language. Indeed, our methodology for designing large scale sophisticated connectionist models would prove useful if one intends to simulate some other aspect of human cognition which involves an interplay of a variety of processes.

As Bechtal and Abrahamsen have argued that 'the goal of psychological modelling is to account for empirical data about behaviour and to provide some understanding of how that behaviour is produced. For connectionist modelling to satisfy this goal, specific models must be developed that can provide a superior account of a variety of data sets. They should also suggest fruitful new approaches to unsolved problems, and point the way to new questions and new experiments to perform' (1991:264). We believe that ACCLAIM certainly can be regarded as a sophisticated connectionist model that provides a plausible account of a variety of psycholinguistic data sets, raising the level of psycholinguistic inquiry into child language development from mere speculations and observations to a connectionist implementation of various issues which can then be simulated in accordance with a variety of constraints, i.e., representation types, learning mechanisms, environmental affects and so on. Even the simulations are time-oriented whereby we demonstrate that the behaviour of a connectionist network improves with time: incorporation of the notion of learning by experience and learning through evolution. We believe that the development of ACCLAIM and the connectionist simulations performed through ACCLAIM, in the next chapter, are both a first step in the addressing old questions about language development and pave the way for new questions to be asked.
Chapter 5

A Connectionist Simulation of Child Language Development:

One-word and Two-word Language

5.1. Introduction
Child language development theories have motivated the collection of substantial amounts of observational data. The data collection exercises involve a number of interesting hypotheses about, the categorisation of objects and events in the 'real world', and conjectures about the possible 'templates' and mechanisms through which a child might represent the 'world' around himself or herself. Language makes possible the expression of the 'world' which may include the child's beliefs, desires and feelings. In order to express such internal states, the language development phenomenon need to account for the development of concepts, lexica, semantics, syntax and discourse. Therefore, it is possible to argue that if one were successful in synthesising psycholinguistic observations, particularly during the various stages of 'language development' (spanning the onset of language, i.e., c. 9-12 months, the vocabulary spurt, and the transition to multiword speech around the 'critical period', i.e. c. 18-24 months) with the developmental neurobiological observations, then one might have a psychologically plausible and neurobiologically tangible description of the child language. We would like to claim that such a synthesis would be possible within a connectionist framework.

The development of a child's language is an subtle interplay of its biological potential, psychological makeup and environmental input. The above is a crude paraphrase of Piaget's notion of assimilation, accommodation and organisation. Our conjecture is that, aspects of child language development that can be construed to be innate development can be simulated using unsupervised learning regimes, whereas environmentally determined aspects of language development can be simulated using supervised learning mechanisms. For instance, on one hand one can argue that the development of concepts is by and large self-motivated and therefore such an important parameter of a child's language can be simulated using
unsupervised learning network. On the other hand there is some evidence that an understanding of word-order can be in part environmentally inspired and, for that matter one might need to use a supervised learning algorithm.

In order to evaluate the efficacy of connectionism, vis-à-vis child language development, we propose the following: first, to simulate the self-motivated learning aspects of child language (e.g. concept development, word acquisition, ostensive naming, and so on) using unsupervised learning algorithms in conjunction with child language data (as provided by child language researchers), and second to evaluate the learnt behaviour of the connectionist network. Similarly, the learning of conceptual relations leading to the production of one word language and the learning of word-order can be deemed to be environmentally sensitive and for that purpose it would be quite a useful strategy to also evaluate the efficacy of supervised learning algorithms for child language development.

In this chapter, we demonstrate how connectionist networks can be used to simulate or more accurately 'trained' to mimic various aspects of children's language in the background of both environmental influences and innate abilities. We intend to use ACCLAIM to perform a series of realistic connectionist simulations of various aspects of child language development within the age group 9 - 24 months.

The connectionist simulations performed using ACCLAIM are language informed such that the data used in 'training' the connectionist networks was derived from the archives compiled by child language researchers. The 'learnt' performance of ACCLAIM would be quantified against Bloom's (1973) data which reports children's utterances with details of the situation in which the utterance was made, also there are references to the cues provided to children by adults. For training the connectionist networks we have used Bloom's characterisation of conceptual relations (for one-word utterances), the conceptual representation schemata incorporates 'semantic features' based on the lines of Nelson's (1973) 'semantic structures', and notions of 'semantic relations' proposed by Brown (1973). We therefore argue that by way of our connectionist simulations we help to operationalise child language data and also provide opportunities to test child language theories.
If one accepts the cognition hypothesis, that is, development of concepts precedes that of language itself, then the development of concepts can be deemed as the starting point for the development of language. The child’s communicative abilities are first demonstrated when he or she starts to produce one-word utterances which can be regarded as a manifestation of their internal concepts, ideas or intentions. The learning of various conceptual relations among entities remains central to the child’s language at the one-word stage which appears to be a supervised affair. Children’s one-word utterances comprise a small, yet growing vocabulary of words which are learnt around the same time. Having acquired a suitable repertoire of concepts and words the child, according to the cognition hypothesis, would then go on to learn to associate concepts with words. This is the equivalent of ostensive naming. So having a semblance of language where there are concepts and words and they are interrelated the child can now acquire neologisms and novel concepts. During the next stage of development the child appears to retrieve appropriate concepts and lexical items when faced with the need to do so, and goes on to generalise concepts. The concept generalisation is quite an abstract process and involves not only the presence of a number of concepts but also some kind of self-motivated development of categories of concepts. The presence of various concept categories leads to an association among them, or more precisely the development of ‘semantic relations’, marking the onset of two-word language. Such self-motivated learning of concepts (and their lexicalisation) can be simulated as a self-organisation process. However, the learning of a word-order, which leads to the production of two-word sentences, requires an interplay with the environment in terms of instruction from adults. Table 15 summarises our child language simulation strategy, showing aspects of child language that are to be simulated both at the one-word and two-word stage of child language development.
### Table 15: List of simulations characterising child language development, together with the connectionist architecture used to simulate them and the type of learning incorporated within the connectionist architecture

<table>
<thead>
<tr>
<th>Simulation Task</th>
<th>Learning Type</th>
<th>Connectionist Architecture</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>One-word language</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concept development, organisation and categorisation</td>
<td>Unsupervised</td>
<td>Concept memory (Kohonen maps)</td>
</tr>
<tr>
<td>Concept generalisation, neologisms and novel concepts</td>
<td>Unsupervised</td>
<td>Concept memory (Kohonen maps)</td>
</tr>
<tr>
<td>Word development, organisation and categorisation</td>
<td>Unsupervised</td>
<td>Word lexicon (Kohonen maps)</td>
</tr>
<tr>
<td>Learning conceptual relations - production of one-word utterances</td>
<td>Supervised</td>
<td>Conceptual relation network (Backpropagation network)</td>
</tr>
<tr>
<td><strong>Two-word language</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concept lexicalisation - development of ‘naming connections’</td>
<td>Unsupervised</td>
<td>Concept lexicalisation network (Hebbian connections)</td>
</tr>
<tr>
<td>Concept and lexical retrieval</td>
<td>Unsupervised</td>
<td>Concept memory + Concept lexicalisation network + Word lexicon (Kohonen maps + Hebbian connections + spreading activation mechanism)</td>
</tr>
<tr>
<td>Learning semantic relations - determining the semantic relation between conceptual categories</td>
<td>Unsupervised</td>
<td>Semantic relation network (additive Grossberg network + spreading activation mechanism)</td>
</tr>
<tr>
<td>Learning word-order - production of two-word sentences</td>
<td>Supervised</td>
<td>Word-order testing network (Backpropagation network)</td>
</tr>
</tbody>
</table>

In the above table one can almost detect a linear order in which child language develops from concepts right on to the word-order\(^{23}\). The child language is demarcated in to two distinct stages - the one word language stage and the two-word language stage.

We present the simulations very much like psychological experiments reported in psycholinguistic literature. Using psychological terminology we illustrate the method of experiment and the results observed. The method of the experiment includes a description of the subject - connectionist networks including Kohonen maps, Hebbian learning connections, backpropagation networks and so on; stimuli - input and output patterns that encode the linguistic, perceptual and conceptual information; and the procedure - the manner and the number of times the learning stimuli is to be presented to the subject, whereas the results describe (a) the learning effects and organisation of information in the connectionist network and (b) the performance of the learnt connectionist network. We would show at the end of this discussion that by and large our conjecture was quite a positive one in the sense that these simulations yielded a close resemblance to the results observed in child language studies.

---

\(^{23}\) The sequence in which the connectionist experiments are reported may have some relevance to the course of language development among children. If one assumes that, development at one stage is constrained by the knowledge acquired during a previous stage than, the order in which we perform our simulations coincides well with this assumption.
5.2. Experiment 1: Development of a Concept Memory

In connectionist terms, children's concept memory, or the 'semantic store' where the acquired conceptual knowledge is 'stored', can be characterised by (a) the concept representation scheme, (b) the organisation of stored concepts, (c) the means for learning new concepts, and (d) the mechanisms for retrieving stored concepts. Recall that we regard the learning of new concepts can perhaps be regarded as an unsupervised process, whereby children appear to detect the salient 'semantic features' of an concept without any guidance. The storage of concepts is effected by categorising them on the basis of perceived semantic features (cf. Section 2.4.1).

In the past, attempts to simulate the development of concepts have been undertaken by connectionists including Hinton (1986), McClelland and Rumelhart (1985). These simulations were conducted in a supervised learning environment; a concept was learnt by repetitively associating its semantic feature based representation with the concept's lexical label. Categorisation was achieved by merely learning the label of the category to which it belonged. It appears that the strategy to learn concepts adopted by earlier connectionists seem more as a case of 'rote learning'.

We have both agreements and disagreements with earlier simulations of concept development. Elaborating on the agreements first:

Agreements

- We agree that concept development is to be simulated in a repetitive manner, such that the concept to be learnt is presented or exposed to the learning connectionist network more than once. This brings into relief the incremental nature of concept development; learning that is a function of time. One can then observe that at an earlier stage of learning the connectionist network was not able to discriminate various concepts, however as it comes across the same concepts more than once the connectionist network is not only able to learn the concept but also to appreciate the subtle differences between various other learnt concepts.

- We agree with the concept representation scheme which is based on the theory of 'semantic features'. We believe that by adopting an established psycholinguistic theory for concept representation we are in fact rendering a degree of psychological plausibility to the simulation.
Disagreements

- Our general disagreement is with a learning strategy which regards concepts to be developed through rote learning. On the contrary, we understand that concept development is a creative activity whereby children form an understanding of their environment through active interaction with the environment and not by just 'memorising' concepts. We disagree with a supervised learning strategy, that relies on an external teacher to provide feedback about the correctness of the learnt concept. Rather, development of concepts should be simulated in an unsupervised manner, based on inference, observation and discovery by the learner. As a general principle of concept development therefore, the assumption that a teacher is involved appears to be rather weak and less psychologically plausible.

- We disagree with the prerequisite of having the concept's lexical label present whilst it is being learnt. This constraint implies that language is a prerequisite for learning concepts, which then contradicts Cromer's 'cognition hypothesis' (cf. Section 2.2.3), i.e., non-linguistic concepts are a precursor of 'language' and that the development of concepts may be independent of their lexical tagging.

- We disagree with the manner in which categorisation of concepts is achieved by explicitly learning the category to which each concept belongs. We believe that categorisation of concepts is rather an indirect affair; categorisation should be seen as a behaviour that emerges as a side-effect during the learning of concepts. Categorisation of concepts results when the connectionist network, is able to discriminate the subtle differences amongst various concepts and is then able to group or categories concepts on the basis of their similarities, and not just by learning category labels.

In our simulation of concept development we improve on earlier simulations and adopt an 'unsupervised' learning approach which is independent of the lexical tagging of concepts, also categorisation emerges as a result of concept learning and not just the mere tagging of category labels.

5.2.1. Experiment 1a: Simulation of the Development of Concept Memory

Method

Network: A Kohonen map implementing the concept memory described in Section 4.4.1. is chosen to simulate the development and the subsequent storage of concepts. The parameters of the concept memory are:

- Input Layer: 20 units
- Competitive (Output) Layer: 121 units
The Kohonen map contains no a priori knowledge. This is ensured by initialising the Kohonen map with random weights in the range 0 - 1.

Stimuli: The stimuli comprises 42 children's concepts selected from the range of concepts reported in child language literature (Bloom, 1973). Each concept is represented by a 20-dimensional semantic feature vector comprising both defining and individual features. A typical semantic feature vector for the concept 'dog' is as follows:

<table>
<thead>
<tr>
<th>Defining Features</th>
<th>Individual Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 1 0 0 0 1</td>
<td>1 0 1 1 0.5 0.0 1.0 0.0 1.0 0.5 0.0 0.0 0.0</td>
</tr>
</tbody>
</table>

Procedure: The simulation of the development of the concept memory is carried out in an iterative manner, such that in each iteration a different concept is presented to the concept memory. Individual concepts are presented more than once in a random order to ensure that the 'learning' taking place is not biased in any way and does not reflect a predefined course of development. The repeated presentation of the concepts over a number of iterations is analogous to the child's increased appreciation and knowledge of the concept over a period of time, and perhaps for the child it is this frequent repetition of information which leads to its assimilation.

At the start of the simulation the Kohonen map contains no a priori knowledge. This can be illustrated by mapping the concepts to be learnt on to the randomly initialised Kohonen map. One may observe that potentially close concepts are mapped sparsely, indicating the absence of any prior categories (see Figure 19).
Learning in a Kohonen map is determined by two parameters - (i) activation level (ACT) of the desired concept's unit when retrieved\(^{24}\), and (ii) the Euclidean distance (ED) between the desired concepts' unit and the most highly active unit. In fact, as learning progresses, ED is minimised by the self-organisation mechanism inherent in Kohonen maps, and at the same time the ACT of the desired concept unit increases. A concept is deemed to be learnt when, upon presentation of its semantic feature vector, the ACT of its representative unit is the highest (approaching unity) amongst all other units and its ED is the lowest (close to zero).

Consider the learning of the concept 'dog', one of the 42 concepts to be learnt by a Kohonen map. The Kohonen map was presented the 20-dimensional feature vector of each of the 42 concepts in a random order, during a learning period spanning 8000 iterations. In order to elaborate on how the Kohonen maps learns, we took a snapshot of the evolving concept memory for the 'dog' concept. This snapshot records

\(^{24}\text{Concept retrieval in a connectionist network involves the presentation of a semantic feature vector - input pattern to the concept memory. This results in all units acquiring some activation level. The unit with the highest activation level best represents the input pattern, hence the concept associated with this unit is considered retrieved. If adequate learning has been achieved, the concept retrieved corresponds to the input pattern, otherwise other similar concepts may be undesirably retrieved.}\)
at every 500 iterations the value of ACT and ED. For each snapshot we presented the Kohonen map the semantic feature vector of the concept 'dog'. This resulted in all the Kohonen map units acquiring some activation level. The unit with the highest activation level was regarded as the concept retrieved. Table 17 gives the learning profile comprising the concept unit retrieved (RU), and the activation level (ACT) and the Euclidean Distance (ED) of the concept unit 'dog'. The changes in the ACT and ED is also graphically illustrated in Figure 20.

<table>
<thead>
<tr>
<th>Iteration Range</th>
<th>Concept 'DOG'</th>
<th>RU</th>
<th>ACT</th>
<th>ED</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - 500</td>
<td>pig</td>
<td></td>
<td>-0.27</td>
<td>0.335</td>
</tr>
<tr>
<td>501 - 1000</td>
<td>dog</td>
<td></td>
<td>-0.29</td>
<td>0.372</td>
</tr>
<tr>
<td>1001 - 1500</td>
<td>duck</td>
<td>0.170</td>
<td>0.366</td>
<td></td>
</tr>
<tr>
<td>1501 - 2000</td>
<td>duck</td>
<td>-0.29</td>
<td>0.096</td>
<td></td>
</tr>
<tr>
<td>2001 - 2500</td>
<td>dog</td>
<td>0.45</td>
<td>0.247</td>
<td></td>
</tr>
<tr>
<td>2501 - 3000</td>
<td>dog</td>
<td>0.966</td>
<td>0.003</td>
<td></td>
</tr>
<tr>
<td>3001 - 3500</td>
<td>dog</td>
<td>0.998</td>
<td>0.001</td>
<td></td>
</tr>
<tr>
<td>3501 - 4000</td>
<td>dog</td>
<td>1.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>4001 - 4500</td>
<td>dog</td>
<td>1.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>4501 - 5000</td>
<td>dog</td>
<td>1.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>5001 - 5500</td>
<td>dog</td>
<td>1.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>5501 - 6000</td>
<td>dog</td>
<td>1.0</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>6001 - 8000</td>
<td>dog</td>
<td>1.0</td>
<td>0.0</td>
<td></td>
</tr>
</tbody>
</table>

Table 17: Learning profile for concept 'dog'

Figure 20: Graphs for ACT and ED (concept 'dog')

Table 17 shows that at the very first iteration, the ED between the (random) weight vector of all the units and the input stimulus is computed. The unit that has the minimal distance to the stimulus is 'assigned' the stimulus label. Subsequent iterations involve the computation of the ED and the reassigning of concepts to the units. After 500 iterations, when the stimulus 'dog' was presented to the concept memory, it retrieved the concept 'pig' - the Kohonen map has not yet learnt to discriminate between a 'dog' and a 'pig' and can easily confuse the two. This 'confused' behaviour of the Kohonen map can be explained as follows: the semantic feature representations of both concepts - 'pig' and 'dog' share a number of features.

The retrieval of the proximate concept 'pig' instead of the concept 'dog' clearly indicates that, at this stage, the Kohonen map has acquired an understanding of a category structure, i.e., the defining features have been learnt. However, the Kohonen map is still not able to discriminate among the individual features of the concepts 'dog' and 'pig' (both concepts belong to the same category) and therefore confuses the stimulus 'dog' with the close concept 'pig'.

127
At the end of 1000 iterations, the stimulus 'dog' retrieves the unit labelled 'dog', but the value of the ED is quite large (0.372) and the activation level is very low, in fact it is negative (-0.29): this retrieval may yet turn out to be a 'fluke'. This is confirmed at the end of 1500 and 2000 iterations; the Kohonen map now confuses the concept 'dog' with 'duck'. But after 2500 iterations, we see in Figure 20 a positive activation and a reduction of the ED in the learning profile for the concept 'dog'. Subsequent iterations do show that the network is becoming more 'stable' in its response to the stimulus 'dog': a doubling of the activation level between 2500 and 4000 iteration and a 200 fold reduction in the ED (see Table 17). At iteration 4000, the criteria for adequate learning have been satisfied, i.e., the activation level has approached unity and the ED has decreased to zero (see Table 17). In Table 18, we give the learning profile for three other concepts - 'Juice', 'Dad' and 'Cow'.

<table>
<thead>
<tr>
<th>Iteration Range</th>
<th>'JUICE' RU</th>
<th>'DAD' RU</th>
<th>'COW' RU</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - 500</td>
<td>--</td>
<td>dad</td>
<td>cow</td>
</tr>
<tr>
<td>501 - 1000</td>
<td>juice</td>
<td>mum</td>
<td>horse</td>
</tr>
<tr>
<td>1001 - 1500</td>
<td>juice</td>
<td>mum</td>
<td>horse</td>
</tr>
<tr>
<td>1501 - 2000</td>
<td>juice</td>
<td>mum</td>
<td>horse</td>
</tr>
<tr>
<td>2001 - 2500</td>
<td>juice</td>
<td>mum</td>
<td>cow</td>
</tr>
<tr>
<td>2501 - 3000</td>
<td>--</td>
<td>dad</td>
<td>cow</td>
</tr>
<tr>
<td>3001 - 3500</td>
<td>--</td>
<td>dad</td>
<td>cow</td>
</tr>
<tr>
<td>3501 - 4000</td>
<td>cookie</td>
<td>dad</td>
<td>cow</td>
</tr>
<tr>
<td>4001 - 4500</td>
<td>juice</td>
<td>dad</td>
<td>cow</td>
</tr>
<tr>
<td>4501 - 5000</td>
<td>juice</td>
<td>dad</td>
<td>cow</td>
</tr>
</tbody>
</table>

Table 18: Learning profile showing the development of concepts - 'juice', 'dad' and 'cow'. The legend for the column is as follows: RU indicates the 'Retrieved Unit', Act is the 'Activation level of the actual concept unit, and ED is the Euclidean Distance.

The learning profile for the other three concepts - 'juice', 'dad' and 'cow' follow a similar trend as noted in the development of the concept 'dog', such that the activation level starting from a low value increases towards unity and an initially high ED is reduced to zero. Note that for the concepts 'dad' and 'cow' during the iteration range 2000-4000 (shaded grey in Table 18) an interesting behaviour is observed. When presented with the semantic feature vector for a concept, say 'dad', two concepts are retrieved: the concept 'dad' and another close concept - 'mum'. This rather atypical behaviour predicates the fact that during this iteration range the Kohonen map is not able to differentiate between close concepts in a category. The retrieval of all the close concepts clearly indicates that at this stage the Kohonen map has learnt a category structure, i.e. defining features, and is exploiting this information when deciding what
concepts are to be retrieved. However, the individual features of concepts are not yet to be learnt and therefore no one concept is more familiar than the other. Note that, after 4000 iterations the discriminating features between the two concepts 'dad' and 'mum' are actually learnt. Interestingly, this indecisive behaviour of the Kohonen map seemingly coincides with children's over-generalisation effects.

Figure 21 shows two graphs depicting the learning profile for concept development, both in terms of activation level and ED.

![Figure 21: Learning profile in terms of activation level and Euclidean distance](image)

**Results and Discussion**

After a learning session spanning 8000 iterations the 42 concepts presented to the concept memory are learnt such that each concept is represented by a unique unit. During learning the concept memory has been organised, starting from a randomly organised Kohonen map shown in Figure 19 the learnt concept memory is highly organised such that similar concepts are stored in proximity. Figure 22 shows the learnt concept memory.
Evidence of an indirect simulation of category learning - 'automatic categorisation': Most connectionist simulations of category learning have been carried out using 'supervised' learning regimes, where a tutor provides category labels and monitors the feedback relevant to the learning criterion. This is indeed not entirely the case with category learning amongst humans. Many categories that humans learn in real life are acquired in observational conditions without feedback from a tutor (Clapper & Bower, 1994).

The organisation of the concept units in the concept memory (shown in Figure 22) reveals that concepts that have close semantic feature representations are stored in proximity, thus forming a global organisation into conceptual regions or, more appropriately, 'categories' of concepts. It may be observed in Figure 22 that, the 'learnt' concept memory is divided into seven broad concept categories suggested by Bloom (1973) - objects, agents, locations, attributes, prepositions, events and function words.

---

25 Similar categorisation effects was reported in psychological experiments by Rip, Shoben and Smith (1973). Categorisation was again based on related semantic features, that is, the more features two concepts had in common, the greater the proximity between the two concepts.
Effectively, self-organisation in Kohonen maps demarcates the possible input space into hierarchical sub-areas which are then mapped on to the two-dimensional Kohonen map. Recall that the semantic feature representation (cf. Section 2.4.1) of each concept is based on a hierarchical structure: 'defining features' (containing category information) and 'individual features' (distinguishing individual concepts). The development of concept memory involved not just the satisfaction of the learning criteria, but additionally detected similarities in the semantic feature representation of the concepts to be learnt. Whilst learning the concepts, the Kohonen map exploited the category information encoded in each concept's representation and collected concepts with similar 'defining features'. These semantically close concepts were then stored in proximity to each other, resulting in clusters of concepts that resemble 'categories'. In this way the Kohonen map not only learnt the concepts, but also simulated an 'automatic categorisation' of the concepts.

It is interesting to note that during learning the connectionist network was not provided any category information nor explicit definition of the semantic features and the possible relationships among them. Nonetheless, the Kohonen map itself deduced the similarity among the 'defining features' of various concepts and 'automatically' created clusters or categories of close concepts. The boundaries of these emergent categories are 'flexible' and this is for two reasons. Firstly, the representations originate from experiences within the environment, which may be 'noisy' at times. Secondly, the members of different categories share some features, allowing the categories to overlap, thereby incorporating the notion of 'family resemblance'.

**Local organisation of concepts inside a category:** The same categorising principle which earlier formed concept categories based on 'defining features' is again responsible for creating a local organisation or 'sub-categories' of concepts within a category based on the 'individual features' of various concepts belonging to the same category. Put simply, the Kohonen map's self-organising learning algorithm analyses the finer distinctions in the semantic feature vector of concepts belonging to the same category and then organises close concepts in proximity. For instance, in Figure 22 the agent category includes concepts dad, mum, Mary, and man that share a number of 'individual features' hence these concepts are stored in proximity to each other thus forming a sub-category, say 'humans'. Also, within the same agent
category, concepts for animals such as dog, pig, cow and horse are in proximity to each other, forming another sub-category 'animals'.

**Prototypical concept within a category or sub-category:** The category structure learnt by the Kohonen maps is such that it is possible to extract prototypical information across many exemplars, while simultaneously storing idiosyncratic information about individual exemplars. The most prototypical concept unit within a category can be found by presenting an input pattern that just provides information about category membership.

**Addition of new concepts to a 'learnt' concept memory:** Child theorists have speculated that the learning of a new concept is constrained by children's prior information about the environment. The categorisation of concepts helps in the learning of new concepts as a new concept can be perceived in terms of an existing concept, i.e., the features of a new concept are compared with the features of concepts in a particular category. For instance, the child may identify a new concept 'cat' in terms of a known and similar concept 'dog', in that the new concept 'cat' has features such as 'animal', 'has tail', 'has furry coat', 'roams in the house', 'is pet', etc. which are common to the child's existing concept of a 'dog'.

In our connectionist concept memory addition of new concepts takes into account the prior existence of a taxonomy of earlier learnt concepts. We demonstrate this aspect by adding a new concept 'cat' to the previously learnt concept memory (shown in Figure 22).

It may be noted that the new concept 'cat' (shaded dark in Figure 23) is learnt and mapped in the immediate proximity of the concept 'dog' by the Kohonen map's self-organising learning mechanism. This indicates three things: (a) the connectionist learning mechanism is aware of the existence of a category structure, (b) the learning mechanism not only 'automatically' determined the category of the new concept but also determined the sub-category to which it belonged, and (c) within the sub-category the concept 'cat' was placed next to the concept which bears greatest resemblance to it (i.e. the concept 'dog').
An explanation for the above behaviour is that the Kohonen map's self-organising learning mechanism earlier tuned the units in the neighbourhood of the concept 'dog' towards its semantic feature representation, thereby creating an area where the neighbouring units of the concept 'dog' have an internal representation that is close to it. Later, when the new concept 'cat' was presented to the Kohonen map to be learnt, it was mapped onto one of the neighbouring units of the concept 'dog' due to its similarity with the concept 'dog'.

The connectionist learning mechanism for adding new information has then the following characteristics: unsupervised learning, no re-organisation of the initial memory structure to accommodate new information, implementation of Piaget's notions of 'assimilation' and 'accommodation' (cf. Section 3.3), automatic and intelligent detection of the category of the new concept, and storage of new concept in proximity of similar concepts.

5.2.2 Experiment 1b: Concept Generalisation
In the previous experiment we showed how a connectionist network, using unsupervised learning algorithm, can simulate the development of concept memory ab initio. In this section we show how a
simulated concept memory deals with new concepts and with erroneous information. Ward and Vela (1989) have reported that the manner in which children generalise from a novel or partially visible category exemplar to other members of the category is likely to be influenced by many factors, including the children’s prior knowledge of previously learnt categories and the mode of processing. In connectionist terms, especially for the concepts stored in the concept memory, the former factor reported by Ward and Vela seems to be related to the category defining semantic features that are associated with members of a category. Whereas, the latter factor is related to the way those semantic features might be used to define a concept, i.e. either singly or in combination with one another.

We simulate two concept generalisation cases: first, concept generalisation when presented an incomplete representation of a learnt concept - ‘dog’, and second concept generalisation when presented a novel concept - ‘fox’. We argue that the manner in which the concept memory generalised for both the above cases seems much closer to what Ward and Vela had suggested to be the case of concept generalisation by children. The connectionist concept memory during concept generalisation first determined the appropriate conceptual category to which the novel or partially represented concept may belong and then selected one member of the candidate conceptual category that most resembled the input stimuli.

**Method**

**Network:** The learnt concept memory (developed in simulation 1) containing 42 concepts.

**Stimuli:** This experiment required two stimuli: (a) an incorrect semantic feature vector of the learnt concept ‘dog’ and (b) the semantic feature vector of a novel concept ‘fox’. We give below both input stimuli as used in this simulation. For comparison purposes the actual semantic representation of ‘dog’ is also given.

<table>
<thead>
<tr>
<th>Incorrect semantic feature vector of the concept ‘dog’</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Semantic feature vector for a novel concept ‘fox’</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Correct semantic feature vector of the concept ‘dog’</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
</tbody>
</table>
Procedure: In this simulation we do not perform any learning, rather we use the recall mechanisms of Kohonen maps to retrieve a concept. The input stimuli was presented (one at a time), to the input layer of the Kohonen map used to simulate the concept memory. This results in the concept unit most representative of the input semantic feature vector to be highly activated, and the corresponding concept is deemed to be retrieved.

Results and Discussion

The ability of the concept memory to generalise to an incorrect semantic feature representation of the concept 'dog' can be observed in the response of the concept memory (shown in Figure 24a). Figure 24b shows how the concept memory has generalised when presented the novel concept 'fox' by retrieving the closest concept - 'dog'.

We noted the activation level of the retrieved concept unit 'dog' in response to the incorrect representation and found it to be 0.74. Similarly, the activation level of the 'dog' unit when generalising the novel concept 'fox' was found to be 0.81. When presented with the a correct representation of the concept 'dog' the same 'dog' unit acquires a higher activation level of value 0.98. This decrease in activation level is understandable as when presented both incorrect and novel input patterns the 'dog' concept unit is not able to generate the maximum activation because of a slightly higher value of
Euclidean distance. Although the activation level of the ‘dog’ concept unit was not as high as when presented its correct representation, yet the fact that its activation level was still highest amongst all other concept units validates the efficacy of the Kohonen map to generalise and furthermore verifies the claims associated with learning in Kohonen maps.

Generalisation in Kohonen maps is at two levels: coarse and fine grained. Coarse-grained generalisation helps to determine (a) the possible category and (b) some close concepts, whereas fine-grained generalisation helps to retrieve a particular concept from the many category members that best represents the novel or incorrect input pattern. It may be noted that generalisation at first is coarse-grained; determining the category, and then there is the fine-grained generalisation that narrows the search to one member of the category that satisfies the most constraints in the input feature representation.

5.3. Experiment 2: Development of a Word Lexicon

One significant manifestation of the development of language amongst children is their ability to comprehend and produce spoken language. One can model this aspect of language development by arguing that, children can analyse acoustic input in terms of its constituent phonemes, the claim is that ‘children are predisposed to make many of the acoustic discriminations relevant to their subsequent language acquisition’ (Wales, 1993: 52). Else where it has been suggested that ‘young children may well be capable of perceiving fine phonetic detail, but they may represent this detail as a set of only loosely organised phonetic features’ (Charles-Luce, 1990: 206). When children either speak or understand spoken language they are relating the perceived phonetic stimuli to words stored in a kind of mental word lexicon which stores the words known to a child.

The ability to ‘spot’ words in continuous speech and the subsequent learning or storing of these words in the word lexicon has much to do with the presence of the ‘similarity neighbourhoods’ in the word lexicon26. The concept of this neighbourhood relates to the fact that similar sounding ‘words’ would be represented in a cluster and further away from words that do not have similar sounds. For instance,

26 Similarity Neighbourhood is defined by Charles-Luce as ‘a set of words that differ from a given target by a phoneme substitution, addition or deletion’ (1990:207).
Charles-Luce (1990: 207) has argued that, the similarity neighbourhood for the word *pit* would include the words *bit, pot, pig, spit, and it*, amongst others.

From a connectionist standpoint, then, one can argue that given a phonetic input to an unsupervised learning connectionist network, such as a Kohonen map, the output from the connectionist network construes to be a set of words corresponding to different phonetic inputs. Also, the organisation of these words in the connectionist *word lexicon* predicates a discrimination of phonetic information leading to 'similarity neighbourhood'. Put simply, these similarity neighbourhoods seem analogous to the categorisation of the word lexicon, which results as a consequence of the temporal organisation of phonetic information. In this simulation we develop a word lexicon which stores the words known to a child at any particular stage.

**Method**

**Network:** We use the connectionist architecture - Kohonen map., implementing the word lexicon (specified in Section 4.4.1). Recall that the parameters of the word lexicon were as follows:

- Input layer: 5 units
- Competitive (output) layer: 121 units

Prior to any learning the word lexicon is initialised with random weights in the range 0 - 1.

**Stimuli:** The stimuli for developing a word lexicon are 42 words that correspond to the 42 concepts taken from Bloom's archive and stored in the concept memory in experiment 1 (Section 5.2). We argued earlier that children recognise words in terms of their constituent phonemes, therefore we believe that a 'phonetic feature vector' is relevant to the representation of words in terms of their constituent phonemes.

The phonetic representation of each word is taken from Oxford Advanced Learner's Dictionary, which identifies 44 (20 vowel and 24 consonants) phonetic symbols. According to our representation scheme

---

27 The connectionist literature contains examples of works where a network has been trained to recognise 'words' when the network is presented with an input which comprises largely of syllables, phonemes and so forth. For instance, Dell (1992) in his model of phonological retrieval represents each word by the number and kinds of syllables and phonemes it contains. Elman's (1992) TRACE model of speech perception, again identifies words at a phonemic level. Phonemic features are represented as individual nodes, such that the node's activation reflects the degree of presence in a word. Kawamoto (1987) has earlier used 48 phonemes to represent words and the presence and absence of each phoneme constitutes a word's representation.
each individual phoneme is assigned a unique numerical value in the range 0 to 1. The 'phonemic feature vector' for a word is then formed by concatenating the encoded value of its constituent phonemes in a vector notation. The length of our 'phonemic feature vector' is based on Charles-Luce's (1990) experiments which suggest that 80% of the words spoken by children below 5 years age have a word length (in phonemes) of five phonemes or less. Therefore, we restrict the length of our phonemic feature vector to five phonemes (For words having less than five phonemes, the phonemic feature vector is completed by adding a null value of 0.0). For illustration purpose the phonemic feature vector for some words is given Table 19.

<table>
<thead>
<tr>
<th>Word</th>
<th>Phonemic Feature Vector</th>
</tr>
</thead>
<tbody>
<tr>
<td>dog</td>
<td>[0.45, 0.60, 0.65, 0.0, 0.0]</td>
</tr>
<tr>
<td>bag</td>
<td>[0.25, 0.40, 0.65, 0.0, 0.0]</td>
</tr>
<tr>
<td>pig</td>
<td>[0.15, 0.20, 0.65, 0.0, 0.0]</td>
</tr>
<tr>
<td>dad</td>
<td>[0.45, 0.40, 0.45, 0.0, 0.0]</td>
</tr>
<tr>
<td>baby</td>
<td>[0.25, 0.42, 0.25, 0.20, 0.0]</td>
</tr>
</tbody>
</table>

Table 19: Exemplar phonetic feature vectors for a few words to be learnt by the word lexicon

It may be noted that, our encoding scheme takes advantage of the fact that a connectionist 'distributed' representation can have values ranging between 0-1, therefore eliminating the need for a lengthy 'binary valued' representation vector as seen in Kawamoto's (1987) model. Furthermore our encoding scheme ensures that similar sounding words have close phonemic feature vectors, which leads to their storage in proximity. In this way, the word lexicon can be categorised on the basis of the length and elements of the phonemic feature vectors of the learnt words.

Procedure: Simulation of the development of the word lexicon is performed in a similar manner as that the development of the concept memory. Again the key to learning is the repetitive presentation of the phonemic feature vectors that are to be learnt. Starting with a random Kohonen map, during the learning session the phonemic feature vectors of the 42 words are presented in a random order to the word lexicon. The learning criteria is again the increase of activation level towards unity and the corresponding decrease of the Euclidean distance to zero. Figure 25 and Figure 26 show a similar learning profile for the word lexicon as that of the concept memory.
It may be noted that, for the development of the word lexicon, these criteria are satisfied rather earlier, i.e. around 4500-5000 iterations. This can be attributed to the reduced dimension and complexity of the phonemic feature vector, as opposed to the semantic feature representation of concepts. Therefore imposing less self-organising demands on the learning algorithm.

Results

After a learning session the initially random word lexicon (shown in Figure 27) has evolved to store the 42 words presented to it. The learnt word lexicon is shown in Figure 28. Words are stored in an orderly manner which reflects a categorisation of words on the basis of the length and content of their phonemic feature vectors.
Like we did for the learnt concept memory, we have marked regions of the Kohonen map that store words of similar phonetic lengths. These regions resemble the 'categories' or 'similarity neighbourhoods' suggested by researchers to be present in children's lexicons. It can be seen that the Kohonen map can be demarcated into four broad regions based on the length of the phonemic feature vector of the words: the top left stores words that are two phonemes long, right to this region is the large area occupied by words comprising three phonemes; the bottom left region contains words comprising four phonemes, and finally the single five phoneme word blanket occupies a small area between the two phoneme and five phoneme regions. There is a slight misclassification involving the two phoneme words as three of them are stored near the bottom right side of the Kohonen map.

As we know that in the self-organisation process, the distribution of the weight vectors becomes an approximation of the input vector distribution. This construes that the most frequent areas of the input space are represented by a large number of units, i.e. categories with a large number of members occupies a large area on the map. This behaviour is evident by the fact that in the word lexicon, three phoneme words being in majority occupy the largest area on the map.

The 'learnt' word lexicon also discriminates words on the basis of their phonetic content, and also within categories similar sounding words are stored in proximity, for instance note that similar sounding words - 'bag', 'dog', 'pig', 'big', 'dad' and 'duck' are stored close to each other. This observation is certainly in
accordance with Charles-Luce’s earlier mentioned definition of ‘similarity neighbourhoods’ as the above words differ from each other by a phoneme substitution, addition or deletion.

By way of this simulation we have learnt a word lexicon which stores words corresponding to the concepts learnt earlier. The resultant word lexicon satisfies the similarity neighbourhood constraint, and its effects can be further exhibited when the word lexicon is required to generalise when subjected to novel words.

5.4. Experiment 3: Learning Conceptual Relations and Associating Them With One-word Utterances

From earlier discussions of the referential purpose of children’s early words (Section 2.5) it appears that the task of learning conceptual relations, which lead to the production of one-word utterances, actually involves the learning of two associations: first, an association between conceptual relations and the so-called perceptual entities and second an association between conceptual relations and the (mainly functional) words which express the learnt conceptual relations.

Although Bloom does not consider that one-word utterances are a kind of telegraphese for two-word sentences or that such utterances in themselves comprise the so-called one-word sentences, she does regard such utterances as an important developmental indicator. These utterances are used by children to express a variety of complex conceptual relations, like recurrence, request, existence, disappearance and so on in a behavioural and social context. For instance, after finishing a glass of milk, the child might ask for more milk by merely saying more. And, then go onto extend the use of more by asking a caretaker, who was tickling the child, to tickle more.

Most child theorists focus on those one-word utterances that involve function words like more, away, gone, no, etc. (cf. Section 2.5.1) rather than substantive single words. These utterances indicate the

28 We regard perceptual entities to be aspects of the environment that the child can perceive, or more appropriately visualise and then talk about them. From Bloom’s (1973) account the prominent perceptual entities that are frequently talked about by children can be categorised as objects, people and events (such as eating, playing, moving, etc.). We understand that in reality there may exist a more elaborate set of perceived entities, however for our purpose this limited set suffices as it encompasses a major portion of children’s communicative ‘intentions’.
child's grasp of conceptual relations and perceptual entities. These perceptual entities embrace objects, people and events.

The successful use of one-word utterances depends crucially on the responses of the caretakers in particular and the child's environment in general. To put it more strongly, one-word utterance, a highly abbreviated symbolism, flourishes under supervision of caretakers. Moreover, this verbal one-word response to a stimulus that involves conceptual relations between perceptual entities is a mapping task. The successful simulation of the learning of conceptual relations between perceptual entities therefore requires a supervised learning network. We have chosen backpropagation network because it has capabilities of pattern association and can also generalise (cf. Section 4.4.3). That a pattern association network can generalise has implications here in that the network can automatically discern associations which are currently encoded in the input patterns: something we believe, and the literature shows, that children do.

We have trained a BP network on a corpus of 32 input patterns, derived from Bloom's archive (cf. Table 6) of Allison during the one-word stage (c: 9-16 months), that led to the utterance of the relevant function words again based on Allison's utterances.

**Method**

**Network:** The conceptual relation network (cf. 4.4.3) implemented as a backpropagation (BP) network has the following parameters:

- Input Layer: 25 units  
- Hidden Layer: 5 units  
- Output Layer: 18 units

The BP network is initialised with random values in the range -1.0 to 1.0, this ensures that it does not contain any prior knowledge.

**Stimuli:** Since the connectionist simulation reported here aims to mimic the child's production of one-word utterances the input stimuli should comprise two components: the underlying 'intention' to communicate (a conceptual relation), the perceptual entity which is the subject of the so-called
communicative act. The output response is the linguistic manifestation of the underlying intention, i.e., a one-word utterance. In its entirety, the input stimuli depicts a situation in the child's experience which represents the child's communicative 'intention' in terms of a conceptual relation and a perceptual entity. The output response is a speech act: the child's one-word utterance in response to the input stimuli.

Consider a situation (see Table 20) reported by Bloom when the child is reported requesting for some juice. Here, one can assume that the underlying communicative 'intention' is the demand or request of a perceptual entity, i.e. juice. The input stimuli then is the conceptual relation recurrence and the perceptual entity objects. Furthermore, the child is reported to utter the word functional 'more' to express his or her demand for more juice, therefore the verbal response is the one-word utterance 'more'.

<table>
<thead>
<tr>
<th>Situation</th>
<th>Input Stimuli</th>
<th>Verbal Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Mother pours herself juice)</td>
<td>recurrence</td>
<td></td>
</tr>
<tr>
<td>(Allison picking up empty cup)</td>
<td></td>
<td>more</td>
</tr>
<tr>
<td>(Allison putting her cup aside)</td>
<td>object (juice)</td>
<td></td>
</tr>
</tbody>
</table>

Table 20: An exemplar situation involving Allison and her one-word utterance taken from Bloom's data.

Bloom's archives contain numerous such situations together with the child's response and we have selected 32 situations to formulate our set of training patterns for this experiment.

To represent the input stimuli and output response we have adopted a localist representation scheme where one unit represents one conceptual relation, perceptual entity or word. Input and output patterns are created by indicating the presence of the relevant conceptual relation and perceptual entity in the input pattern and the associated word in the output pattern. We use the binary digits 1 and 0 to indicate presence or absence of some information, respectively. Table 21 shows the 25-dimensional input pattern for the situation given in Table 20: the presence of the conceptual relation recurrence and the perceptual entity object is indicated by the binary digit 1. Similarly, the desired output response is represented by giving a value 1 to the functional word more in the 18-dimensional output pattern.
Table 21: Exemplar input and output patterns for the situation given in Table 20. The legend for the perceptual entities illustrated alongside the conceptual relations is as follows: O = Object, P = People and E = Event.

<table>
<thead>
<tr>
<th>Unit No.</th>
<th>Conceptual Relation</th>
<th>Input Pattern</th>
<th>Unit No.</th>
<th>One-word Utterance</th>
<th>Output Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Disappearance (O, P)</td>
<td>0</td>
<td>1</td>
<td>GONE</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>Disappearance (O, P)</td>
<td>0</td>
<td>2</td>
<td>AWAY</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>Request (O, E)</td>
<td>1</td>
<td>3</td>
<td>MORE</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>Recurrence (O, E)</td>
<td>0</td>
<td>4</td>
<td>THIS</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>Existence (O, P)</td>
<td>0</td>
<td>5</td>
<td>THERE</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>Existence (O, E)</td>
<td>0</td>
<td>6</td>
<td>UH OH</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>Existence (E)</td>
<td>0</td>
<td>7</td>
<td>NO</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>Non-occurrence (E)</td>
<td>0</td>
<td>8</td>
<td>PERSON'S NAME</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>Failure (E)</td>
<td>0</td>
<td>9</td>
<td>STOP</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>Rejection (E)</td>
<td>0</td>
<td>10</td>
<td>BIG</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>Non-existence (O, P)</td>
<td>0</td>
<td>11</td>
<td>SMALL</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>Greeting (P)</td>
<td>0</td>
<td>12</td>
<td>DIRTY</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>Cessation (E)</td>
<td>0</td>
<td>13</td>
<td>UP</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>Pointing (O, P)</td>
<td>0</td>
<td>14</td>
<td>DOWN</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>Actor (O, E)</td>
<td>0</td>
<td>15</td>
<td>OBJECT'S NAME</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>State-Large size (O)</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>State-Small (O)</td>
<td>0</td>
<td>16</td>
<td>OBJECT (O)</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>State (P)</td>
<td>0</td>
<td>17</td>
<td>PEOPLE (P)</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>Upness (O)</td>
<td>0</td>
<td>18</td>
<td>EVENT (E)</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>Downness (O)</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Substantive (O)</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Substantive (P)</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Perceptual Entities</th>
<th>23</th>
<th>OBJECT (O)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>24</td>
<td>PEOPLE (P)</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>EVENT (E)</td>
</tr>
</tbody>
</table>

Procedure: During learning a training (input-desired output) pattern pair is presented to the conceptual relation network such that the 25-dimensional input pattern maps on to the 25 units of the input layer and similarly the 18-dimensional output pattern maps on to the 18 output units of the conceptual relation network. The learning sequence involves the gradual development of an association between the conceptual relation and the corresponding desired one-word utterance. Also, in a rather covert manner the connectionist network learns a relationship between the conceptual relation and perceptual entity encoded in the input pattern.

Learning in a BP network progresses through a number of iterations. In each iteration, a training pattern is randomly selected from the ensemble of training patterns (32 in total) and presented to the conceptual relation network to be learnt. The error signal generated by the current training pattern is then used to adjust the connection weights. For updating the weights we prefer the 'pattern update
schedule’ as opposed to the ‘batch update schedule’ because it seems that children are unlikely to monitor an average error of their training environment, rather they are more likely to monitor the error associated with each training pattern as it is encountered.

We know that the backpropagation learning algorithm (Appendix A) is based on an error minimisation process, where the value of the error at any stage is then regarded as the learning criteria. In our simulation, the cut-off error value was set to 0.05, and as soon the error produced by the BP network was below 0.05 the learning sequence was ceased and the connectionist network was deemed to have learnt the training patterns. In total about 5110 iterations were required to minimise the error to the cut-off value of 0.05.

Results

In order to show how our BP network successfully simulated one-word utterances we present a few test patterns comprising a learnt conceptual relation and perceptual entity as input to the input layer of the learnt conceptual relation network and note its response in terms of the activation level of the output units. The output unit acquiring the highest activation level is deemed to represent the response of the learnt conceptual relation network. Note that the output layer of the conceptual relation network is divided into two assemblies - the one-word utterance assembly and the perceptual entity assembly), therefore the most highly active output unit in the one-word utterance assembly corresponds to a one-word utterance produced and like-wise the most active unit in the perceptual entity assembly represents the perceptual entity associated with the conceptual relation. Having said that, we believe that it is useful to consider all output units with activation levels close to the highest activation level when analysing the goodness of learning as it provides indicators to the global knowledge learnt by the connectionist network:

---

29 In the batch update schedule the error signals are averaged over a batch of training patterns before the connection weights are adjusted.

30 The one-word utterance assembly quantifies the learning of an association between conceptual relations and words, and the perceptual entity assembly quantifies the other association established during learning, i.e., the association between conceptual relations and perceptual entities.
specifically output units with slightly less activation levels should be seen as relatively less strongly associated with the input conceptual relation\textsuperscript{31}.

Table 22 shows the performance of the learnt conceptual relation network. The testing pattern comprise a conceptual relation and perceptual entity, and in response to the input pattern the associated function word is produced. By comparing the response of the conceptual relation network with the possible relationships between conceptual relations and word given in Table 6, we conclude that the conceptual relation network has successfully learnt to map the input pattern to the relevant function word. This fact is further validated by the high activation level of the output unit representing the relevant function word, whereas the other output units are shown to have a relatively low activation level.

<table>
<thead>
<tr>
<th>Input Pattern</th>
<th>Output Produced</th>
<th>Activation Level of Retrieved Unit</th>
<th>Other Highly Active Output Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Request + Object</td>
<td>More</td>
<td>0.949</td>
<td>Big (0.074), Object name (0.027), Away (0.024), Up (0.021)</td>
</tr>
<tr>
<td>Greeting + People</td>
<td>Person name</td>
<td>0.953</td>
<td>Dirty (0.056), Away (0.027), Stop (0.020), There (0.015)</td>
</tr>
<tr>
<td>Cessation + Event</td>
<td>Stop</td>
<td>0.851</td>
<td>More (0.071), Dirty (0.034), Uh-oh (0.026), Small (0.022)</td>
</tr>
</tbody>
</table>

Table 22: Result of the simulation of learning conceptual relations. The input pattern comprises a conceptual relation and a perceptual entity. The output produced is a functional word corresponding to the conceptual relation.

The input patterns presented during learning encoded an implicit relationship between conceptual relations and perceptual entities, for instance the \textit{owner} of an object belongs to the perceptual entity \textit{people}, \textit{cessation} is relevant only to ongoing \textit{events} and so on. Learning conceptual relations did not involve the learning of such associations between conceptual relations and perceptual entities, however we argue that a successful simulation of the learning conceptual relations should have a semblance of the relationship encoded in the training input pattern. We show that our conceptual relation network has implicitly learnt to associate conceptual relations with the co-occurring perceptual entities (Table 23) by presenting as input just a conceptual relation and noting the perceptual entity activated in response.

\textsuperscript{31} When demonstrating the production of one-word utterances on the computer terminal we illustrate the activation levels of the output units on a colour scale ranging from red to yellow. The most active unit in each assembly is illustrated by a red colour whereas other slightly less active units are shown by lighter shades of red moving towards the colour yellow.
Again, by comparing the retrieved perceptual entity unit with the conceptual relation-perceptual entity relationships mentioned in Table 6, it can be argued that the conceptual relation network has indirectly learnt the knowledge about what conceptual relation can co-occur with which perceptual entity(ies).

We argued earlier that the learning of conceptual relations has implications for the production of one-word sentences. Researchers have argued that children have the ability to generalise in novel situations, i.e., to produce an appropriate response based on learnt knowledge when encountered with novel situations. We would like to test whether our conceptual relation network is able to exhibit generalisation affects. This can be achieved by presenting novel situations\(^\text{32}\) to the conceptual relation network and observing whether an appropriate function word is produced. The generalisation results (see Table 23) show that our conceptual relation network can certainly generalise in novel situations, as it produces responses that are both appropriate and based on learnt knowledge.

Determining the prototypical functional word (or one-word utterance) associated with a perceptual entity is an interesting observation as it indicates which functional word is most strongly associated with a perceptual entity (shown in Table 23). We show prototypical one-word utterances by presenting as input just a perceptual entity, and the note that the prototypical functional word associated with the perceptual entity. For instance, in one test we note that the function word corresponding to \textit{person name} is the prototypical word for the perceptual entity 'people', this implies that whenever the child encounters a person the most likely response is uttering the name of the person. Similarly, the prototypical word associated with the perceptual entity objects is \textit{big}, which mean that talking about the size of an 'object' (perceptual entity) as opposed to other aspects of an object is prevalent to the conceptual relation network.

\(^{32}\) A novel situation for the conceptual relation network would be as follows: the network has been trained on input patterns which comprise a conceptual relation and a perceived entity, for instance in the training pattern \textit{recurrence+object or event} the conceptual relation network learns that the conceptual relation \textit{recurrence} always occurs in conjunction with the perceived entities \textit{objects} and \textit{events}. A novel situation involving the conceptual relation \textit{recurrence} would be the recurrence of the perceived entity - \textit{people}. 
Production of child-like one-word utterances: We argued earlier that the learnt conceptual relation network is to be used to simulate the production of child-like one-word utterances. Here we make use of the retrieval mechanism of the conceptual relation network to produce a one-word utterance. Recall that the input stimuli for this simulation is taken from the archives of child language recorded by Bloom (1973), which corresponds to a real-life situation which instigates an utterance by the child and the output response expected to correspond to an one-word utterance. Table 24, illustrates a comparison of the one-word sentences produced by both Allison and the conceptual relation network in similar situations.
<table>
<thead>
<tr>
<th>Situation</th>
<th>Allison's Response</th>
<th>Conceptual relation network's input pattern</th>
<th>Conceptual relation network's response</th>
</tr>
</thead>
<tbody>
<tr>
<td>(M pointing to chair)</td>
<td>chair</td>
<td>Pointing + Object (chair)</td>
<td>obj name (chair)</td>
</tr>
<tr>
<td>What is this?</td>
<td></td>
<td></td>
<td>(output unit 17)</td>
</tr>
<tr>
<td>(M pours self juice)</td>
<td>more</td>
<td>Request + Object (juice)</td>
<td>more</td>
</tr>
<tr>
<td>(A picking up empty cup)</td>
<td></td>
<td></td>
<td>(output unit 3)</td>
</tr>
<tr>
<td>(A putting her cup aside)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(M pours juice; A drinks juice, looks into empty cup, M taking cup) Where's the juice?</td>
<td>gone</td>
<td>Disappearance + Object (juice)</td>
<td>gone</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(output unit 1)</td>
</tr>
<tr>
<td>(A tries to get cookies out of bag; can't; pointing bag to M)</td>
<td>Mama</td>
<td>Pointing + Object (Mum)</td>
<td>Mum</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(output unit 18)</td>
</tr>
<tr>
<td>(A holding picture to photographer's assistant, off camera) Where's the girl?</td>
<td>girl</td>
<td>Non-existence + People Existence + People</td>
<td>no</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(output unit 7)</td>
</tr>
<tr>
<td>(A turns picture over so she can't see the girl) (A turning it back to picture side)</td>
<td>no there</td>
<td></td>
<td>there</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(output unit 5)</td>
</tr>
<tr>
<td>(A pushing car) (car stops)</td>
<td>hmmm</td>
<td>Cessation + Event</td>
<td>stop</td>
</tr>
<tr>
<td>(A walks toward car; to urge to move) (A pushes car; then car stops)</td>
<td>stop</td>
<td></td>
<td>(output unit 7)</td>
</tr>
<tr>
<td></td>
<td>more</td>
<td>Request + Event</td>
<td>more</td>
</tr>
<tr>
<td></td>
<td>stop</td>
<td>Cessation + Event</td>
<td>(output unit 3)</td>
</tr>
</tbody>
</table>

Table 24: A comparison of the sentences produced by the conceptual relation network with actual child data. The letters A and M in the description of the situation refer to the child Allison and her Mother, respectively.

Table 24 shows that when the conceptual relation network is subjected to stimuli corresponding to real-life situations in the child’s language it is capable of producing child-like one-word utterances.

### 5.3.1. A Novel Modification of a Backpropagation Network to Realise Bi-directional Associations

In psychological parlance the connectionist simulation reported above aims to mimic the child’s understanding of conceptual relations and how to use them to express certain internal ‘intentions’ in terms of various functional words, the so-called one-word utterances. For that matter, the simulation mainly focused on learning a uni-directional association between conceptual relations and words. It remains of interest to speculate that, although the conceptual relation network can successfully associate a conceptual relation with a word, does the same conceptual relation network has any understanding of a reverse association, i.e., an association between words and conceptual relations. Child theorists may argue that in producing one-word utterances the child has an understating about what conceptual relation
can manifest a particular aspect of the environment or an internal ‘intention’, and additionally which word is best suited to express the relevant conceptual relation. However, one can argue that the child’s knowledge can not be restricted to just a uni-directional association, the argument goes that when the child learns to associate a conceptual relation with a word he or she is also developing an understanding that a particular word can be used to express certain conceptual relations, i.e. the referential purpose of words (cf. Section 2.5). This implies that whilst learning to associate conceptual relations with words the child is also inherently associating words with related conceptual relations.

Our argument implies that learning a uni-directional association between conceptual relations and words does not reflect true psychological plausibility, rather what actually need to be simulated is the establishment of bi-directional associations between conceptual relations and words. Here, we are confronted with a functional and structural limitation, in that, a typical backpropagation network is only capable of learning uni-directional associations. The straight forward approach to implement the required bi-directional associations is to use two BP networks: one BP network learning to associate conceptual relations with words and the other BP network associating words with conceptual relations.

The above approach does not impress us for two reasons: (a) it requires the implementation of an extra connectionist network, and (b) the same training patterns need to be stored in two separate connectionist networks, i.e., redundancy of knowledge within the connectionist system. To avoid the shortcomings of the above approach we have devised a novel modification to the design of the learnt conceptual relation. Our hypothesis here is that, if the conceptual relation network has learnt the knowledge about how conceptual relations are associated with words, then there is no need to re-learn the converse of this knowledge, rather the learnt knowledge can be made transitive by modifying the typology of the connectionist network. In this way we would be able to demonstrate the existence of an association between words and conceptual relations within the same connectionist network, such that, given a word as input the modified conceptual relation network retrieves the associated conceptual relation. We explain below a modification to the structure of the conceptual relation network to realise a modified conceptual relation (MCR) network.
Recall that the backpropagation network comprises three layers - an input layer representing conceptual relations, an output layer representing words, and a hidden layer representing the network's internal representation. The input layer is connected to the hidden layer by a set of weights (stored in a weight matrix, say $W_1$), that transform the input pattern containing conceptual relation information into an internal representation of the network. Similarly, the hidden layer is connected to the output layer by another weight matrix, say $W_2$, which transforms the network's internal representation to the desired output pattern containing word information. It therefore seems that two different kinds of knowledge are contained in the conceptual relation network; the weight matrix $W_1$ stores knowledge about how to transform conceptual relations to an internal representation, and weight matrix $W_2$ contains knowledge about how to transform an internal representation to produce words.

Our modification of the structure of the conceptual relation network is guided by the demands of the simulation, in that we want to retrieve as output a conceptual relation when a word is given as input. We give below the tenets of our scheme to modify the conceptual relation network to realise a MCR network.

- Since in the MCR network we would be presenting a word as input and retrieving a conceptual relation as output, we first reverse the role of the input and output layers of the conceptual relation network. The actual output layer of the conceptual relation network therefore becomes the input layer of the MCR and the actual input layer is to be treated as the output layer of the MCR network.
- Next, we plug the appropriate weight matrices between the layers of the MCR network. We know that weight matrix $W_2$ (having dimension $[5, 18]$) stores knowledge about how to transform an internal representation of the conceptual relation network to an associated word. Since, now we want to perform the opposite action, we transpose weight matrix $W_2$ to form a weight matrix $W_2^T$, which can transform a word oriented input pattern to an internal representation of the MCR network. The matrix transformation is as follows:

$$W_2 = [5, 18] \quad \rightarrow \quad W_2^T = [18, 5]$$

Similarly, weight matrix $W_1$, containing knowledge about how to transform a conceptual relation oriented input pattern to an internal representation, is transposed to realise a weight matrix $W_1^T$, which can now perform the opposite operation, i.e., transform an internal representation to generate an output pattern containing conceptual relation information. Again the matrix transformation is as follows:
In this way, we achieve a MCR network which can take as input a word and produce as output the associated conceptual relation. Figure 29 shows the topology of the MCR network originating from the typology of the conceptual relation network.

Figure 29: Topology of the actual conceptual relation network is shown on the left, while the modified conceptual relation network is shown on the right. Note that both networks are implemented as backpropagation networks.

Now that we have modified the conceptual relation network we examine the efficacy of the resultant MCR network in terms of its ability to activate appropriate conceptual relations in response to functional words presented as input. We argue that if the MCR network retrieves correct conceptual relations in response to input words, our initial hypothesis that the conceptual relation network has implicitly learnt bi-directional associations between conceptual relations and words would be satisfied. We test our hypothesis by presenting a number of functional words combined with perceived entities as input to the MCR network and observing the conceptual relation(s) retrieved together with its activation level (see Table 25). To determine whether the retrieved conceptual relation is the correct one, Table 6 illustrating relationships between conceptual relations and would be consulted.
There exist no non-existence person name actor existence (event) small cessation

Table 25: Performance of the MCR network. We also give other highly active conceptual relations, and one may note that in cases when a word is associated with more than one conceptual relation the other related conceptual relations are the ones next highly active, for instance consider the next highly active conceptual relations for the words 'there', 'no' and 'person name'. For these words all conceptual relations that are associated with them are shown to acquire high activation levels, thus signifying that they are strongly associated with the input word.

Our results show that the conceptual relation network whilst learning to associate conceptual relations with words did learn a corresponding association between words and conceptual relations, thus proving our earlier hypothesis to be true. In view of our results, it can be argued that this knowledge was resident in the weights of the learnt conceptual relation network and could be exploited by accessing it in the right way. By modifying the typology of the conceptual relation network we have provided a mechanism to explicate this implicit knowledge, in fact here we suggest a general-purpose mechanism which could be used to achieve transitive mappings or associations in learnt BP networks.

5.5. Experiment 4: Ostensive Naming-Development of 'Naming Connections'
Recall that we have already developed a concept memory (Section 5.2) and a word lexicon (cf. section 5.3) in the previous simulations. Having these two networks at hand we believe that we are ready to attempt a connectionist simulation of ‘ostensive naming’ of concepts by linking these two networks.

In child language literature, ‘naming’ of concepts is regarded as the mapping of children’s linguistic knowledge to their conceptual knowledge. Child theorists including Nelson et al (1978), Levine and Carey (1982), Callanan (1985) and Ward (1989) have argued that learning to ‘name’ or ‘lexicalise concepts is largely conducted through a process of ostensive naming. Ostensive naming of concepts can be envisaged as a relationship between concepts and words, it links the concept memory with the child’
early word lexicon such that concepts and words can be retrieved transitively: if a word is given then a corresponding concept can be retrieved and vice versa. Such a relationship is learnt over time during the earlier parts of the second year of infancy, between 10-20 months.

Ostensive naming then appears to be a good candidate for connectionist simulation of a kind not yet attempted within connectionism. Recall that we have individual networks to learn some aspect of child language: like the word lexicon (Section 5.3.) and the concept memory (Section 5.2.) - each containing 42 items out of Bloom's archive. Now we will attempt to link the two connectionist networks in order to simulate a kind of co-operative behaviour that might take place in a child's concept memory and word lexicon. Of course, the simplest way would be to merely link the corresponding units in the two networks. But we wish to demonstrate the learning capabilities of connectionist networks. For this purpose we intend to connect the two networks via the naming connection network (cf. Section 4.4.2.) and train the latter to establish the right connections between concept memory units and word lexicon units.

At least three ostensive naming stages identified by Nelson et al (1978) can be simulated by connectionist networks: the first (c: 10-13m) relates to the assignment of a known word to a known concept. The second (c: 13-15m) relates to the assignment of a word to a 'known concept' where the child has a concept of an object or event but lacks the appropriate word to express it: the child then first learns the word and then associates it with the known concept. The third (c: 16-20m) relates to the assignment of a word to a 'novel concept': the child hears a novel word referring to a novel object or event, then the child relates the novel word to the new concept. In all these three stages the 'naming' decision is that of the child - unsupervised learning to a certain extent. We believe that a successful simulation of these three stages using an unsupervised learning network (based on Hebbian learning) will lead to a better understanding of the role of ostensive naming in child language development on the one hand and on the other will demonstrate the effectiveness of connectionist simulations of child language development.

In this experiment we will simulate the development of an association between a lexical label (word) with the corresponding concept: essentially a 'naming connection' is learnt between a concept unit in a
concept memory to the corresponding word unit in the word lexicon. Each unit in the concept memory would be connected, with varying connection strengths, to all units in the word lexicon and vice versa, thereby establishing a many-many relationships between concept units and word units. In the following discussion we will present a simulation of the first stage of ostensive naming (due to lack of space stages 2 and 3 will not be discussed).

**Method**

**Network:** For the ostensive naming simulations we use the naming connection network (cf. Section 4.4.2) based on the two Kohonen maps (reported respectively in Section 5.2 - concept memory and Section 6.2 - word lexicon) each having 121 units in their output layers. Since the output layers of the two Kohonen maps are to be connected there are 14621 (121 * 121) Hebbian connections between the two Kohonen maps. These Hebbian connections are used to spread the activations from one Kohonen map to another such that a localised activity pattern in either Kohonen map will cause a corresponding localised activity pattern on the other Kohonen map.

**Stimuli:** Two types of stimuli were used in this experiment: (1) a perceptual stimuli, i.e. a 20-dimensional semantic feature vector representing a concept, and (2) a phonemic stimuli, i.e. a 5-dimensional phonemic feature vector representing the corresponding word. The entire training set comprises Bloom's 42 concepts and words learnt earlier in Section 5.2 and Section 5.3, respectively.

**Procedure:** Ostensive naming can be simulated in a developmental manner by simultaneously presenting a concept (the perceptual stimuli) to the concept memory and the corresponding word (phonemic stimuli) to the word lexicon. Presentation of the respective stimuli to each Kohonen map results in a group of units to become highly active. Naming connections can then be established between these highly active units based on the Hebbian learning algorithm. Consider an exemplar situation for ostensive naming: an adult points towards a 'dog' and utters the sentence 'That is a dog', thus both the verbal and perceptual stimuli corresponding to 'dog' are presented to the child almost at the same instance. We assume that the child possesses both the concept and word 'dog'. The presentation of the perceptual stimuli to the
concept memory forms a localised pattern of activity around the 'dog' concept unit on the output layer of the Kohonen map simulating the concept memory. Since the concept 'dog' was previously learnt therefore the 'dog' concept unit acquires the highest activation level amongst all other concept units. In a similar manner, the presentation of the verbal stimuli 'dog' to the word lexicon results in the word unit 'dog' acquiring the highest activation level. Ostensive naming is then achieved by applying the Hebbian learning algorithm to establish inter-map naming connections among the 'dog' concept and word units in the concept memory and word lexicon, respectively.

Ostensive naming is carried out in an iterative manner, where in each iteration a concept-word pair is presented to the naming connection network and learning involves slight increments to the strength of the Hebbian connections between the concept and word units. In this way, over a period of several iterations strong naming connections are established between concepts and their corresponding lexical labels (words).

An important parameter of the Hebbian learning algorithm is the learning rate whose value determines the strength of connection between two units: if the value of the learning rate is high a stronger Hebbian connections are established whereas with a low learning rate less strong connections are established. For this experiment at the start of the simulation the learning rate was 0.2, however with increasing number of iteration the learning rate was proportionally decreased. Our reason for this progressive decrease in learning rate is as follows: at the initial stages we wanted to establish strong Hebbian connections between a large number of units, however at a later stage once naming connections were established between relevant units we wanted to just further strengthen the developed naming connections between the most active units and this was achieved by minimising the learning rate with increasing number of iterations.

The development of the naming connections was quantified at regular intervals of 500 iterations. This was achieved by temporarily stopping the learning sequence and using the partially developed naming connections to retrieve a word given a concept. Quantification of the simulation involved presenting a concept to the learnt concept memory, this results in all concept units acquiring an activation level and
the concept unit corresponding to the input concept having the highest activation level. The spreading activation mechanism (cf. Section 3.7.1) was then used to spread the activations of the concept memory via the current naming connections to the word lexicon, resulting in the word units acquiring some activation level. The word unit acquiring the highest activation level was regarded to be retrieved in response to the input concept. Ostensive naming of a concept is deemed to be achieved when its corresponding word unit acquires the highest activation level. This can be determined by noting the activation level of the most highly active word unit (i.e. the retrieved word unit RU) and the activation level of the (actual) word unit (AU) corresponding to the concept. The difference (DIFF) between the activation level of these two units is a measure of the learning achieved at any stage, as with an increased learning of ostensive naming the difference between the activation levels of these two units decreases. Naming connections are deemed to be established when the DIFF between the activation level of the RU and AU is zero. Table 26 gives the learning profile of four words - 'dog', 'cow', 'juice' and 'dad', that are represented in the learnt word lexicon by units 76, 88, 55 and 86, respectively (Recall that each unit in the Kohonen map has been assigned an identification number). Development of ostensive naming is illustrated by showing what word units in the word lexicon are retrieved in response to the above-mentioned concepts. Figure 30 graphically illustrates how ostensive naming is achieved in terms of a high value of DIFF being reduced with increasing iterations.

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Dog (76)</th>
<th>Cow (88)</th>
<th>Juice (55)</th>
<th>Dad (86)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range</td>
<td>RU</td>
<td>RU</td>
<td>RU</td>
<td>RU</td>
</tr>
<tr>
<td>1 - 500</td>
<td>2</td>
<td>36</td>
<td>70</td>
<td>17</td>
</tr>
<tr>
<td>501 - 1000</td>
<td>36</td>
<td>88</td>
<td>56</td>
<td>17</td>
</tr>
<tr>
<td>1001-1500</td>
<td>36</td>
<td>88</td>
<td>56</td>
<td>17</td>
</tr>
<tr>
<td>1501-2000</td>
<td>36</td>
<td>88</td>
<td>91</td>
<td>17</td>
</tr>
<tr>
<td>2001-2500</td>
<td>114</td>
<td>88</td>
<td>91</td>
<td>17</td>
</tr>
<tr>
<td>2501-3000</td>
<td>114</td>
<td>88</td>
<td>56</td>
<td>17</td>
</tr>
<tr>
<td>3001-3500</td>
<td>114</td>
<td>88</td>
<td>56</td>
<td>17</td>
</tr>
<tr>
<td>3501-4000</td>
<td>114</td>
<td>88</td>
<td>91</td>
<td>17</td>
</tr>
<tr>
<td>4001-6000</td>
<td>114</td>
<td>88</td>
<td>91</td>
<td>119</td>
</tr>
<tr>
<td>6001-6500</td>
<td>76</td>
<td>88</td>
<td>91</td>
<td>119</td>
</tr>
<tr>
<td>6501-7000</td>
<td>76</td>
<td>88</td>
<td>91</td>
<td>119</td>
</tr>
<tr>
<td>7001-7500</td>
<td>76</td>
<td>88</td>
<td>55</td>
<td>86</td>
</tr>
<tr>
<td>7501-8000</td>
<td>76</td>
<td>88</td>
<td>55</td>
<td>86</td>
</tr>
</tbody>
</table>

Table 26: Learning profile for ostensive naming

Figure 30: A graph showing ostensive naming
Table 26 shows what word units are associated with a concept at any stage, and this is determined by the word retrieved in response to corresponding concepts, based on the current naming connections established between the concept memory and word lexicon. For instance, at iteration 500 the concept 'dog' is associated with unit 2 in the word lexicon. This turns out to be an incorrect association since the actual word unit representing 'dog' is 76. During subsequent iterations the connectionist network is again incorrectly associating the concept 'dog', first with word unit 36, and then later with word unit 114. It is only after 6000 iterations that the connectionist network has correctly learnt to name the concept 'dog', as now the concept 'dog' is associated with word unit 76, which represents the word 'dog'. The learning profile for the other three concepts show a similar trend where first incorrect associations are established between concept and word units in the concept memory and word lexicon, respectively. This inaccuracy in retrieval is gradually rectified as learning progresses, and is reflected in the decrease of DIFF, as shown in Figure 30. Learning continues for a fixed duration of 8000 iterations, and our results indicate that at this stage the concepts are properly 'named' - strongly associated with corresponding words.

**Results and Discussion**

At the end of the learning sequence, it can be seen both Table 26 and Figure 30 that the correct word is retrieved in response to the corresponding concept, also the activation level of the retrieved word unit is reasonably high. A complete explication of the performance of the learnt naming connection network is demonstrated as a separate simulation (simulation 4b) concerning the retrieval of a word by presenting a concept.

The ostensive naming simulation was oriented in establishing naming connections emanating from the concept memory to the word lexicon, but due to nature of the Hebbian connections used we have in fact learnt a bi-directional or transitive relationship between concepts and words. Now we can not only retrieve words given concepts but alternatively we can use the same Hebbian connections to retrieve a concept associated with a word. In this scenario, by presenting a phonetic stimuli at the word lexicon the word unit is retrieved, and its activations can be spread through the naming connections, in the opposite direction, to activate the corresponding concept. Retrieval of a concept upon 'listening' its lexical label
is an important activity in the learning of semantic relations and word-order, and would be exploited in later simulations. We argue that our choice of using Hebbian connections for the ostensive naming of concepts has further implications as we are relieved of implementing a separate connectionist structure that links the words with concepts, rather in our case the Hebbian connections incorporated in the naming connection network would suffice to realise a relationship between words and concepts.

It may be noted that during the ostensive naming of a particular concept, say 'dog', not only the concept 'dog' is associated with the word 'dog' but also other close category members, such as 'pig', 'horse' and 'cow' also get associated with the word 'dog', although with a relatively less strong connection. This implies that strong naming connections are established between close category members and less strong naming connections exist among other not so close category members. We argue that these lesser weighted naming connections with the word 'dog' have a useful purpose: in case the naming connection between the concept and word 'dog' is damaged, the naming connections of the concept units in the immediate proximity of the concept 'dog' would then help in spreading the right amount of activation to the word lexicon such that desired word 'dog' acquires the highest activation level to be retrieved.

Our work is novel in that we have attempted to interrelate two connectionist networks, the concept memory and the word lexicon, through Hebbian connections in an unsupervised manner- a training regime that has empathy with the developmental paradigm of language development. Furthermore, we have demonstrated the transformation of one kind of representation to another representation; the perceptual representation at the concept memory gets transformed to a phonetic representation at the word lexicon, and vice versa.

5.5.1. Experiment 4b: Retrieval of a Concept-Word Pair using Naming Connections

The connectionist simulation of ostensive naming establishes a relationship between the child's concepts and words such that given a concept the corresponding word is retrieved. In this experiment we examine the efficacy of naming connection network by attempting to retrieve a concept-word pair by presenting a concept. Such a behaviour is relevant in the production of child language, as the psycholinguistic
assumption is that children's utterances are a manifestation of their concepts. We argue that our mechanism for concept retrieval based on matching semantic features has much similarity with Smith, Rips and Shoben (1974) 'feature comparison' model reported in psycholinguistic literature. We give below (Figure 31) a simplified version of Smith's concept retrieval scheme together with ACCLAIM's concept retrieval scheme.

![Diagram of Smith's and ACCLAIM's concept retrieval schemes](image)

Figure 31: Concept retrieval schemes for Smith's feature comparison model (shown at the left of the picture) and the scheme adopted by ACCLAIM (shown at the right of the picture)

Both in Smith's model and in ACCLAIM, concept retrieval is based on the presentation of a feature oriented stimuli to the concept memory, next a feature comparison mechanism, in our case the calculation of the Euclidean distance, compares the input stimuli with the feature representation of various learnt concepts and the concept which best matches the input stimuli is retrieved.

**Method**

**Subjects:** For this simulation we require three connectionist structures - the learnt concept memory, word lexicon and naming connection network storing 42 items taken from Bloom's (1973) data.
Stimuli: We would like to demonstrate the retrieval of the word 'dad' when given the corresponding concept 'dad', therefore the stimuli is the 20-dimensional semantic feature vector representing the concept 'dad'.

Procedure: This simulation does not involve any learning, rather it is can be seen as a vehicle to explicate the knowledge learnt by the concept memory, word lexicon and naming connection network. To begin with, the semantic feature vector for concept 'dad' is presented at the input layer of the concept memory. This brings into relief the information retrieval mechanism of Kohonen maps; the output units become activated and the unit with the highest activation level is taken to represent the input stimuli. Since the concept memory has earlier learnt the concept 'dad' therefore the concept unit 'dad' acquires the highest activation level.

To retrieve the lexical label of the retrieved concept 'dad' the naming connection network is next used. By employing the spreading activation mechanism the activation level of all concept units is spread through the naming connections to the word-lexicon. The amount of activation transmitted to each word unit is a function of the activation level of the connected concept units combined with their connection weights to the word unit. This flow of activation results in the emergence of localised patterns of activations on the word-lexicon, such that word units that are strongly connected with highly active concept units acquire a high degree of activation level. Again, the retrieval of a word is based on the same principle, i.e., the word unit which acquires the highest activation level is regarded to be retrieved.

Results

Figure 32 shows the state of the concept memory with the concept 'dad' being retrieved. Figure 33 shows the word lexicon after the spread of activations from the concept memory, resulting in the word 'dad' acquiring the highest activation level and being retrieved.

It may be observed (Figure 32) that the presentation of the perceptual stimuli to the concept memory has resulted in varying activation levels for the concepts belonging to various categories. Units in the conceptual category-'agents' have a high activation level as compared to the other units in different
categories. This suggests the overall selection of the 'agent' category by the presentation of just one member of the category. It may therefore be argued that during the retrieval of the concept first the broad category was selected and subsequently the selection was narrowed down to one category member that best represented the perceptual stimuli. This is a useful observation, as information retrieval in such a manner has some psychological relevance, as usually argued by psychologists.

Figure 32: State of the concept memory when presented the semantic feature representation of concept 'dad'. The degree of activation level is depicted by shades of grey, the higher the activation level the darker the shade of grey. Note that the concept unit 'dad' has the highest activation level.

Figure 33: State of the word lexicon after activations are spread from the concept memory. The degree of activation level is depicted by shades of grey, the higher the activation level the darker the shade of grey. Note that the word unit 'dad' is retrieved as it has the highest activation level.
In the word-lexicon (Figure 33) it may be observed that apart from the highly active word unit 'dad' the word unit 'mum' is the next most highly active unit. In connectionist terms this implies that the connectionist network has deduced that the concepts 'dad' and 'mum' are very similar to each other, which indeed is the case. Also, the words corresponding to concepts in the 'agent' concept category, for instance the words 'baby' and 'Allis', have higher activation levels as compared to other word units. The activation of words corresponding to agent concepts suggests a competition for selection among various words labelling the members of the same category, a behaviour observed in human lexical recall.

We argue that this simulation not only demonstrates the information retrieval mechanisms inherent in Kohonen maps but also validates the efficacy of the Hebbian connections implemented in the naming connection network.

5.6. Experiment 5: Learning Semantic Relations

Child theorists such as Brown (1973) and Bloom (1973) have argued that the development of language from the one-word stage to the two-word stage is underpinned by the learning of a semantic relation between the categories of two concepts. Brown (1973) argues the child deduces the underlying semantic relations between categories of concepts without any guidance from a teacher, rather this is achieved by exploiting his or her knowledge of the presence of certain concept categories and an interaction with the environment. This interaction involves comprehending adult sentences and the visual perception of entities such as objects, persons and events.

This experiment is concerned with the learning of a set of semantic relations between concept categories proposed by Brown (1973), presented earlier in Section 2.5. Child language researchers have us understand that no teacher is involved in learning semantic relations, therefore one can assume that they can be learnt in an unsupervised manner. In connectionist terms learning semantic relations between two concept categories can be simulated by developing an association between two concept categories (where each concept category is represented by an individual unit) such that when one concept category unit is highly active the other associated concept category unit also gets activated. This kind of knowledge is
best stored in a memory which comprises individual units representing conceptual categories and the associations between units is developed in an unsupervised manner. Recall that the class of unsupervised connectionist networks - additive Grossberg networks were considered plausible for simulating semantic relations because they characterise a matrix memory and an unsupervised learning mechanism (cf. Section 3.7.1).

Learning of semantic relations is based on the assumption that a child having the knowledge of conceptual categories and the words associated with concepts, listens to adult sentences (in our case adult two word collocations) until he or she come across an instance of a semantic relation between two concepts. The actual stimuli received by the child is an adult two-word collocation, however what is actually required for the learning of semantic relations is the category of the two concepts embedded in the two-word adult collocation. Recall that category information is encoded in terms of the 'defining features' of the concept's semantic feature representation (cf. Section 2.4.1). Therefore, if one intends to perform a psychologically plausible simulation of the learning of semantic relations then a mere association of two units representing concept categories, either by using a BP network or else, may not suffice. Rather, what is needed is a transformation of the input stimuli (i.e. words) to the category information, in a manner that may have some relevance to the processing of the child and may incorporate the previously acquired knowledge of words and concepts.

Transformation of the verbal stimuli is therefore an integral step in the learning of semantic relations. In our simulation this transformation is made possible by an interaction between the earlier learnt concept memory (cf. Section 5.2), word lexicon (cf. Section 5.3) and naming connection network (cf. Section 5.5). The learning scheme can be summarised as followed: learning semantic relations begins with the presentation of a two word collocation to the word lexicon, which results in the retrieval of the corresponding two words. The transitive connections of the naming connection network are then exploited to retrieve the concepts corresponding to the two words in the adult collocation. The category information, i.e. defining features, of the two concepts is obtained from the semantic feature representation of the two retrieved concepts. The above scheme for transforming the verbal stimuli to
concept category information is illustrated in Figure 34 and explained in more detail in forthcoming discussion.

![Diagram](image)

Figure 34: The processing scheme for learning semantic relations

Now that the categories of the two concepts embedded in the two-word collocation are known a semantic relation, i.e. an association between the units representing those categories in the semantic relation network is created by the AGN learning algorithm. In passing we like to point out that such an interaction between a variety of connectionist networks leading to the learning of semantic relations is novel in the connectionist literature, in fact there is no instance of the learning of semantic relation even by a single connectionist network.

**Method**

**Subject:** For this simulation we use the semantic relation network specified in Section 4.4.4. To account for the environmental input this experiment uses the learnt concept memory and word lexicon storing the 42 items from Bloom's archive of child language, along with the naming connection network. The parameters of the semantic relation network, essentially an additive Grossberg network are:

- **Input layer:** 12 units
- **Intermediate layer:** 16 units
- **Output layer:** 12 units

Prior to learning all units are connected with uniform weights, thus ensuring that no associations between conceptual categories initially exist.
Stimuli: The stimuli to learn semantic relations has two components - linguistic and perceptual. The linguistic stimuli is an adult sentence presented as a two-word collocation. Each two-word collocation contains an underlying semantic relation between the two concepts expressed as words. The perceptual stimuli is a semantic feature vector corresponding to a perceptual entity to which reference is made in the adult sentence. Consider a prototypical stimuli for learning the semantic relation recurrence + object:

Linguistic component - two word collocation more milk

Perceptual component - Semantic feature vector for the perceived concept milk

The learning corpora consisted of 200 two-word adult collocations consistent with the set of semantic relations suggested by Brown (1973).

Procedure: The learning sequence for semantic relations comprises two phases: (a) learning associations based on the linguistic stimuli, i.e. two-word collocation; and (b) learning associations based on the perceptual stimuli.

Learning based on linguistic stimuli: The individual words of a two-word collocation are presented, one at a time, to the word lexicon. These words are recognised by the word lexicon such that corresponding word units acquire a high activation level. Recall, that the naming connection network implements a transitive association between the word lexicon and concept memory. To determine the concepts corresponding to the two words presented as input, the activations of the word units are spread from the word lexicon to the concept memory using the naming connection network. This results in the concept units corresponding to the each of the two retrieved words to acquire a high activation level be deemed as retrieved from the concept memory. This process can be envisaged as the reverse of the earlier simulation involving the retrieval of a word given concept (cf. Section 5.5.1).

We know that the semantic feature representation of a concept encodes information about its category, the so-called defining features (cf. Section 4.1.). Once a concept is retrieved by the naming connection network, its category is determined by examining the defining features of its semantic feature vector.
The category information of the two concepts corresponding to the two-word adult collocation form the input to the semantic relation network. For illustration purposes, we would call the category of the first concept in the two-word collocation as \textit{category} \textsubscript{1} and similarly the category of the second concept is called \textit{category} \textsubscript{2}.

Learning the semantic relation is now conducted as follows: At the input layer of the semantic relation network the activation level of the category unit corresponding to \textit{category} \textsubscript{1} is raised to a higher level, i.e. 1.0, whereas the activation level of other category units is kept at a low activation level of 0.1. Similarly, at the intermediate and output layer, the category unit corresponding to \textit{category} \textsubscript{2} is given a high activation level of 1.0. Next, the additive Grossberg network's learning algorithm, which is based on Hebbian notions of learning, is applied to learn or strengthen two different associations: (a) an association between the \textit{category} \textsubscript{1} and \textit{category} \textsubscript{2} units in the input and intermediate layers; and (b) an association between the \textit{category} \textsubscript{2} units in the intermediate layer and output layer. Since the \textit{category} \textsubscript{1} and \textit{category} \textsubscript{2} units are highly active the learning mechanism implements a strong connection between them, thereby associating two category units in different layers of the semantic relation network.

\textbf{Learning based on perceptual stimuli:} The perceptual stimuli available for this simulation corresponds to concepts known to a child and stored in the concept memory. The perceptual stimuli is directly presented to the concept memory, resulting in the activation of the corresponding concept. Again, the category of the retrieved concept is determined through the defining features of its semantic feature representation. In this case the concept can belong to one of the four categories - object, agent, location or attributes.

Learning the perceptual stimuli involves highly activating (a) the perceptual unit in the perceptual assembly of the intermediate layer corresponding to the category of the perceived concept, and (b) the \textit{category} \textsubscript{2} (determined whilst learning the linguistic stimuli) unit in the output layer. Again, the additive Grossberg learning mechanism is used to create an association between the perceptual unit in the intermediate layer and the \textit{category} \textsubscript{2} unit in the output layer. Learning of this association can be regarded as similar to the learning of an association between perceived entities and conceptual relations whilst learning conceptual relations in experiment 3 (cf. Section 5.4).
The above learning procedure can be understood by the following example: Consider the linguistic stimuli corresponding to the semantic relation *possessor + possession*, where *possession* relates to a perceptual entity - *object*. The linguistic stimuli is learnt by establishing two associations: (i) an association between the *possessor* unit in the input layer and the *possession* unit in the intermediate layer; and (ii) an association between the *possession* unit in the intermediate layer and the *possession* unit in the output layer. The perceptual stimuli is learnt by establishing an association between the *object* unit in the perceptual assembly of the intermediate layer and the *possession* unit in the output layer.

The entire learning sequence spanned 6000 iterations, where in each iteration an input pattern comprising a two-word collocation and perceived entity, was randomly chosen from the 200 training patterns.

**Results**

The result of the simulation is a semantic relation network that has learnt the possible semantic relations that are used by children in their two-word sentences as suggested by Brown (1973).

In order to highlight the performance of the semantic relation network we present an example: The intention here to determine the semantic relationship between the conceptual category *agent* (for example 'daddy') and an *object* (for instance 'chair'). The conceptual category 'agent' is presented to the input layer of the semantic relation network. The activation of the unit is spread from the input layer to the intermediate layer. At this stage we observe that the units *object* and *action* in the intermediate layer are highly active due to the input. This is because the category *agent* has semantic relationships with both the *object* and *action* categories. The high activation of two or more units in the intermediate layer makes the situation ambiguous in that it is not possible to determine which category is actually associated in this situation with the category *agent*. The perceptual stimuli is used at this stage to resolve the ambiguity that whether the input stimuli is represented by the semantic relation - *agent+object* or *agent+action*. In this example, the perceptual stimuli corresponds to the *object* category and maps on the object unit in the perceptual assembly of the intermediate unit. The highly active units in the
intermediate layer further spread their activations to the output layer, consequently activating a number of output units. Now, a process of competition takes place among the active units in the output layer, such that the most active unit inhibits the activations of the other units so that it becomes the unit with the highest activation level. At the output layer the object unit acquires the highest activation level, this is because both the highly active object units (one corresponding to linguistic component of the input stimuli and the other to the perceptual stimuli) in the intermediate layer passed their activation to it. The high activation level of the object unit in the output layer qualifies it to represent the actual second conceptual category, thus the ambiguity is resolved and the semantic relation is determined to be agent+object. This connectionist simulation therefore illustrates how semantic relations can be learnt and then exploited to produce two-word sentences.

5.7. Experiment 6: Learning Word-order
The objective here is to simulate children’s learning of word-order by listening to adult sentences. Learning of word-order is simulated in the background of Braine’s suggestions who predicates the presence of certain syntactic formulae which children use to combine two concepts to form a two-word sentence (cf. Section 2.6.2). Our simulation conforms to the developmental behaviour observed in children’s language, whereby children first learn a 'groping pattern' which is improved with time to reflect the word-order observed in adult's speech (Braine, 1976).

This simulation is first to observe and learn the inherent word-order in adult sentences and, second based on the learnt word-order to arrange two concepts, expressed as two words, to form a two-word sentence. From a connectionist perspective, learning of word-order can be simulated as a pattern association task: the input pattern comprising a two-word collocation is learnt in a supervised manner. Backpropagation (BP) networks characterise in pattern association and this is achieved in a supervised environment. We therefore use a BP network - word-order testing network (cf. Section 4.4.4.) to simulate the learning of word-order.
Word-order learning involves the learning the order of the two words in an adult two-word collocation - the environmental stimuli encoding word-order information. It has been argued by child theorists, in particular Braine (1976), that the child does not learn the order of two specific words rather he or she learns how to arrange the categories of the concepts which correspond to the words in the adult sentence. In this way, the child can generate a large number of two-word utterances by arranging a variety of words according to generic concept category oriented formulae, which in turn also relieve the child from learning how to order every word he or she knows to form a two word sentence.

The connectionist network learning word-order - the word-order testing network, takes as input two concept categories and learns the order in which the two concept categories are arranged. As, in the case of semantic relation learning, we notice that the input stimuli is a two-word adult collocation whereas the stimuli for learning word-order is category information. This calls for the transformation of the input stimuli to a representation that is compatible with the word-order testing network, i.e. category information. Such a transformation, again, necessitates an interaction between the other earlier simulated connectionist networks - concept memory (cf. Section 5.2), word lexicon (cf. Section 5.3), naming connection network (cf. Section 5.5) and semantic relation network (cf. Section 5.6). Each of these networks store information that has its origins in child language literature, in particular Bloom's (1973) 42 concepts and words. The simulation strategy can be illustrated by Figure 35 and is significant in the sense that ours is an initial attempt within connectionism to exploit the knowledge learnt in diverse connectionist networks in a unified manner to learn word-order. Conversely, one could have simulated word-order by directly presenting category information to a BP network, however our argument is that such a scheme would not hold much psychological plausibility as compared to a simulation scheme which brings into relief the underlying mechanisms that may be involved in the child's learning of word-order.
Once a connectionist network has learnt a certain word-order between two words, it is presented a sequence of two words, in response the connectionist network would generate an error value. If the two words are in an incorrect ordered sequence the error value would be high, whereas for a correctly ordered sequence of words the error would be low. Furthermore, if the error value is very high then the connectionist network will reverse the order of the two words so that now it conforms with the learnt word-order.

**Method**

**Network**: The word-order testing network communicates with the previously learnt word lexicon, concept memory, naming connection network and the semantic relation network in order to learn the word-order of words in the child's language. The word-order testing network is implemented as a back-propagation learning network and has the following parameters (cf. Section 4.4.5):

**Input layer**: 12 units  
**Hidden layer**: 5 units  
**Output layer**: 12 units
Stimuli: The stimuli - 200 two word adult collocations derived from Brown's data are transformed by the four networks to category information which is learnt by the word-order testing network. The stimuli received from the environment corresponds to a two-word adult collocation; the two constituent words represented in terms of their phonemic representation - a phonemic feature vector. Word-order information is assumed to be inherent in the order of the words comprising the two-word collocation.

The stimuli presented to the word-order testing network is a transformation of the initial phonetic stimuli; first the phonetic stimuli is transformed to the semantic feature representation of the concepts corresponding to the two words in the input, and next the semantic feature representation gets transformed to 'conceptual category representations' based on the semantic relation existing between the two concepts. Recall, that category information is embedded in the defining features (the first six semantic features) of a concept's semantic feature representation (cf. Section 4.1). In this way, the input to the word-order testing network comprises two six feature vectors (12 binary features in total) corresponding to the conceptual category representations of the two concepts corresponding to the input words. Since learning word-order is regarded by us as a case of pattern association, therefore the desired output pattern is a copy of the input pattern.

In order to explicate the above discussion of how the stimulus is presented to the word-order testing network consider as input a two-word adult collocation more juice. The phonemic feature vectors of the two words more and juice are then the input phonetic stimuli. The phonetic stimuli first gets transformed to the semantic feature vectors of the concepts 'more' and 'juice', determined to be recurrence + object. Next the conceptual category vector for the two concepts is derived from their semantic feature vectors. In this case the two conceptual categories are 'recurrence' and 'object'. Finally, the input stimuli to the word-order testing network would be the conceptual category recurrence followed by the category object, where each conceptual category is represented by a six feature long conceptual category vector.
Procedure: As mentioned earlier, the complete simulation of word-order learning is divided into two stages. In the first stage, when children do not have an understanding of a definite word-order, a 'groping pattern' is simulated by associating two semantically related conceptual categories in all possible order, i.e., concept A followed by concept B, and concept B followed by concept A. For instance, if we have two categories agent and object then to learn groping patterns the input stimuli would be two patterns: (a) agent + object and (b) object + agent.

In the second stage, the actual word-order as observed in adult sentences is simulated by presenting two semantically related categories in the order corresponding to the word-order observed in adult sentences. An explanation of the learning procedure (illustrated earlier in Figure 35) is as follows: The individual words of the adult sentence (two-word collocation) are presented one at a time to the word lexicon. This results in the retrieval of the corresponding words. The naming connection network is then used to spread the activation of the word lexicon to retrieve the corresponding concepts from the concept memory. The conceptual categories of the two retrieved concepts are determined by examining their defining features. Next, the semantic relation network is used to determine the possible semantic relation between the two concepts. Note that each conceptual category unit in the semantic relation network is sensitive to a unique feature representation which corresponds to the defining features of concepts. The (conceptual category) representation of the two units representing the semantic relation is used to formulate a training pattern for learning word-order. Learning in both stages is carried out in an iterative manner, such that in each iteration a two-word adult sentence is randomly selected from the entire input data set comprising 200 adult, and presented to the word-order testing network to be learnt. The conceptual categories of the two concepts are concatenated in the order in which they appear in the adult sentence. The resultant 12 feature pattern is the training pattern, i.e., the input and desired output pattern for the BP network simulating the learning of word-order.

We first simulated the 'groping pattern' stage, for which we chose a cut-off error value of 0.30, i.e., we allowed the network to learn until the error produced was above 0.3. It maybe noted that this is a very high value of cut-off error, however, we have chosen a high error value to make the network's response flexible so that it can exhibit a high degree of 'groping' in the initial stages of learning. Starting with a
randomly initialised network, 1875 iterations were required to learn the so-called groping patterns. In the second stage, we simulated the learning of the actual word-order found in adult sentences. At this stage we did not initialise the word-order testing network, rather we built upon the 'groping' knowledge earlier learnt by the network to learn the actual word-order. In this way we mimic a developmental profile - acquiring knowledge by building on previous knowledge. To learn the actual word-order the cut-off error was reduced from 0.3 to 0.1 to ensure that we fine tune the connectionist network based on the training patterns. In the end, a further 2568 iterations were required to learn the entire training patterns.

Results and Discussion

The modular architecture in Figure 36 was used to learn word-order from 200 two-word adult collocations. Once the word-order was learnt, the architecture was presented with test two-word sequences to check whether or not the network can cope with inconsistent word-order. In order to elaborate on the above, consider two words 'juice' and 'more' presented in the same order. Concept 'juice' belongs to the 'object' category, whereas 'more' belong to the conceptual category - 'recurrence'. An test pattern constituting the conceptual category vectors of both these concepts is presented to the input layer of the word-order testing network; the test pattern is object (concept1) + recurrence (concept2).

![Figure 36: The word-order testing network showing the word-order assemblies - WO1 and WO2](image)

Recall that the output layer of the word-order testing network comprises two assemblies: assembly WO1 maps the two concepts in the order concept1 followed by concept2, whereas assembly WO2 implements the opposite mapping, i.e., concept2 followed by concept1 (see Figure 36). When an input pattern is presented to the input layer of the word-order testing network, each of these two output assemblies
produce an error, whose value reflects the word-order knowledge learnt by the word-order testing network. For instance, if the correct word-order of the two input words (concepts) is `concept1 followed by concept2` then assembly `WOI` would produce a low error value whereas assembly `WO2` would produce a high value, signifying that the word-order `concept2 followed by concept1` is not known to the network. Therefore, the correct word-order corresponds with the assembly producing the lowest error. For our example, the output error produced at assembly `WOI` was greater than the one produced at assembly `WO2`. This indicates that a two-word sentence in which the word-order is `object + recurrence` does not conform with the word-order experienced in adult sentences. Therefore the correct word-order for the two words is `recurrence + object`, and therefore the two-word sentence is `more juice`.

The above result is in accordance with the word-order observed in adult language. Thus, we argue that after a learning sequence which first learns groping patterns and next the actual word-order, the word-order testing network has adequately learnt the word-order inherent in adult sentences. It can be argued that, the word-order testing network not only learns the word-order between two words but also corrects the sequence of the two words if they are given in an order not previously experienced by the word-order testing network.

Another important aspect of the learning is that the connectionist network not learns just instances of two-word collocations, rather it learns the ordering of conceptual categories, as Brown (1973) argues. Thus ensures that the word-order learning is not just the memorisation of an ordering between specific words, rather it is the acquisition of the knowledge of how to order the conceptual categories. Therefore, even newly acquired words or words not presented in adult's two-word collocations can be appropriately ordered to form a two-word sentence.

5.8. Experiment 7: Production of Two-word Sentences

Children use (two-word) sentences to talk about themselves, their needs, beliefs, desires, state or otherwise to express some aspect of their environment. Children's utterances can be regarded as a means to an end: the child communicates to achieve some desired goal. Whenever a child produces a sentence
he/she always has an underlying 'intention' on which the utterance is based. Development in intention based communication - 'intentional communication' is related with the child's cognitive development. For instance, at an early age the child's act of crying is reflexive, but with increasing age it takes on an intentional communicative function. The child’s transition from Piaget's sensori-motor substage 5 to stage 6 (18-24 months) marks a change in his/her communicative behaviour: now a child conveys his or her intention to communicate by combining attention to the caretakers with attention to the derived goal. Communication, then, is instrumental in reaching the child's end goal (Small, 1990: 133-5). One may argue that the child expresses his/her 'intention', together with related concepts, by using words strung together in a meaningful manner to form a sentence. Based on the above assumption, we consider the child's two-word sentence to comprise two concepts: one concept representing the child's communicative 'intention', and another concept (corresponding to an external perceptual stimuli) representing some aspect related with the 'intention'. We explain below a simulation of the production of a two-word sentence which corresponds to a real situation (see Table 27) reported in Bloom's (1973: 235) data.

<table>
<thead>
<tr>
<th>Situation</th>
<th>Two-word Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allison reaching for cookie box in bag.</td>
<td>there cookie</td>
</tr>
<tr>
<td>Allison takes out box of cookies</td>
<td></td>
</tr>
</tbody>
</table>

Table 27: A real-life situation concerning Allison, and her response in terms of a two-word sentence. (Bloom, 1973: 235).

Given the above situation, it can be inferred that Allison's (the child) communicative 'intention' is to indicate the location of the object 'cookie'. Therefore, the two concepts which form the input to the simulation are:

**First concept:** The child’s communicative 'intention' represented by the concept 'there'. Note that the concept 'there' is a member of the category 'location'.

**Second concept:** The perceptual stimuli corresponding to the concept 'cookie'.

We mentioned earlier that ACCLAIM is a synthesis of various connectionist modules, each simulating some aspect of child language development. When each module has learnt its designated knowledge, be it concepts, words, semantic relations or else, then two-word sentences can be produced by exploiting the knowledge learnt by each module. In fact, the production of child-like two-word sentences requires an
interaction of the knowledge learnt by the various modules of ACCLAIM: concepts to be uttered are retrieved from the concept memory, and their corresponding lexical labels are retrieved from the word lexicon; the semantic relation between the two concepts is determined by the semantic relation network; and the hypothesis about the correct word-order is evaluated by the word-order testing network. The ability to produce two-word sentences, which in fact is the final output of ACCLAIM, is a measure of the success of our connectionist simulation of child language development. The production of two-word sentences also has implications on the efficacy of our hybrid design of ACCLAIM, because in order to produce two-word sentences different modules communicate with each other by either passing their output as input to other modules or by making processing decisions based on the activity in other modules. Production of child-like two-word sentences is a three stage task. We explain below the various stages in the processing sequence whilst simulating the production of a two-word sentence in response to the situation given in Table 27.

**Stage 1: Retrieve the two concepts and the corresponding words**

The input to the activity of two-word sentence production is two concepts - 'there' and 'cookie'. Since the child's concepts are transformed into words which are then combined together to form a sentence, in the first step we retrieve the two concepts from the concept memory and their corresponding words from the word lexicon. First, the concept 'there' is retrieved by presenting to the concept memory a semantic feature vector corresponding to the concept 'there'. The concept 'there' is retrieved as it acquires the highest activation level. To retrieve the corresponding word - 'there' the naming connections established between the concept memory and word lexicon are utilised. Activations are spread through the naming connections from the concept memory to the word lexicon. A strong naming connection between the concept and word 'there' ensures that, in response to the high activation of the concept 'there' the corresponding word unit 'there' acquires the highest activation level amongst all other word units; thus the word 'there' is regarded to be retrieved.

In a similar manner the second concept 'cookie' is retrieved along with the corresponding word 'cookie'. At this stage we therefore have two concepts and their corresponding words.
Stage 2: Determine the semantic relation between the two concepts

In the next step, the semantic relation between the two retrieved concepts need to be determined. The knowledge of semantic relations has earlier been learnt by the semantic relation network. The input to the semantic relation network consists of the concept categories of the two concepts 'there' and 'cookie'. At the start of the simulation we specify the category of the first concept, which was 'location', as it was necessary to provide the category of the concept that corresponds to the child's internal intention and it is that concept which determines what the child actually wants to speak about. However, since the second concept corresponds to an external perceptual stimulus, which in this case is the object 'cookie', we do not need to specify its concept category, rather we rely on ACCLAIM's learnt knowledge to determine the correct concept category for the second concept. Recall that the category information is defined by a concept's defining features, and is manifested in the organisation of the concept memory. Based on the location of the concept 'cookie' in the concept memory, ACCLAIM deduced that the concept 'cookie' belongs to the category 'object'. We now proceed to describe the processing of this information in the semantic relation network, shown in Figure 37 and Figure 38. The units are shaded with degrees of grey with respect to their activation level. Therefore, the higher the activation level of a category unit the darker the shade of grey.

The category 'location' of the first concept 'there' is mapped onto the first layer of the semantic relation network. This results in the 'location' category unit acquiring a high activation level (dark shade of grey in Figure 37). Activation spreads across the connections between the input and intermediate layer to all units in the intermediate layer. This results in all other category units that may possibly have a semantic relation with the category 'location' acquiring a high activation level. In this case, the category units - 'action' and 'entity' are activated with equal magnitude of activation as shown in Figure 37. Since both the active category units in the intermediate layer have the same activation level, ACCLAIM is not able to determine the second concept category that is related to the category 'location' to realise a semantic relation between two concept categories.
Figure 37: The input and intermediate layer of the semantic relation network, showing the highly active (input) category 'location' and equally active category units - 'action' and 'entity' in the intermediate layer.

At this stage, to determine the actual second concept category among a number other highly active candidate categories, the perceptual input corresponding to the second concept category is provided to the perceptual assembly in the intermediate layer of the semantic relation network. In this case, the perceptual input corresponds to the category 'object' and therefore the 'object' unit in the perceptual assembly is acquired a high activation level (shown in Figure 38). Once again activations are spread from the semantic relation network's intermediate layer to the output layer through the connections between these two layers. Activations from the highly active units ('Action' and 'Entity') in the intermediate layer and the perceptual category unit 'object' are spread to the output layer, resulting in various semantically related category units in the output layer acquiring a high activation level. In the output layer, the unit with the highest activation level amongst all other units is considered to correspond to the category of the second concept in the semantic relation. Note that in Figure 38, the category unit 'Entity' (Entity represents both objects and agents), has acquired the highest activation level. This is because the category unit 'Entity' received combined activations from the 'Entity' unit in the intermediate layer and the perceptual input's 'Object' unit. We now have the category 'location' in the input layer and the category 'Entity' in the output layer with the highest activation levels. Therefore, we deduce at the end of this stage, that the semantic relation between the two concepts 'there' and 'cookie' is Location <-> Entity.
Stage 3: Determine the correct word-order

In the final stage we need to determine the word-order, i.e., how the two concepts 'there' and 'cookie' are to be arranged in a sentence so as to reflect the word-order observed in everyday adult language. The word-order knowledge learnt by the word-order testing network is used to determine the correct word-order.

The words 'there' and 'cookie' corresponding to the first concept 'there' and the second concept 'cookie', respectively, are presented at the input layer of the word-order hypothesis testing network. The network tests whether the correct word-order should be (a) the first concept followed by the second concept or (b) the second concept followed by the first concept. Evaluation of either hypothesis is done by noting the error produced for each hypothesis. The hypothesis which produces the least error is considered to represent the correct word-order. Finally, the two words are placed according to the determined word-order to constitute the two-word sentence.

In this case, the word-order hypothesis testing network evaluated both hypotheses, i.e., 'there' followed by 'cookie' and 'cookie' followed by 'there'. The word-order hypothesis - 'there' followed by 'cookie' produced a low error value, as compared to the other hypothesis (see Figure 39). Finally, ACCLAIM
arranges the two words 'there' and 'cookie', according to the determined word-order, to produce a child-like two-word sentence - 'there cookie'.

Figure 39: Evaluating word-order hypothesis. The high error is shown by a darker shade of grey, whereas the low error has a lighter shade. The two-word sentence produced is there cookie'.

Table 28 gives a comparison of Allison's and ACCLAIM's response to some situations taken from Bloom's data (1973).

<table>
<thead>
<tr>
<th>Situation</th>
<th>Allison's Response Two-word Sentence</th>
<th>Inferred Semantic Relation</th>
<th>ACCLAIM's response Two-word Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Allison reaching for cookie box in bag)</td>
<td>there cookie</td>
<td>location + entity (Entity = object)</td>
<td>there cookie</td>
</tr>
<tr>
<td>(Allison takes out box of cookies)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Mother pointing to chair.) What is this?</td>
<td>that chair</td>
<td>demonstrative + entity (Entity = object)</td>
<td>that chair</td>
</tr>
<tr>
<td>(Mother pours herself juice). (Allison picking up empty cup)</td>
<td>more juice</td>
<td>recurrence + object</td>
<td>more juice</td>
</tr>
<tr>
<td>(Mother pours juice; Allison drinks juice, looks into empty cup. Mother taking cup)</td>
<td>gone juice</td>
<td>negative + object</td>
<td>gone juice</td>
</tr>
<tr>
<td>(Allison holding picture to photographer's assistant.) Where's the girl??.</td>
<td>there girl</td>
<td>location + Entity (Entity = agent) negative + agent</td>
<td>there baby</td>
</tr>
<tr>
<td>(Allison turns picture over so she can't see the girl)</td>
<td>gone girl</td>
<td></td>
<td>gone baby</td>
</tr>
</tbody>
</table>

Table 28: A comparison of Allison's and ACCLAIM's response, i.e. the two-word sentences produced in the given situation.
5.9. Conclusions
In this chapter we presented nine major simulations of the development of aspects of child language, particularly those aspects that can be regarded to be involved in the one-word stage and two-word stage of child language.

We argued earlier that the one-word stage can be characterised by the development of concepts, acquisition of words and the learning of conceptual relations. In simulating the development of concepts, we first devised a concept representation scheme that was based on the so-called semantic feature theory that is extensively used in child language literature. To represent concepts we incorporated the notion of defining features that were relevant to a broad classification of concepts and individual features that define the individual features of a concept. We would like to stress here that the origins of our scheme for representing concepts was not intuitive, as was the case in other simulations of concept development reported in the literature. The simulation of the learning of words was based on a phonological level of representation, and we demonstrated the learning of words encoded as phonetic features. Our contribution lies in the fact that we have operationalised the theoretical arguments by child theorists and child language observations in a computer system.

Both our simulations of the development of concept memory and word lexicon, using Kohonen maps, were a shift from the traditional approach to simulate the development of concepts and words in a supervised environment by explicitly lexicalising the concept and directly categorising them. We improved from previous approaches of concept memory and word lexicon development by using self-motivated unsupervised learning algorithms that take into account the distinction between perceptual and phonetic input and the innate learning mechanisms of a child. Categorisation of concepts and words automatically emerged as a consequence of an increased understanding of the learnt concepts and words, and new concepts and words could be learnt without altering the existing structure of the concept memory and word lexicon, respectively. The overall learning mechanism incorporated Piaget’s notions of assimilation and accommodation.
In another but related simulation (Experiment lb) we demonstrated that the learnt concept memory had
generalisation capabilities, much like a child who utilises his or her learnt knowledge to respond to novel
concepts and to incomplete or incorrect perceptual stimuli. We were able to show that the concept
memory learnt earlier did possess the ability to adequately generalise in the above situations to produce a
correct response.

We simulated the learning of conceptual relations in a supervised learning environment providing stimuli
that has its origins in child language data. We not only learnt to associate conceptual relations with
words but also our learning strategy ensured that an association between conceptual relations and
perceptual entities was also realised. In the context of generalising to novel situations, we were able to
demonstrate that our conceptual relation network was able to handle unfamiliar situations by producing a
response based on learnt knowledge, as indeed is the case with developing children. Our modification of
the original structure of the conceptual relation network was a novel attempt in connectionism to exploit
the uni-directional association learnt by a BP network to demonstrate the existence of bi-directional
associations. By way of our modification we were able to make the learnt knowledge transitive, i.e., if
there exists an association between conceptual relations with words then by modifying the structure of the
BP network we can have an association between words and conceptual relations.

Our simulation of the production of child-like one-word utterances was a case of using connectionism as
a vehicle for operationalising child language data. We believe that our exercise of arranging child
language data in a meaningful manner so that it can be used for computer simulations is a contribution in
itself.

Ostensive naming of concepts within connectionism was demonstrated by our simulation of the naming
of concepts. Concept naming was simulated by devising a novel connectionist architecture which
connected two Kohonen maps via Hebbian connections in an unsupervised manner. This simulation was
a case of exploiting the previously acquired conceptual and lexical knowledge of a 'child' to associate
concepts with words in three separate stages. The efficacy of the simulation was demonstrated by
retrieving a word when presented a concept at the concept memory, or else retrieving a concept when a word was presented to the word lexicon.

Learning of a set of semantic relations proposed by Brown (1973) and the ordering of two words in an adult-like manner to form a two-word sentence, marked the transition from one-word to two-word language. Semantic relations were deemed to be learnt by the child himself or herself with no apparent input from adults, however learning word-order was regarded to be aided by specific cues from adults as when the child expressed the words in the wrong order it provided an opportunity for adults to correct him or her. For that matter, semantic relations were learnt in an unsupervised environment using additive Grossberg network and word-order was learnt by the supervised learning backpropagation network. An important processing aspect was implemented in both these simulations: the stimuli received for each simulation was a two-word adult collocation which was then transformed to a representation that was relevant for the learning, and this was achieved by exploiting the knowledge learnt in various other connectionist networks. In this way, in both these simulations we presented a unique interaction amongst structurally divergent connectionist networks including two Kohonen maps, Hebbian networks and additive Grossberg network, enabling a transformation of the input stimuli. Although, both these tasks could have been simulated in a much simpler manner, which did not require the manipulation of input stimuli by various connectionist networks, but that would have seriously diminished the psychological plausibility of the simulation. Therefore, we argue that our simulations are unique, and more psychological well-grounded as the processing is much like the child’s processing of environmental input to learn semantic relations and word-order.

The simulation of the production of child-like two-word sentences is a unique effort in connectionist literature, both in the sense that incorporation of knowledge learnt in various stages of development indicates that connectionism can simulate developmental learning that may involves the satisfaction of multiple simultaneous constraints, and also in the operationalisation of child language data. The two-word sentences produced by ACCLAIM were in accordance with the data reported by Bloom and the manner in which they were produced was a validation in itself of our psycholinguistic framework for child language development.
Ours was an initial attempt to incorporate the various aspects involved in the child's language at the one-word and two-word stage in a unified connectionist architecture to simulate a behaviour that has resemblances with children's spoken language. The question about the significance of and the existence of more semantic features, conceptual relations, semantic relations and other in the child's language is one to be addressed by child theorists. However, for our purpose we have shown how such knowledge can be learnt and associated with each other within the connectionist paradigm to produce a variety of child-like one-word and two-word sentences. Indeed, with an improved exposition of these psycholinguistic aspects in child language literature, connectionists can pursue further to built larger connectionist networks that can handle an enhanced set of semantic relations, conceptual and semantic relations, word-ordering rules and encompass a larger variety of child utterances.

The principles presented here for a connectionist simulation of child language production would still hold valid and future connectionists can benefit from our methodology.
Chapter 6

Conclusions and Future Work

6.1 Conclusions
Child language development is an interdisciplinary subject involving subjects as diverse as psychology, neurolinguistics, psycholinguistics, neurobiology, anthropology and sociology. Connectionism is just as diverse a subject and has inputs from computer science, neurobiology, mathematics, cybernetics, artificial intelligence, linguistics, and developmental psychology. Given that child language development studies and connectionism are both interdisciplinary, the researchers in both disciplines occasionally have to exaggerate to make a point, omit some details here, sometimes confuse theoretical artefacts with experimental data.

Connectionists claim to work with more realistic 'real-world' data than the classical artificial intelligence community. However, the connectionist community very seldom, if at all, uses longitudinal developmental data for their simulations. Longitudinal studies emphasise *growth*. A good exemplar of this data are the various child language development studies and the associated corpora referred to in this thesis. Learning is a continuous process and growth distinguishes learning from other cognitive activities. And, if connectionists are serious about simulating learning in general and serious in their claims about using 'real-world' data, then the use of data gathered in longitudinal studies becomes essential in connectionist simulations. We believe, that we have simulated the growth of concepts, growth of vocabulary, growth of ostensive naming, growth of conceptual relations, growth of semantic relations, and growth of word order. This simulation of growth is psychologically plausible because no child is born with a complete repertoire of words, concepts, conceptual relations and semantic relations, or at least the child does not show any signs of having such a repertoire in his or her early infancy (c. 0-9 months). One can quote child language researchers to substantiate this claim about the child's repertoire of language or seek the opinions of caretakers of the child. (Note, that the non-connectionist simulations
of child language development reported in the literature also do not refer to any extant corpora of child language development cf. Section 1.4).

Another interesting question that one can ask about the real-world nature of connectionist simulations of language learning is related to the theoretical assumptions that underpin such simulations. The child language literature includes discussion about the primacy of cognition over language in early infancy (c.0-36 months). Such discussions do provide a theoretical framework in which one can operationalise and evaluate the efficacy of such claims in child development literature. Again, the connectionist community appear to side-step such questions.

Connectionist literature contains some profound observations on learning: the distinction made between supervised, graded and unsupervised learning is profound in that such distinctions, we believe, mirror distinctions made in psycholinguistics between behaviouristic learning of language and the innate ability of humans to learn language. But when one looks at the reported simulations of learning in general no reference is made to this very fundamental distinction in approaches to learning. In our work we have selected supervised connectionist architectures where it is clear that the environmental/caretaker input is crucial to a certain aspect to child language developments. Specifically, we have used supervised learning algorithms and architectures for successfully simulating the learning of word order and the learning of conceptual relations. Again, we have used unsupervised learning algorithms to simulate those aspects of child language development where the child's own initiative is crucial to learning. Specifically, we have used unsupervised learning algorithms for the learning of words, concepts, ostensive naming and in the simulation of the learning of semantic relations.

Connectionist literature is focused on learning by experience, yet the most popular connectionist architecture, the back-propagation network, is based on fairly simplifying assumptions, almost Skinnerian assumptions, about learning: no learning is possible without a tutor. One can simulate self-motivated learning using the BP architectures but the psychological plausibility of such simulations leaves much to be desired. At least thirteen obvious short-comings of the BP networks were noted earlier when attempting to simulate child language development (cf. Chapter 4). This is not to argue that BP networks
should not be used in child language development simulations, rather that they should be used to simulate those aspects of the developments that do require the presence of a tutor. The advent of hybrid connectionist models, models that include different architectures and learning algorithms, will certainly improve on the existing simulations, and we hope that the hybrid network we have developed will add weight to the current work in hybrid simulations.

The simulations of aspects of child language development reported in this thesis are based on the characteristics of children's early language, that is, by taking into account naturalistic data - data collected from children's spontaneous speech in naturally occurring situations. The connectionist networks used in our work were integrated as a hybrid - a hybrid of two modes of learning, supervised and unsupervised.

The co-operative nature of various cognitive faculties in child language development was simulated by training individual networks for individual tasks and then integrating the inputs and outputs of these different networks by using yet another connectionist network. Thus we were able to simulate two major milestones in child language development - one word language and two-word language. The training data used in our work was taken from extant child language corpora, Brown and Bloom, design of semantic feature vectors was again based on the work of a child linguist, Nelson, and notions of conceptual and semantic relations were derived from Bloom's and Brown's works, respectively. Theoretical underpinnings of our simulation were based on the cognition hypothesis advanced by Cromer, a distinguished Anglo-American child linguist. We were encouraged by the arguments in the connectionist literature that claim that Piagetian notions of assimilation, accommodation and learning have some synergy with key parameters used in connectionist simulations. Indeed, we feel that connectionism does provide a practical framework to evaluate the effectiveness of claims made by Piaget.

The aspects of child language development that were successfully simulated included connectionist networks for: concept memory (cf. 5.2), word lexicon (cf. 5.3), conceptual relations (cf. 5.4), ostensive naming (cf. 5.5), semantic relations (cf. 5.6) and word order (cf. 5.7). The success of our simulation can
be judged in three different ways. First, we did not overtly try to train the networks to either learn the defining features of concepts or to learn the individual features: a Kohonen map trained on Nelson-specified feature vectors automatically learnt these features of the 42 concepts noted by Bloom and also categorised them into distinct categories; similarly, 44 phonemes were learnt through another Kohonen map for simulating the sound of 42 different words stored in a word lexicon.

Second, we have used a connectionist network to inter-relate the behaviour of two networks, each was trained to simulate a very specific aspect of child language. The 42 learnt concepts were ostensibly named by establishing Hebbian connections between the concept memory and word lexicon. Both the memory and the lexicon were simulated by using a Kohonen Map respectively and the two maps were connected together with a Hebbian Connections network that was trained to recognise the relationship between a word and its corresponding concept and vice versa.

Third, when we compared the one-word and two-word outputs with the outputs of infants, the results were very much in agreement. Our choice of input data was plausible, and more 'real-world like', in that we have used data from longitudinal studies of child language development. This plausibility was further reinforced at the output stage of our simulation.

By way of summarising our contribution to the connectionist literature, and possibly to the child language development literature (Abidi and Ahmad, forthcoming), we can argue that a major contribution of this thesis is to show the efficacy of a hybrid connectionist model, a model which helps to simulate child language development in a psychologically and neurologically probable manner.

Our results have encouraged us to introduce a typology of theories of learning (Table 29) in terms of the connectionist simulations that can be used to simulate key points of these theories, and indeed to critically examine the results of these simulations as a vehicle for evaluating the theories:
Table 29: Correspondence between psychological theories of learning and connectionist learning networks

<table>
<thead>
<tr>
<th>Psychological Theory of Learning</th>
<th>Possible Connectionist Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unsupervised Learning</strong></td>
<td></td>
</tr>
<tr>
<td>Competitive &amp; Co-operative Learning (Marr, D.)</td>
<td>Competitive and Co-operative Networks</td>
</tr>
<tr>
<td>Semantic Feature Based Learning (Nelson, K.; Clark, E.)</td>
<td>Kohonen Maps</td>
</tr>
<tr>
<td>Attention &amp; Reinforcement (Pavlov, I.)</td>
<td>Hebbian Learning Networks or Grossberg Networks</td>
</tr>
<tr>
<td>Behaviouristic Learning (Skinner, B.)</td>
<td>Hebbian Learning Networks or Grossberg Networks</td>
</tr>
<tr>
<td><strong>Supervised Learning</strong></td>
<td></td>
</tr>
<tr>
<td>Trial-and-Error Learning (Thorndike)</td>
<td>Backpropagation Networks</td>
</tr>
<tr>
<td>Paired - Associated Learning</td>
<td>Pattern Associators</td>
</tr>
<tr>
<td>Probabilistic Learning</td>
<td>Hopfield Networks</td>
</tr>
<tr>
<td></td>
<td>Boltzmann Machines</td>
</tr>
<tr>
<td><strong>Unsupervised and Supervised Learning - Hybrids</strong></td>
<td></td>
</tr>
<tr>
<td>Assimilation &amp; Accommodation (Piaget, J.)</td>
<td>Backpropagation Networks, Kohonen Maps</td>
</tr>
<tr>
<td>Contrastive Learning (Clark, E.)</td>
<td>Backpropagation Networks, Kohonen Maps</td>
</tr>
<tr>
<td>Rote Learning</td>
<td>Associative Memories, Backpropagation Networks</td>
</tr>
<tr>
<td>Learning by Categorisation</td>
<td>Kohonen Feature Maps, Backpropagation Networks</td>
</tr>
</tbody>
</table>

6.2. Future Work

The connectionist methodology for simulating aspects of human cognition, particularly language development, can be used for answering questions in two distinct yet overlapping disciplines. First, connectionist simulations can be used to evaluate psychological hypothesis about the 'human cognitive system' and its development (Bechtal, 1991: 262). Second, connectionist simulations can be used to study the neuro biological aspects of human brain (Levine, 1991), addressing issues such as physical neural structures, specialisation of the brain regions, plasticity and learning. But complex connectionist architectures would be required to even simulate the simplest and most well-observed psychological tasks and neural processes. Such an undertaking would be outside the limits of expertise of experts in cognitive and neural sciences and that is where the notion of a connectionist workbench has its appeal.

A connectionist workbench is essentially a suite of well-integrated connectionist programs. Each program allows its end users to choose appropriate connectionist network(s) to simulate a specific cognitive or perceptual task and then allows the users to inter-link these individual networks to simulate a
complex cognitive or perceptual task. The author is currently involved in the specification and prototyping of such a connectionist workbench. The user-surface of such a connectionist workbench is shown in Figure 40.

![Connectionist Workbench - the user surface](image)

The connectionist workbench being currently developed at the University of Surrey was developed in response not only to provide a workbench to the scholars of child language development but also to other researchers involved in studying language disorder.

**Child language development and the workbench**

Once the workbench is successfully implemented, it would be possible conduct more extensive experiments on language development. There is a wealth of child language data available through the CHILDES database (Child Language Data Exchange System, MacWhinney 1991) and the simulation of such data will provide a quantitative commentary on the data and indeed on the theories that have been used to explain such data.

The workbench can also be used to simulate data related to reading acquisition amongst older children and young adults (Laxon, et al., 1991, 1992 & 1994) where a number of factors that influence reading,
comprehension and writing abilities can be simulated as individual connectionist networks in a manner that is more plausibly than what is reported in the literature. For instance, much of the simulations carried out in these studies is based on the use of IAC networks, networks that little or no learning potential (McClelland & Rumelhart, 1981; Dell, 1986; Brown & Watson, 1991).

Lesion Studies - Simulation of Developmental and Acquired Disabilities

Due to the immense complexity of the human brain, neurologists regard disabilities acquired during development (developmental disabilities) or due to damage to certain brain areas (acquired disabilities) as a means for gaining insight into the organisation and functioning of the various components of the human cognitive system. Generally, a selective impairment in an ability or degradation of performance implies a lesion in the component of the cognitive system dedicated to that ability. The performance of a lesioned cognitive system is then expected to provide useful information about its components, their organisation, function and relationship with other components.

Neurolinguists and neurologists have argued that, brain lesions or abnormal brain development can effect linguistic and cognitive performance, particularly if the lesions (and abnormal growth) are found to be in or around the areas of the brain responsible for 'language' and 'cognitive activities'. Human subjects possessing mental disabilities exhibit particular ability profiles which do not correspond to normal behaviour. These ability profiles of disabled subjects are used by researchers to establish a correlation between the damaged brain area and its function. For example, lesion to the brain site - left tertiary pariento - occipital zones, results in the loss of the selective naming ability. The language of patients with lesions to the above site contains frequent appearances of irrelevant words that resemble the required word in morphology, meaning or phonemic composition. Such disabilities provide neurolinguists a basis to conclude that lesion to the left tertiary pariento - occipital zones causes the removal of inhibitory constraints which are employed in a competitive process of selection of words (Luria, 1970).

Connectionist simulation of language disorder are gaining currency (Hinton and Shallice, 1989). Much like language learning the use of fairly simplistic architectures, like the IAC networks, has prompted
neuropathologists to claim an inter-relationship between synaptic decay times and the correlation between formal and semantic errors in aphasic patients (Martin and Saffran, 1992). Any language disorder is a result of interactions, or lack of interactions, between the language producing areas of the human nervous system. A plausible model of language disorder would require an interaction between various modules - or networks in a connectionist simulation. The connectionist workbench proposed above will be used by Wright (personal communication) in order to simulate aphasia, dyslexia and related disorders. A brief summary of how a workbench, comprising a number of different network architectures can be used to simulate language disorders has been presented in Wright and Ahmad (1994).

Connectionism provides a plausible computational framework for simulating acquired and developmental disabilities. The complete simulation may involve two phases. In the first phase a connectionist network need to be designed and trained to produce unimpaired performance. Design of the connectionist network should be constrained by available neurobiological information. In the second phase, the connectionist network is to be lesioned by two distinct mechanisms; (1) systematically removing some connections or units or (2) adding noise to the connections. The performance of the connectionist network after such artificial lesions can suggest information about the constraints encoded in the connections, and the correlation of the lesioned site to the network’s performance.

**Visual cognition and language development**

In much of the discussion of child language development the role of object permanance is regarded as crucial to the child’s ability to make abstractions about the world around him or her. Visual cognition should play a key role in the development of object permanance. Child language literature regards object permanance as a kind of linguistic task, where phonemes are related to semantic primitives avoiding all discussions of icons that may have played some role in the development of the child’s cognitive faculties. We have not addressed this question directly either, but indirectly, through our discussions on semantic feature vectors of concepts we have attempted to incorporate symbolically what in effect must be treated much more carefully by drawing upon visual cognition in general and the connectionist simulation of visual cognition in particular.
Visual cognition is a term used to refer to those 'mental' processes which occur in the human brain when visual events are recognised, categorised and interpreted. These capabilities for recognition and interpretation rest upon established processing facilities and accessible stores of information concerning essential semantic properties of words and objects. (Seymour, 1979: 7). Pinker has emphasised the relevance of computer systems in elucidating theoretical and empirical aspects of visual cognition: 'After a period of relative stagnation, researchers in visuospatial cognition are striving to synthesise a large number of new empirical findings, theoretical constructs, and external constraints. [...] It has also witnessed a burgeoning of experimental data on imagery and recognition made possible by [...] the development of explicit computations and neural models of processes and structures that were previously characterised only in vague metaphors' (Pinker, 1984: 56).

Future research in terms of a connectionist simulation could usefully speculate on the relationship between vision and early language, in particular the development of visual cognition and its influence on the origination of concepts and their categorisation. Furthermore, analysis of the visual component of language can lead to the elucidation of certain properties of the underlying neurobiological system. A characteristic of the underlying neurobiological system is to accept a visual symbol and to generate the name of that symbol as speech output - vocal naming of visually presented symbols.

One neurobiological system that can be given a connectionist interpretation is the so-called 'Pictorial Memory' (Seymour, 1979). In a psychological sense pictorial memory is regarded as the neurological system which deals with the execution of common tasks, such as the identification and use of objects, the construction and interpretation of spatial configurations of objects, understanding of the function of objects and their classification into taxonomies and hierarchies and other characteristics. Interaction of the pictorial memory with the verbal system is manifested in tasks which require a pictorial - verbal transformation (the naming of objects or description of scenes), a verbal-pictorial transformation (drawing, construction, or visual search for a verbally designated object). Prominent mental functions associated with the pictorial memory include representation and storage of pictorial input; retrieval of
'semantic codes' (object descriptions) comprising of dimensional attributes such as size, shape, colour and function; and inspection of semantic codes against prototypical 'icons' prior to their storage.

Our connectionist workbench with relationships between connectionism, neurobiology, and psychology could be envisaged as follows (Figure 41):

Figure 41: A Connectionist Workbench: synthesis of psychology and neurobiology through connectionism

We suggest that such a connectionist workbench would provide the following features:

(i) A psychologically and neurobiologically plausible simulation of various aspects of human development, such as language, vision, attention, motor movement and perception.

(ii) Implementation of connectionist models to test various psychological theories against psychological data. This would take into account the relationship between psychological theories and the various psychological tasks adhering to the theory.

(iii) Observe the neurobiological implications of connectionist networks towards the specification of brain regions corresponding to psychological tasks, i.e. role of Broca and Wernicke's regions in speech; (b) implications of neural plasticity deemed responsible for human learning. This can be done by altering the structure of a so-called 'neurologically' plausible connectionist networks; (c) observation of the effects of brain lesions towards various psychological tasks.

(iv) Verify the relevance of various neurobiological structures towards both psychological tasks and theories.

(v) Explore the correspondence between the psychological behaviours of generalisation, categorisation, organisation, feature analysis and others with the 'neurobiological' architecture in which they originate.
(vi) Investigate the interaction of a variety of 'neural subsystems' (Arbib et al., 1987) leading to the explication of a 'high-level' psychological task. This can be achieved by firstly developing connectionist networks simulating simpler tasks and then integrating them in a meaningful manner to simulate a 'high-level' psychological task.

**Postscript**

The notion of using a connectionist workbench to speculate on the relevance of psychological and neurobiological theories is a bold one. But there is no shortage of researchers showing keen and positive interest in promoting the idea of using a connectionist workbench. Consider Bechtal's arguments here: 'As one illustration, there are large data sets in these fields that can be used to suggest and evaluate particular simulations. As a second illustration, development is in part a process of building more complex systems from simpler ones. The particular simple and complex systems of the developing child are ones that will be highly relevant to model, and the characterisations of those systems that already exist in developmental psychology should provide at least some clues as to how to proceed' (1991: 278).

If there is any impact of our work, we believe that it would be to encourage others to exploit all what connectionism offers and to examine theories in cognitive science that have hitherto escaped close and objective examination.
Appendix A

Connectionist Algorithms and Structures

A.1. Introduction
Recall that connectionism is a research discipline that aims to understand the nature of human intelligence by simulating aspects of human behaviour through a collection of idealised neurons. These neurons communicate with each other in such a way that in some cases the behaviour of individual neurons can affect the behaviour of the entire collection. The idealised neurons are provided with a variety of 'stimuli' and are expected to 'respond' in a manner that mimics aspects of human behaviour.

Learning in connectionist networks is effected through changes in the strength of connections between individual processing units in a connectionist network. Put simply, given a set of inputs \( x_1, \ldots, x_n \) (a vector symbolically denoted as \( X \)) to a system, the system generates a set of outputs, \( y_1, \ldots, y_s \) denoted symbolically as (a vector) \( Y \). This is achieved computationally by relating \( X \) and \( Y \) through a matrix of connection weights \( w_{11}, \ldots, w_{ns} \) denoted as \( W \). This interrelationship matrix assumes that one, some or all the inputs influence individual outputs (see Figure):

\[
    y_i = w_{i1}x_1 + w_{i2}x_2 + \ldots + w_{in}x_n;
\]
\[
    y_j = w_{j1}x_1 + w_{j2}x_2 + \ldots + w_{jn}x_n;
\]

(and similarly for \( y_2, \ldots, y_s \)).

Learning in the above simplification is then the change of the weights \( W \).

Figure 42: A connectionist network, showing processing units and connections

Connectionist networks are broadly classified on the basis if their learning algorithm which could either be supervised or unsupervised. Below we discuss prominent supervised and unsupervised connectionist networks.

A.2. Unsupervised Learning Connectionist Networks
Learning in an unsupervised manner is characterised by the fact that such learning does not rely on the feedback of an external 'teacher' (as is the case with supervised learning) verifying the goodness of
learning. Rather, unsupervised connectionist networks learn on their own without any explicit supervision; a type of learning which is seemingly more akin to some aspects of exploratory and spontaneous learning observed in a developing child.

A.2.1. Kohonen Maps
Trevo Kohonen (1984) has introduced a connectionist unsupervised learning architecture - the so-called Kohonen maps, based on the theory of self-organising feature maps. Kohonen regarded the self-organising feature mapping theory as both a method of organising complex knowledge and a model of learning in biological systems.

Structure: A Kohonen map consists of two distinct layers of processing units: an input layer and an output layer also called the competitive layer. The input layer is used to present an input vector to the Kohonen map and is n-dimensional, i.e., it consists of n number of units, where n is the number of features in the (n dimensional) input vector X, where X = [x₁, x₂, ..., xₙ]. The dimensionality of the input layer implies that each input unit represents a single feature (or dimension) of the n-dimensional input vector. The output layer, usually referred to as a 'two dimensional map', consists of m number of output units arranged in a two-dimensional format, i.e. rows and columns. The output layer maps the n-dimensional input vector to a lower (two) dimensional representation. Units in the output layer are best understood as competitive units as during learning they compete with each other to represent the learnt knowledge. Both the input and output layers are connected by weighted connections, such that each output layer unit is connected to all input layer units. Associated with each output unit, Oᵢ, is an n-dimensional weight vector W that stores the strength of the connections from the output unit Oᵢ to all units in the input layer. The weight vector W for unit i would be given as Wᵢ = [wᵢ₁, wᵢ₂, ..., wᵢₙ]. It should be noted here that the dimension of the weight vector W of an output unit O always equals the dimension of the input layer. This is because the weight vector of each output unit encodes the n-dimensional information supplied to the output unit by the input layer (see Figure 43).

![Figure 43: A typical Kohonen map connectionist network](image-url)
**Network initialisation:** Before learning, a Kohonen map is initialised to ensure that it does not contain any prior information. Initialisation is achieved by assigning a random weight value in the range of 0-1 to the components of the weight vectors of all competitive units, resulting in random weighted connections between the input and output layers.

**Input presentation:** Learning in Kohonen maps is carried out over a number of iterations. In each iteration, an input pattern is randomly chosen from the ensemble of input patterns and presented to the input layer of the Kohonen map. This random selection of the input patterns ensures that the learning taking place does not observe a pre-determined course and is also not biased in any way.

**Learning algorithm:** Learning in a Kohonen map is based on a process of 'self organisation' which realises a topological mapping of the training (input) patterns. This is achieved by presenting to the Kohonen map's input layer an ensemble of training patterns based on which the connection weights between the input and output units are modified. To realise a topological mapping of the input patterns Kohonen maps incorporate the notion of *topological neighbourhoods* - a region with a dynamically changing boundary (radius) which contains a number of proximate output units that are affected by an input pattern. Each output unit may have a topological neighbourhood $n$, where the size of the neighbourhood is determined by the parameter *neighbourhood radius* (see Figure 44).

![Neighbourhoods for an output unit](image)

**Figure 44:** Three different neighbourhoods for an output unit. The radius determines the size of the neighbourhood.

An important variable used in the Kohonen map's learning algorithm is the so-called Euclidean distance - a kind of goodness of fit between the input and the output pattern. We will refer symbolically to this variable as $ED$ throughout the discussion.

The Kohonen map's learning algorithm comprises the following steps:
Appendix A

i. **Initialize** the Kohonen map to ensure that each output unit has a unique weight vector so that no similar topological regions may initially exist.

ii. **Present** a training (input) pattern \( I \) to the input layer of the Kohonen map which is to be projected on the output units, resulting in each output unit to acquire some activation level based on its weight vector \( W \). In each iteration a different training pattern is randomly chosen from the ensemble of training patterns.

iii. **Determine** the output unit which best matches the training pattern \( I \). This is achieved by firstly calculating for all units the ED between the n-dimensional training pattern \( I \) and the weight vector \( W \) of each unit:

\[
ED_I = \| I - W \|
\]

Next, **select** the output unit \( O \) with the least ED based on the criterion that:

\[
ED_I = \min \{ ED_m \} \quad \text{where} \quad m = \text{all output units}
\]

The output unit with the least ED is regarded as the input pattern's 'image unit'.

iv. **Determine** the image unit's 'neighbourhood' \( N \), i.e., other output units that are in proximity to the image unit.

v. **Make** the 'image unit' and the units in its neighbourhood \( N \) more representative of the input pattern by moving their weight vectors closer to the input pattern. This is to ensure that these units produce an even stronger response to the same input pattern when it is presented again. The weight change is based on the equation below.

\[
\hat{w}_{i,h}(t + 1) = \alpha(t) \left( I^h - w_{i,h}^h \right) \quad \text{if Output unit } O \in N_c
\]

\[
\hat{w}_{i,h}(t + 1) = 0 \quad \text{otherwise}
\]

where \( N \) is the neighbourhood with radius \( c \), alpha is a learning function at time \( t \), \( w_{i,h}^h \) refers to the \( h^\text{th} \) component of weight vector on output unit \( O \), \( I^h \) is the \( h^\text{th} \) component of input pattern \( I \). As learning progresses the weight vector of the image unit moves closer and closer to the input pattern.

vi. **Repeat** steps ii - v for a number of iterations, where in each iteration a different input pattern is randomly chosen from the ensemble of training (input) patterns.

During learning the image unit of a particular training pattern has its ED gradually minimised by the self-organisation process, whereas on a reciprocal basis its activation level is increased. In effect, the Kohonen map's self-organising learning algorithm merely moves the weight vector such that the ED distance between the weight vector and the related input vector is minimised, i.e., the unit's weight vector is closely aligned with the input vector. This results in the Kohonen map incrementally learning the training patterns. At the end of the learning sequence, the training patterns are learnt and represented by the output layer such that each training pattern is represented by a unique output unit and that similar training patterns are placed in proximity. Learning is to be continued until certain learning criteria are satisfied.

**The learning criteria:** In a Kohonen map each learnt concept, feature, word or any other item of knowledge in the input pattern is represented by a unique unit known as its 'image unit'. In connectionist terms, any information, say a concept, is assumed to be learnt when: (a) the activation level of its image unit is the highest amongst all other units and is approaching unity; (b) its ED is minimal, i.e. close to zero, and (c) the image unit's weight vector is very close to the input pattern it is representing.

**Self-organisation process and the emergence of topological neighbourhoods:** The arrangement of similar training patterns in proximity is a direct consequence of the creation of topological neighbourhoods which in turn is due to the self-organisation learning process. During learning a
Topological neighbourhood is created around the image unit by changing the weight of the image unit and all units in its neighbourhood so that they are tuned towards the training pattern. Weight changes in this manner ensure that units in the neighbourhood of an image unit also to a certain degree recognise the same pattern as the image unit. Topological neighbourhoods therefore provide an 'average' representation of a class of the training pattern, where similar training patterns are mapped on proximate output units on the map. At the onset of learning, the neighbourhood is kept large to ensure a global influence of the input patterns, however as learning progresses the neighbourhoods are shrinked so as to restrict the influence of the input to a small number of units, implying a move from generic to specific.

**Dynamics of the self-organisation learning mechanism:** The creation of topological neighbourhoods due to the self-organisation process can be understood as a two-step process. The first step involves a sub-process of 'feature detection' whereby distinguishing features among the input patterns are detected and patterns comprising of similar features are related. Secondly, based on the distinguishing features the n-dimensional input space is segregated into distinct regions or 'topological maps', where each region contains patterns with similar features. This is termed as the realisation of a 'topological mapping' of the input patterns. In the connectionist literature this is regarded as the 'automatic categorisation' of the input data. In this way, learning in Kohonen maps provide a topology preserving mapping of a high-dimensional input space to a low-dimensional output space. This enables one to understand the complexity of a high-dimension input space by viewing it on a two-dimensional graphical map.

Figures 45 a-c show the self-organisation process at work. Beginning with a randomly organised map (the weight vectors of the units are symbolised as arrows pointing in a particular direction) which has the image unit pointing upwards (Figure 45a). The aim of the self-organising learning algorithm is to organise the map through a series of iterations such that units in the immediate neighbourhood of the image unit have arrows pointing in almost the same direction as that of the image unit. After a number of iterations (Figure 4b) the self-organisation process has slightly changed the original direction of the units in the neighbourhood of the image unit towards the image unit's direction. Finally, in Figure 45c the Kohonen map has self-organised such that the arrows of neighbouring units are pointing in the same direction as that of the image unit.
Figure 45: Self-organisation in Kohonen maps

Information retrieval from Kohonen maps: Stored information is retrieved from a Kohonen map by presenting an input pattern and observing the response of all the output units. The output response is the instantiation of activations over all the units. The unit with the highest activation is regarded as the image unit that has the highest activation and therefore best represents the input pattern. Also, the output response smoothly decreases with the distance from the image unit, forming a localised region on the map in which units in the neighbourhoods of the image unit have some activation. Response of units in the neighbourhood N of an image unit is computed by eq.(a)

\[
\text{Act}_{ij} = 1 - \frac{||I - W_{ij}|| - d_{\text{min}}}{d_{\text{max}} - d_{\text{min}}} \quad \text{if} \ (i,j) \in N_c
\]

\[
\text{Act}_{ij} = 0 \quad \text{otherwise} \quad \text{eq (a)}
\]

where \(\text{Act}_{ij}\) is the response of image unit \(o_{ij}\), \(I\) the input vector, \(W_{ij}\) is the weight vector of unit \(o_{ij}\), \(N\) is the neighbourhood with radius \(c\), \(d_{\text{max}}\) is the largest distance of \(I\) to a unit in the neighbourhood, similarly \(d_{\text{min}}\) is the least distance.

A.2.1.1. Neurological plausibility of Kohonen maps

Kohonen (1982), has provided some explanation regarding the possible of embodiment of self-organisation in and the existence of topological maps in the human brain. According to Kohonen, topological maps of structured distributions of signals can be formed in one or two dimensional arrays of processing units (neurons) that did not previously possess such a structure. Kohonen argues that his principle is a generalisation of earlier work done by Willshaw and Malsburg (1976) known as retinotectal mapping, concerning the formation of direct topographic projections between two laminar structures. It is suggested that there exists many kinds of maps or images of sensory experiences in the brain. The
most familiar ones are the retinotopic, somatotopic, and tonotopic projections in the primary sensory areas, as well as the somatotopic order of cells in the motor cortex. According to Kohonen, Lynch (1978) has provided some evidence that topographic maps of the exterior environment are formed in the hippocampus, and Rolls (1989a) has detected structure of cells that selectively respond to face recognition. Kohonen argues that, if the ability of the brain to form maps is established, then one could explain the power of the human brain to operate on 'semantic items', such that some areas of the brain would simply create and order specialised cells in conformity with some high-level features and their combinations.

Researchers such as Martindale (1991) and Massaro (1988) have argued that the brain is composed of a large number of 'modules'. These so-called modules are all constructed in more or less the same way, and comprises several layers, where each layer is made of a large number of nodes. Martindale (1991) argues that each such layer can be imagined as a two-dimensional map. The principle of arrangement on any layer of a module is based on similarity, that is, two nodes representing similar things are closer to each other. It is interesting to note that, the Kohonen map incorporates a similar arrangement principle, which in fact originates from its unsupervised learning mechanism, and results in the categorisation of input patterns. In most cognitive modelling, categorisation of the stored knowledge is desirable, as it helps in efficient knowledge retrieval and also 'generalisation' of novel input. Ritter has suggested that Kohonen maps 'are also directly able to create in an unsupervised process topographical representations of semantic relationships implicit in linguistic data' (1989: 242). In particular, if the data are clustered hierarchically, a very explicit localised representation of the same structure is generated. In this respect, one can argue that Kohonen maps not only provides categorisation of the input, but achieve it in an unsupervised manner through an 'innate' capacity for feature detection. These attributes, along with its neurological plausibility, make Kohonen maps a suitable connectionist network for psychological modelling, particularly for tasks that require unsupervised learning and categorisation.

A.2.1.2. Learning in Kohonen maps: An exemplar simulation

We stated earlier that the Kohonen map connectionist network is capable of learning n-dimensional training (input) patterns in an unsupervised manner. Given a set of input patterns (the so-called environment of the connectionist network), learning also involves establishing relationships between common features in various input patterns which leads to the grouping or the so-called 'categorisation' of similar input patterns.

Consider a simple simulation of the learning of four patterns, which correspond to the concepts - 'dog', 'cat', 'apple' and 'ball'. The concepts to be learnt are represented as a four component input vector, given in Table 30. Note that, for simplicity, the input vectors are arbitrarily created, however, in actuality a pre-defined scheme (cf. Section 5.2.) is always adopted to create input patterns. The Kohonen map used is shown below in Figure 46, comprises nine competitive units, four input units and associated with each
Appendix A

A competitive unit is a four component weight vector. At the end of the training session the Kohonen map would represent or more precisely simulate a small scale concept memory. The notion of concept memory plays a major role in child language development. We have, therefore, simulated a concept memory storing A2 concepts using a larger Kohonen maps comprising 121 output units. But more of that complex map later (cf. Section 5.2), for the present we focus on the simulation of a simple concept memory.

At the start of the simulation, no competitive unit is associated with any particular concept; rather, all competitive units are more or less associated to some extent with all the input patterns. As learning progresses, each concept becomes more and more associated with a particular competitive unit. Therefore, at the end of this exemplar simulation, we expect the four concepts to be represented by four individual competitive units.

### The Learning Profile

The simulation of learning the above mentioned four concepts was conducted by applying the learning mechanism described in Section A.2.1. The learning sequence spanned over 2000 iterations, after which the learning criteria were adequately satisfied.

How the Kohonen map progressively learnt the four concepts can be understood by observing the state of the Kohonen map, i.e. the learning criteria at various stages during learning. We have taken five snapshots of the Kohonen map learning the four concepts, noting the values of the various learning

<table>
<thead>
<tr>
<th>'Concept'</th>
<th>Input Vector</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dog</td>
<td>[1 1 1 0]</td>
</tr>
<tr>
<td>Cat</td>
<td>[1 0 1 0]</td>
</tr>
<tr>
<td>Apple</td>
<td>[0 0 1 1]</td>
</tr>
<tr>
<td>Ball</td>
<td>[0 1 0 1]</td>
</tr>
</tbody>
</table>

Table 30: Input vectors of the four concepts to be learnt
Appendix A

criteria, at regular intervals of 500 iterations. Figure 47 illustrates the minimisation of the ED, and Figure 48 shows the increase in the activation level, with respect to the increasing number of iterations. The reader may note how the ED is minimised as the number of iterations increases, and at the end of the learning sequence the ED is very small, indicating that the concepts have been learnt by the Kohonen map.

Table 31 shows the transformation of the initially random weight vectors to the 'ordered' representation of the input concepts.

<table>
<thead>
<tr>
<th>Concept</th>
<th>Initial Weight Vector</th>
<th>Final Weight Vector</th>
<th>Input Vector</th>
<th>ED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dog (unit 2)</td>
<td>[0.35, 0.93, 0.65, 0.02]</td>
<td>[0.92, 0.90, 0.97, 0.0]</td>
<td>[1, 1, 1, 0]</td>
<td>0.137</td>
</tr>
<tr>
<td>Cat (unit 3)</td>
<td>[0.51, 0.20, 0.9A, 0.20]</td>
<td>[0.9A, 0.09, 0.98, 0.05]</td>
<td>[1, 0, 1, 0]</td>
<td>0.120</td>
</tr>
<tr>
<td>Ball (unit 7)</td>
<td>[0.6A, 0.47, 0.30, 0.61]</td>
<td>[0.07, 0.91, 0.09, 0.9]</td>
<td>[0, 1, 0, 1]</td>
<td>0.157</td>
</tr>
<tr>
<td>Apple (unit 9)</td>
<td>[0.56, 0.70, 0.95, 0.92]</td>
<td>[0.09, 0.10, 0.97, 0.95]</td>
<td>[0, 0, 1, 1]</td>
<td>0.146</td>
</tr>
</tbody>
</table>

At the end of the learning sequence, units 2, 3, 7 and 9 are selected to represent the concepts 'Dog', 'Cat', 'Ball' and 'Apple' respectively (shown later in Figure 52), as these units have the least ED and highest activation level for their associated concepts.

A pictorial profile of the learning of the concept 'Dog', represented by unit 2 in the 'learnt' Kohonen map, is shown in Figure 49. Recall that the value of the ED is a measure of learning achieved after a number of iterations. To show how much a unit is associated with a particular concept, we have devised an illustration scheme: we project the image of the concept (a picture of the concept) on the concerned unit and show the degree of association by hiding or greying out a portion of the image. This implies that the
lower the ED, the less the concept's image is greyed out, hence the concept is more visible, and therefore suggesting a strong association with one unit. In Figure 49 the reader may note that at iteration 1 the ED is quite large therefore resulting in a major portion of 'dog' associated with unit 2 to be hidden. At the end of the simulation, i.e. at iteration 2000, the ED is reasonably low enabling 'dog' to be completely visible, hence implying that the concept 'dog' has now been learnt.

Recall that at the initialisation step in a kohonen map learning, each unit is given a random weight vector. Now in our example there are four concepts to be learnt, cat, dog, ball and apple. It is convenient, perhaps to imagine that each output unit has four quadrants (see Figure 50): one quadrant associated with each of the input concepts with random weighting.

Each unit's degree of association with all four concepts is illustrated by greying out a portion of the concept's image corresponding to the value of the ED given in Figure 47. Again, a concept is deemed to be learnt by a unit only when its complete picture is visible. Whereas a partially visible concept does not account for a unique and strong association with the unit, rather this just indicates the extent to which the unit is partially biased towards the concept. Learning the concepts can then be illustrated by showing the initial random memory being transformed into an ordered concept memory after 2000 iterations, where the four concepts 'Dog', 'Cat', Ball' and 'Apple' are represented by individual units numbered 2, 3, 7 and 9, respectively.

1 concept is deemed to be learnt by a unit only when its complete picture is visible. Whereas a partially visible concept does not account for a unique and strong association with the unit, rather this just indicates the extent to which the unit is partially biased towards the concept. Therefore one can determine the relationship among close concepts, as each concept's image unit shows a distinguishable image of other related concepts.
Figure 51: A pictorial illustration of the Kohonen map before learning. It may be noticed that no unit is explicitly representing a concept.

Figure 52: A pictorial illustration of the Kohonen map after learning. The four concepts - 'Dog', 'Cat', 'Ball' and 'Apple' are represented by units numbered 2, 3, 7 and 9 respectively.

It may be noted that during learning the Kohonen map has 'automatically' determined an 'implicit' category structure for the learnt concepts, whereby similar concepts are stored in proximity. Since the concepts 'Dog' and 'Cat' are alike in that they share quite a number of features, they are stored in proximity (units 2 and 3). Similarly, one may argue that the concepts 'Apple' and 'Ball' share a common shape, and therefore they also are stored relatively near to each other (units 7 and 9). This phenomenon is regarded in connectionist literature as the 'automatic categorisation' of the input data.
A.2.2. Hebbian Learning Rule

Hebb proposed the earliest and one of the simplest learning rule for creating a connection between two units, based on the biological notion that in the nervous system learning occurs by strengthening the connections between two neurons, provided they are active at the same time. Quinlan (1991: A-5) reports that Hebb’s ideas originated from his work on explaining human memories in terms of structural and metabolic changes at inter-cellular junctions, i.e., synapses. Typically, a synapse exists when a process emanating from one cell gets connected with another cell: It was the growth of the synapse, in some biological experiments set up to investigate the neurochemical changes during some kinds of learning which Hebb regarded as the fundamental neuropsychological process underlying memory. Hebb further described how neuronal changes are governed by simultaneous and sequential associations. The central idea was that if two neurons are simultaneously active then over time activity in one neuron will cause some activity in the other neuron. This can happen when the two neurons share a synapse, therefore allowing one neuron to map its activity onto the other, leading to the development of an association between the two neurons. In neuropsychological terms Hebb’s rule posits that ‘when two adjacent neurons are repeatedly active then contingent metabolic changes lead to a lowered synaptic resistance between the two cells. This in turn increases the probability that activity in one cell will cause activity in the other’ (Quinlan, 1991: A). Thus, in a connectionist network based on Hebbian learning rule, the connection strength (the weight) between two units can be increased or decreased in proportion to the product of their activations. The Hebbian learning rule specifies this function as:

\[ \Delta w_{ij} = \lambda r \cdot a_i \cdot a_j \]

where \( \Delta w \) is the change in the weight of the connection between two units \( i \) and \( j \), \( \lambda \) is a constant specifying the rate of learning, \( a_i \) & \( a_j \) are the activations units \( i \) & \( j \): Accordingly the change in weight between two units is proportional to the current activation of the two units together with the value of the learning rate. There are two points of note here: First, a change in connection weights happens only when both units are ON, i.e., their activation level is above/below zero. This implies that the rule is correlational in nature because weight changes only accrue on connections between co-active units. In this sense Hebbian connections have an associative characteristic; associating two active units in an unsupervised manner, where the degree of association (connection weight) is internally determined by the learning mechanism. Second, the weight changes are sensitive to the size of the learning rate: bigger weight changes are possible with larger learning rates. Although, for practical purposes larger learning rates are not advocated as then the weights change dramatically and the learning is not gradual, thus not able to incorporate the subtle attributes of connectionist learning. Even when both units are ON, the weight of the connection between them can decrease if both units have opposite signs, i.e., one is in an excitatory (+) mode whilst the other is inhibitory (-). The connections originating from Hebbian learning rule are termed as 'Hebbian Connections'. Essentially, Hebbian connections are useful in simulating an association between two units, such that two highly active units would be connected by a strong Hebbian connection. The Hebbian connections could then be used to spread the activation of one unit to the other connected unit: this can be achieved by utilising the spreading activation mechanism (cf. Section A.2.A).
A.2.2.1. An Exemplar Simulation of Hebbian Learning and Hebbian Connections

In terms of cognitive modelling, a simulation of the lexicalisation of concepts, i.e., labelling a concept with a word can serve as an example for demonstrating the efficacy of Hebbian learning. Given that both a concept and a word are represented by individual units in separate connectionist networks, concept lexicalisation is simulated by associating a concept unit with its corresponding word unit. In connectionist terms this involves the creation or 'learning' of a Hebbian connection between a concept and its lexical label (word) in an unsupervised manner.

For this simulation we can use the concept memory implemented through a Kohonen map, storing the four concepts dog, cat, apple and ball, an equivalent word lexicon storing the lexical labels (words) for the above mentioned concepts. The word lexicon (shown in Figure 53) has a similar structure as that of the concept memory and is also developed in a similar manner. The simulation of concept lexicalisation is carried out in the following manner:

1. Simultaneously present a concept to the concept memory and its corresponding label to the word lexicon (Figure 53). In the below figures we show the lexicalisation of the concept ‘ball’. The pattern representing the concept ‘ball’ in the concept memory is [0 1 0 1] and the word ‘ball’ in the word lexicon is [1 0 1 1].

The presentation of a pattern to a Kohonen map results in the retrieval of learnt knowledge corresponding to the pattern presented. In response to the stimuli of respective input patterns to the concept memory and word lexicon, the concept unit ‘ball’ and the word unit ‘ball’ acquire the highest activation level amongst all other units in their respective Kohonen maps, and are deemed to be retrieved. In response to the input pattern ‘ball’, units other the concept/word unit ‘ball’ also acquire some degree of activation (see Figure 53).

![Figure 53: State of the concept memory and word lexicon after the presentation of corresponding input patterns to each connectionist network. Activation levels are illustrated in terms of shades of grey; the higher the activation level the darker the shade of grey. Accordingly, both the ‘ball’ concept and word units have the highest activation level.](image)
2. Next, we create an association among all units in the concept memory with all units in the word lexicon. The learning mechanism takes one concept unit at a time and associates it with all units in the word lexicon. The strength of the Hebbian connection would depend on the activation level of the concept unit and the connecting word unit. In this manner the learning mechanism creates associations of varying strengths between all concept and word units, and incidentally since the concept and word unit 'ball' have the highest activation level the strongest Hebbian connection is learnt between them. It should be noted that the learning mechanism was not explicitly instructed to associate any two specific units, on the contrary the association between the concept unit 'ball' and the word unit 'ball' was learnt in an unsupervised manner. Since we have 9 units each in the concept memory and the word lexicon, a total number of $9 \times 9 = 81$ Hebbian connections would be created to associate the concept memory with the word lexicon.

3. Steps 1 and 2 are to be repeated for some number of iterations, where in each iteration a concept-word pair is to be presented to the concept memory and the word lexicon. According to the above equation, the strength of the Hebbian connections are incremented as concept-word pairs are repeatedly presented. Concept lexicalisation is achieved after a defined number of iterations, and is marked by the existence of a strong connection between each concept and its corresponding lexical label - word.

We would like to point out that the above simulation of concept lexicalisation is very simple as compared to the actual simulation performed in Chapter 5.

**A.2.3. Additive Grossberg Network**

The additive Grossberg network (AGN) is a single-layer connectionist network that consists of feature sensitive units having two types of connections, a positive feedback connection with themselves and negative lateral connections with other units in the network. The AGN learns analog patterns using the Hebbian learning algorithm in a 'competitive' learning manner. In a network of feature sensitive units competitive learning means that a number of units is comparing the same input patterns with their internal parameters, and the unit with the best match, the 'winner', is then tuning itself to that input. In this way different units learn different aspects from their input, and the learning can be regarded as a very simple form of abstraction. Figure 54 shows the structure of an AGN.

![Figure 54: Structure of an additive Grossberg network](image-url)
AGN are popularly used as 'memory', storing a one-to-one association between two units. This association learning is accomplished by the below equation which is a slight modification of the basic Hebbian learning rule.

$$W_{ij} = -\alpha W_{ij} + \beta S(a_i) S(a_j)$$

where $W_{ij}$ is the symmetric ($W_{ij} = W_{ji}$) connection strength from the $i^{th}$ to the $j^{th}$ unit, and Grossberg refers to them as 'long term memory'. $S()$ is a sigmoid function, alpha and beta are positive constants.

Information recall in AGN is a competitive process, implemented through the self-exciting (+) and neighbouring (-) lateral feedback connections for mutual inhibition, as shown in Figure 54. Grossberg (1982) presents a biological interpretation of the competitive process suggesting that, this mutual inhibition, if combined with a non-linear (e.g. sigmoid) output function of the neurons, can lead to the 'fast firing neurons' producing a large inhibition of other neurons, which enhances the differences in the firing rates of the neurons. These notions, are expressed in the below equation, describing the activation of units during recall.

$$a_i = -a_i + \delta \sum_{j=1}^{n} S(a_j) W_{ji} + I_i$$

where $a_i$ and $a_j$ are the activation levels of the $i^{th}$ and $j^{th}$ units, respectively, $I_i$ is the $i^{th}$ input value, $\delta$ is a positive constant controlling lateral (inter-layer) feedback. Grossberg refers to the activation value of the units during recall as the 'short term memory'.

An AGN adopts a localist representation for the various features in the input pattern, where each unit is represented by an unit. In this way, the AGN can act as a composite memory: whenever an input pattern is presented, positive and negative weight adjustments are made to connections so as to indicate which features did occur and which did not occur. In fact, strong connections based on Hebbian learning are made between the units representing the present features. The information about the feature pairs, can be conveniently represented as the weights of the units, stored in a matrix which has the dimension $[n,n]$, where $n$ is the number of features. Probing the matrix with an input vector produces a corresponding output vector. The network can then be used as an association memory, whereby associations between two features are stored. Recognition of the association between features can be achieved by probing the memory with one of the original input vectors to reproduce its associated pattern on the output nodes. The activation of the output unit then determining how well the input and output vectors are associated.

The biological plausibility of the matrix memory $[n,n]$ has been discussed by Rolls (1989b) in terms of the architecture shown in Figure 55. In the matrix memory, the strengths of the synapses between horizontal axons and the vertical dendrites are initially random. Because of these random initial synaptic weights, different input patterns on the horizontal axons will tend to activate different output neurons. The
tendency for each pattern to select or activate different neurons can then be enhanced by providing mutual inhibition between the output neurons, to prevent too many neurons responding to that stimulus. The effect of mutual inhibition is that the next time the same stimulus is presented, one neuron responds more because of its strengthened synapses, more inhibition is pass to other neurons, and there is further modification to produce even greater selectivity.

Figure 55: A matrix for competitive learning in which the input stimuli are presented along the rows (input axons), which make modifiable synapses with the columns (output dendrites).

It therefore appears that, the learning mechanisms similar to AGN may be present in the human brain. The AGN effectively selects different output units to respond to different combinations of the inputs. It thus perform a type of categorisation, in which many and 'complex' input patterns are encoded economically onto a few output lines. With the above mentioned characteristics of AGN, it appears that, AGN's can be helpful for memory oriented tasks, particularly when an association between two features of a pattern is to be learnt in an unsupervised manner. This is the kind of association which seems akin to the semantic relations found in children's language, in fact we would use AGN's to simulate a memory storing semantic relations between two features - conceptual categories.

A.2.4. Spreading Activation Algorithm
The spreading activation process is basically an content addressable memory retrieval mechanism proposed by Rumelhart and McClelland (1986a). This algorithm is very popular amongst linguists, neurobiologists, neuropathologists and cognitive psychologists. The popularity is due to the reasons that the notion of spreading activation, leading to association of concepts and ideas, is very intuitive on the one hand and on the other hand has theoretical grounding in the associationist literature.

Retrieving information using spreading activation works relies on preset or learnt weights. During the spreading activation process the processing units in a connectionist network receive input from connected units and pass an output signal to other units. The input and output signals are 'activation levels' of the processing units, where the activation level is a continuous value. Spreading of activation is needed in

---

2 There is associated with each unit a momentary level of energy or activation known as the "Activation Level" of the unit. The activation level is a real number, and for unit i at time t it is represented as a(t).

212
most connectionist networks to pass the result or effect of one unit to another, as processing is
constrained by the mutual interaction between two units.

Units communicate with each other by passing their activation level to other connected units. Activation
level of a unit continuously increases/decreases as excitatory/inhibitory activations are received by it.
Change in the activation level of a unit depends upon the current activation level of the unit and the
influence of other units (net input) and external input (if present) to the unit. The influence of a unit \( j \) on
another unit \( i \) is the product of unit \( j \)'s output and the weight of the connection between unit \( i \) and unit \( j \),
given by \( w_{ij} \). The net input to unit \( i \) is given by the below equation A:

\[
\text{net}_i = \sum_j w_{ij} \text{output}_j + \text{extinput}_i \quad \text{eq. A}
\]

The index \( j \) ranges over all units in the network with connections to unit \( i \).

For cognitive modelling, spreading activation provides a useful mechanism to identify the relationships
between units in various connectionist networks. The activation of an 'parent' unit, when spread through
the spreading activation mechanism activates other units, where the activation is relative to the strength
of association between the two units. If a threshold is set, then units acquiring an activation level above
the threshold can be regarded to be associated to the parent unit. Since the amount of activation received
by a unit is calculated by taking into account all its excitatory and inhibitory connections, thereby
ensuring that the constraints among the various units is maintained.

A 2.4.1. Exemplar simulation of spreading activation: Retrieval of a concept's lexical
label

The efficacy of the spreading activation mechanism could be best demonstrated by
simulating the retrieval of a concept's lexical label. This would involve the spreading of activation of the
units in the concept memory to units in the word lexicon through the Hebbian connections (concept
lexicalisation connections) connecting the concept memory and the word lexicon.

In psychological parlance, expression of internal concepts through language appears to be an important
first step in child language development. This may involve the presence of a stimulus to the child, which
instigates the retrieval of the corresponding concept from the so-called child's concept memory. Next, in
order to express the retrieved concept the child needs to retrieve the concept's lexical label (word) from
the child's so-called word lexicon. As we understand, the verbal expression of the retrieved lexical label
corresponds to the child's utterance of a concept.

In a connectionist simulation, which emphasise the role of spreading activation, the above sequence of
operations corresponds to the following steps:

1. Presentation of a stimulus, i.e., a pattern corresponding to the concept 'dog' to the concept memory.
b. This results in the retrieval of the concept 'dog', i.e. the 'dog' concept unit acquires the highest activation level amongst all other units in the concept memory.

c. Next, the spreading activation mechanism enables the transmission of the activation of all concept units to the word units via the Hebbian connections implemented between the concept memory and word lexicon. The amount of activation spread and received by each word unit is determined by the spreading activation equation A.

d. Due to the spread of activation from the concept memory to the word lexicon the word unit 'dog' acquires the highest activation level and is thus deemed to be retrieved.

An exemplar simulation show how spreading activation can be used to retrieve a word from a connectionist word lexicon, linked by Hebbian connections to a connectionist concept memory, by presenting a stimulus to the concept memory. Figure 56 a-d are based on two networks (discussed in section A.2.2) that have not only been trained to recognise concepts and words respectively, but units in each of the network is connected, with different strengths to the other via Hebbian connections. The use of spreading activation for retrieving a word related to a concept. shows not only the role of spreading activation but also gives an idea how various connectionist architectures could be combined to achieve a behaviour that is achieved by the joint processing of more than one connectionist network; what connectionist may regard as a hybrid connectionist architecture.

The concept 'dog' is presented to the concept memory in terms of a concept input vector. Figure 56a, shows the start of flow of activation from the concept memory to the word lexicon. Figure 56b and Figure 56c illustrate the notion of 'spread of activation', and furthermore establishes the role of the Hebbian connections in this simulation. In simple terms, 'spread of activation' can be understood as a mechanism to 'spread' or 'transmit' activations between connected units. The amount of activation spread equals the product of the Hebbian connection strength and the activation of active unit, in this case it is the concept memory units. The activation level of each recipient unit, in this case the word lexicon units, equals the sum of all activations received from various units. This can be mathematically expressed as

$$Activation_{WL} = Activation_{CM_{unit}} + Activation_{CM_{unit}} + ... + Activation_{CM_{unit}}$$

Figure 56d, shows the final state of both the connectionist networks. It may be noted that the word unit 'dog' has acquired the highest activation level in response to the activation spread from the concept memory. The are two reasons for this kind of behaviour.

1. The concept unit 'dog' has the highest activation level.
2. The exists a strong connection between the dog's concept and word unit.
Appendix A

Figure 56: An exemplar simulation of concept lexicalisation based on the spreading activation mechanism. We show how activations are spread from the concept memory to the word-lexicon through the weighted Hebbian connections

A.3. Supervised Learning Connectionist Networks

A.3.1. Backpropagation Networks

Rumelhart and McClelland (1986), influenced by the developmental theories of their time, developed a class of connectionist networks - Backpropagation (BP) networks that exemplify such supervised learning. Loosely speaking, the so-called backpropagation algorithm allows the establishment of arbitrary, non-linear relationships between input and output patterns.
BP networks cited in the literature are essentially variants of a 'generic architecture' which comprises a multi-layered topology, with an input layer, an output layer, and an intermediate layer, the so-called 'hidden layer': the input layer is the network's interface with the outside world as it receives the (input) patterns to be learnt, the hidden layer applies some mathematical functions to the input patterns and passes the transformed pattern to the output layer. In order for information to be passed between these layers are interconnected such that the input layer is connected to the hidden layer, which in turn is connected to the output layer. These inter-layer connections are weighted and are modified during learning. Figure 57 shows the architecture of a typical BP network.

![Figure 57: A typical back-propagation network](image)

The backpropagation learning algorithm

The backpropagation learning algorithm can be envisaged as a complex mathematical function which maps an input vector into an output vector. The mapping from input to output is smooth and non-linear (i.e. a graph relating input to output is a continuous arbitrary curve), and is realised by tuning the connection weight parameter. A learning algorithm based on 'error minimisation' tunes the connection weights such that the mapping is the 'best fit' according to some measure of error to a set of training data. Learning in BP networks aims to achieve a balance between the ability of the network to respond correctly to the learnt knowledge (i.e. memorisation) and the ability to give reasonable responses to an input that is similar, but not identical to that learnt by the network (i.e. generalisation).

Network Initialisation: Prior to learning any information, the backpropagation network is initialised to ensure that it does not contain any prior information. A backpropagation network is initialised by assigning a random weight value in the range of $0 - 1$ to all its connections, i.e., connections between the input to hidden layer, and hidden to output layer. Random values for weights are obtained from a random number generator. Initialisation results in random weighted connections between the input-hidden and hidden-output layers.
The Learning Phase: Learning in backpropagation networks is carried out over a number of iterations, where in each iteration an input pattern and the corresponding 'target' (output) pattern is presented to the input and output layer respectively. The backpropagation learning mechanism is based on an 'error-correction' procedure such that the learning algorithm can be divided into two distinct phases - the forward phase and backward phase (a schematic presentation of the learning algorithm is shown in Figure 58).

Forward phase
An input pattern is presented to the input layer and the desired output pattern is presented to the output layer. From the input layer the input pattern is first propagated to the hidden layer and then to the output layer. During this propagation, the input pattern is transformed according to a built-in learning functions, resulting in the generation of an output pattern at the output layer. Next, the generated output pattern is compared with the desired target pattern, so as to compute the discrepancy or 'error' between the target output pattern and the generated output pattern.

The value of the 'error' is an indicator of how much learning has been achieved, as a minimal or low error value indicates that the network is adequately producing the desired target pattern, whereas a high error value reflects that the network has not yet learnt the association between the input and target pattern, and therefore more learning is may be required.

Backward phase
The error computed in the forward phase is propagated backwards, first to the hidden layer and then to the input layer. Based on the error received, the weights of the connections between the input-hidden and hidden-output layers are updated accordingly, such that the BP network acquires, or loosely speaking 'learns', a set of weights which enables it to produce an output pattern which is more closer to a desired target pattern. It may be noted here that due to the distributed nature of knowledge representation in a BP network, each hidden layer unit receives only a portion of the total error value, which is roughly
equivalent to the relative contribution that hidden unit made to the original output. Learning, then, in BP networks can therefore be considered to be based on this backward 'error-minimisation' procedure.

We give below the mathematical description of the BP learning algorithm:
1. Initialise the network
Initialise with random values in the range [+1, -1] all the weights, that is input layer - hidden
layer: matrix $V$, and hidden layer to output layer: matrix $W$. Also initialise the thresholds $T_i$ for
each hidden unit and thresholds $T_j$ for each output unit.

2. Calculate hidden unit activations
Present an input pattern $A_k$ and desired output pattern $C_k$ to the input layer and output layer,
respectively. ($k = 1, 2, \ldots, m$, where $m$ is the total number of training patterns)
Transfer $A_k$'s values to the input units, pass the input unit activations through weight matrix $V$
and calculate the activation $b$, for each hidden unit using equation 1

$$b_i = f \left( \sum_{h=1}^{n} A_h V_{hi} + T_i \right) \quad \text{eq. 1}$$

where $f()$ is the sigmoid threshold function $f(x) = (1 + e^{-x})^{-1}$.

3. Calculate output unit activations
Pass the activation of the hidden units through weight matrix $W$, to calculate the activation $c$ of
the output units.

$$c_j = f \left( \sum_{i=1}^{n} b_i W_{ij} + T_j \right) \quad \text{eq. 2}$$

A. Compute output unit error
Compute the error $d$ between the computed and desired output using equation 3

$$d_j = c_j (1 - c_j) (c_j^{-k} - c_j) \quad \text{eq. 3}$$

for all $j = 1, 2, \ldots, q$, where $d_j$ is the $j^{th}$ output unit's computed error.

5. Compute hidden unit error
Calculate the error of each hidden unit, relative to each $d_j$ with the equation

$$e_i = b_i (1 - b_i) \sum_{j=1}^{q} W_{ij} d_j \quad \text{eq. 4}$$

for all $i = 1, 2, \ldots, p$, where $e_i$ is the $i^{th}$ hidden unit's computed error.

6. Adjust weights and thresholds
Adjust the connections from the hidden layer to output layer by equation

$$\Delta W_{ij} = \alpha b_i d_j \quad \text{eq. 5}$$

for all $i = 1, 2, \ldots, p$, and the $j = 1, 2, \ldots, q$, where $\Delta W_{ij}$ is the amount of change made to the
connection from the $i^{th}$ hidden unit to the $j^{th}$ output unit.

Adjust the output units thresholds

$$\Delta T_j = \alpha d_j \quad \text{eq. 6}$$

Adjust the connections from the input layer to the hidden layer

$$\Delta V_{hi} = \beta A_h e_i \quad \text{eq. 7}$$

Adjust the hidden units thresholds

$$\Delta T_i = \beta e_i \quad \text{eq. 8}$$

7. Repeat step 2
continue the learning sequence until the error correction value $d_j$ is sufficiently low.

Back-Propagation Recall Mechanism
Back-propagation recall accepts an input at the input layer, and passes it through the hidden layer
to produce a response at the output layer. This feedforward operation is in two step, is performed
by equations 1 and 2, given above in the learning algorithm.
A.3.1.1. Significance of BP networks for cognitive modelling

BP networks have been used extensively in the connectionist simulation literature, particularly cognitive modelling, as BP networks are suggested to exhibit human learning profiles such as staged development, flexible constraints, Piagetian developmental mechanisms - assimilation and accommodation. Bechtal and Abrahamsen (1991) argue that the way a BP network iteratively reduces the error in the output signal might model staged development and the use of flexible constraints as mentioned in psychological literature. Also, it appears that the Piagetian notions of assimilation and accommodation can be modelled, and understood as a kind of gradient descent process.

There are problems when one wishes to ascertain the 'biological plausibility' of BP networks, in that there appears to be no direct evidence, particularly in neurobiology of their plausibility. In fact, Grossberg regards BP networks as 'biologically unrealistic' in that the feedback in these networks is not that of 'electrical signals' but of synaptic weights. The studies of human nervous system shows no evidence of such 'weight transport' (Grossberg, 1987). Additionally, Bechtal and Abrahamsen have noted that 'backpropagation does not map directly onto any known biological processes' (1991:97). This is based on the fact that there is no evidence yet in the human nervous system information is passed backwards to adjust the performance of the system. However, Bechtal and Abrahamsen stress that, the efficacy of backpropagation networks should be evaluated not at the neurobiological level but rather at the psychological level. At the psychological level, backpropagation networks can achieve gradient descent, that is, reduce the error in the output in an iterative manner. This attribute qualifies BP networks as appropriate for psychological modelling.

The significance of the BP learning algorithm is that during learning the hidden units get associated with specific features of the input pattern. It so happens that, in order to reach some global optimum, the hidden units become tuned to represent 'generalisations' or 'abstractions' with the input patterns. For this reason, it is tempting to think of the weight vectors created at the hidden layer, i.e., the network's internal representation as a form of abstract representation that retains some idiosyncrasies of specific input patterns. It has recently been demonstrated that the weight vectors created at the hidden layer may converge to values that encode linguistic items according to their semantic roles (Miikkulainen & Dyer, 1988). In fact, we anticipate that this attribute of BP networks to be helpful in the simulation of the learning of word order. Such a behaviour of the hidden units results in the development of 'internal

---

Researchers are still attempting to provide a biological plausibility to BP. One such researcher Stork (1989) has put forward a variant of a BP network that comprises, what he calls 'fundamental', 'transform', 'error' and 'difference' neurons, including 'modulating' and 'non-Hebbian' synapses. Stork's network has a logical NOR unit, that controls weight changes through his collections of neurons and synapses. Introducing these variations Stork developed, what he hoped to be a 'biologically plausible' BP network that does not depend on synaptic weight transport. But instead he had to introduce a number of inactual assumptions regarding synaptic activity (Levine 1991: 216).
Appendix A

representations of the input patterns. This characteristic of BP networks provides, (a) categorisation of the input patterns, and (b) ability to generalise, that is, to give a satisfactory response for a novel or incomplete input. Generalisation is possible if the new input pattern partially contains features that resemble the network's internal representation, resulting in response that is closest to the novel input pattern.

BP networks can then be characterised by their potential to handle complex pattern-matching problems, i.e., learning a pre-defined set of input-output example pairs, recognition of complex patterns which may be both learnt or novel (generalisation effects), performing non-trivial mapping functions and creating internal representations that capture the idiosyncrasies of the learnt knowledge.
Bibliography


Ackerman, B. (1986)


Anglin, J. (1983)


Au, T. (1985)

Ault, R. (1983)


Barrett, M. (1985)
*Children's Single Word Speech.* Chichester: John Wiley & Sons.
Bibliography

Barsalou, L. (1992)


Beilin, H. (1975)


Bjorklund, D. & Harnischfeger, K. (1990)

Block, H., Moulton, J. & Robinson, G. (1975)

Bloom, L. (1993)

Bloom, L. (1973)
*One Word At A Time.* Paris: Mouton.

Bloom, L., Lightbown, P. and Hood, L. (1975)


Bowerman, M. (1973)

Braine, M. (1976)

223
Brown, R. (1973a)

Brown, R. (1973b)


Bruner, J. (1975)

*Inquiries into Child Language.* London: Allyn & Bacon, Inc.

Callanan, M. (1985)


Caudill, M and Butler, C. (1992)
*Understanding Neural Networks: Computer Explorations.* Cambridge (Mass.): MIT Press.

Charles-Luce, J. & Luce, P. (1990)


Clark, A. & Lutz, R. (1992)
*Connectionism in Context.* London: Springer-Verlag.

Clark, E. (1973)

Clark, E. (1973)

Clark, E. (1978)
Clarke, E. & Dewhurst, K. (1972)
*An Illustrated History of Brain Function.* Sandford Publications.

Cromer, R. (1991)


Dalenoort, G. (1985)

Davis, S. (Ed.) (1992)

Dell, G. (1986)

Dell, G. (1988)


Diedrich, J. (1990)

Dromi, E. (1987)
*Early Lexical Development.* Cambridge: Cambridge University Press.

Elman, J. (1989)

Ervin-Tripp, S. (1973)

Feldman, J. (1986)


Flavell, J. (1963)


Gainotti, G. (1990) 

A Learning Algorithm for Connectionist Concept Classes. Int. Joint Conf. on Neural Networks. IEEE Vol. 1.

Gardener, H. (1978) 


Gopnik, A. (1988) 


Gopnik, A. & Meltzoff, A. (1987a) 

Gopnik, A. & Meltzoff, A. (1987b) 

226

Greco, A. (1990)


Gregory, R. (1987)

Grossberg, S. (1978)


*Studies of Mind and Brain.* D. Reidel Publishing Company.

*Neural Networks and Natural Intelligence.* Cambridge: MIT Press.

Harris, J. (1990)

*Connectionist Approaches To Language Disorders.* Aphasiology, Vol. 7, pp. 221 - 249.


Hebb, D. (1949)

Hecht-Nielsen, R. (1990)

Hill, J. (1992)

Hill, J. (1983)


*Parallel Models of Associative Memory.* Hillsdale: Lawrence Erlbaum Associates.

Hintzman, D. (1990)  

Hopfield, J. (1982)  

Hulme, C., Snowling, M. and Quinlan, P. (1991)  


*The Second Year of Life.* Cambridge, MA: Harvard Univ. Press

Kohonen, T. (1990)  

Kohonen, T (1989)  


*Self-Organisation and Associative Memory.* Berlin: Springer-Verlag
Bibliography


Lippmann, R. (1987)
*An Introduction To Computing With Neural Nets*. IEEE ASSP Magazine, April, pp. 4-22.


Luce, J. & Luce, P. (1990)

Luria, A. (1970)
Macario, J. (1991)  
**Young Children's Use of Color in Classification of Foods and Canonically Coloured Objects.** Cognitive Development, Vol. 6, pp. 17-46.


Macnamara, J. (1972)  

Macnamara, J. (1977)  

MacWhinney, B. (1991)  

Mandler, J. (1993)  


Markman, E. (1989)  


**Vision.** San Francisco: W. H. Freeman.

Martin, N. & Saffran, E. (1992)  
**A Computational Account of Deep Dysphasia: Evidence From a Single Case Study.** Brain and Language, Vol. 43, pp. 240-274


McClelland, J. (1989)  
McClelland, J. L. (1988)


Distributed Memory and the Representation of General and Specific Information. J Experimental Psychology: General, 114, (2), 159-88.


McCulloch, W. & Pitts, W. (1943)

McMaster, I., Sampson, J. and King, J. (1976)


Acquisition of a Lexicon. Contemporary Educational Psychology, Vol. 8, pp. 237 -

Building Distributed Representations Without Microfeatures. Technical Report UCLA-AI-87-17, AI Laboratory, Computer Science Department, University of California, Los Angeles, USA


A Modular Neural Network Architecture for Sequential Paraphrasing of Script-Based Stories. The International Joint Conference on Neural Networks, Washington DC


Morikawa, H. (1988)


Nelson, K. (1977)


Nelson, K. (1973)

Nelson, K. (1973)

Nelson, K. (1973)


Nelson, K., Rescorla, L. & Gruendel, J. (1978)

Ney, J. & Pearson, B. (1990)

Palfreyman, N. (1990)

Pavlov, I. (1927)

Piaget, J. (1952)


Quinlan, P. (1991)
Connectionism and Psychology. New York: Harvester Wheatsheaf


Reilly, R. & Sharkey, N. (1992)
Connectionist Approaches To Natural Language Processing. Hove: Lawrence Erlbaum Associates.

Rhodes, G. (1993)

Rips, L. Shoben, E. & Smith, E. (1973)

Rodgon, M. (1976)

Rolls, E. T. (1989a)
Rolls, E. T. (1989b)

Rosch, E. (1973)

Rosch, E. (1978)

Rosch, E., Mervis, C., Gray, W., Johnson, D. & Boyes-Braem, P. (1976)

Rosenblatt, F. (1962)

Ross, G. (1980)

Rumelhart, D., McClelland, J. & The PDP Research Group (1986a)

Rumelhart, D. & McClelland, J. (1986b)

Schlesinger, I. (1971)

Schlesinger, I. (1977)

Schnelle, H. (1971)


Searle, J. (1991)
Seidenberg, M. (1993)  


*NETtalk: A Parallel Network that Learns to Read Aloud.* Technical report JHU/EECS-86/01, The Johns Hopkins Univ., Baltimore.

Selfridge, M. (1986)  

Seymour, P. (1979)  


Shawley, A. & Schopman, J. (1990)  


Simpson, P. (1990)  

Sinclair, J. (1970)  

Bibliography

Siskind, J. (1990)
Acquiring Core Meaning of Words, Represented as Jackendoff-Style Conceptual Structures, From Correlated Streams of Linguistic and Non-linguistic Input. Proceedings of the Twenty-Eighth Annual Meeting of the Association For Computational Linguistics, University of Pittsburgh, Pittsburgh, Penn.

Slobin, D. (1973)

Small, M. (1990)


Smolensky, P. (1988)
On the Proper Treatment of Connectionism. the Behavioural and Brain Sciences, Vol. 11.

Solovik, M. (1990)


Starr, S. (1975)

Is Backpropagation Biologically Plausible. Int. Joint Conf. on Neural Networks. IEEE. Vol. 2.

Sugarman, S. (1983)

Wales, R. (1993)


_A Connectionist Model of Conceptual Representation._ Int. Joint Conf. on Neural Networks. IEEE. Vol. 1.

Wills haw, D. & Malsburg, C. (1976)  

_Artificial Intelligence._ Reading: Addison-Wesley. 1984.

Wright, J. & Ahmad, K. (1994)  