Learning in a Connectionist Network

Language Acquisition (CS784)

Term-Project Report

by

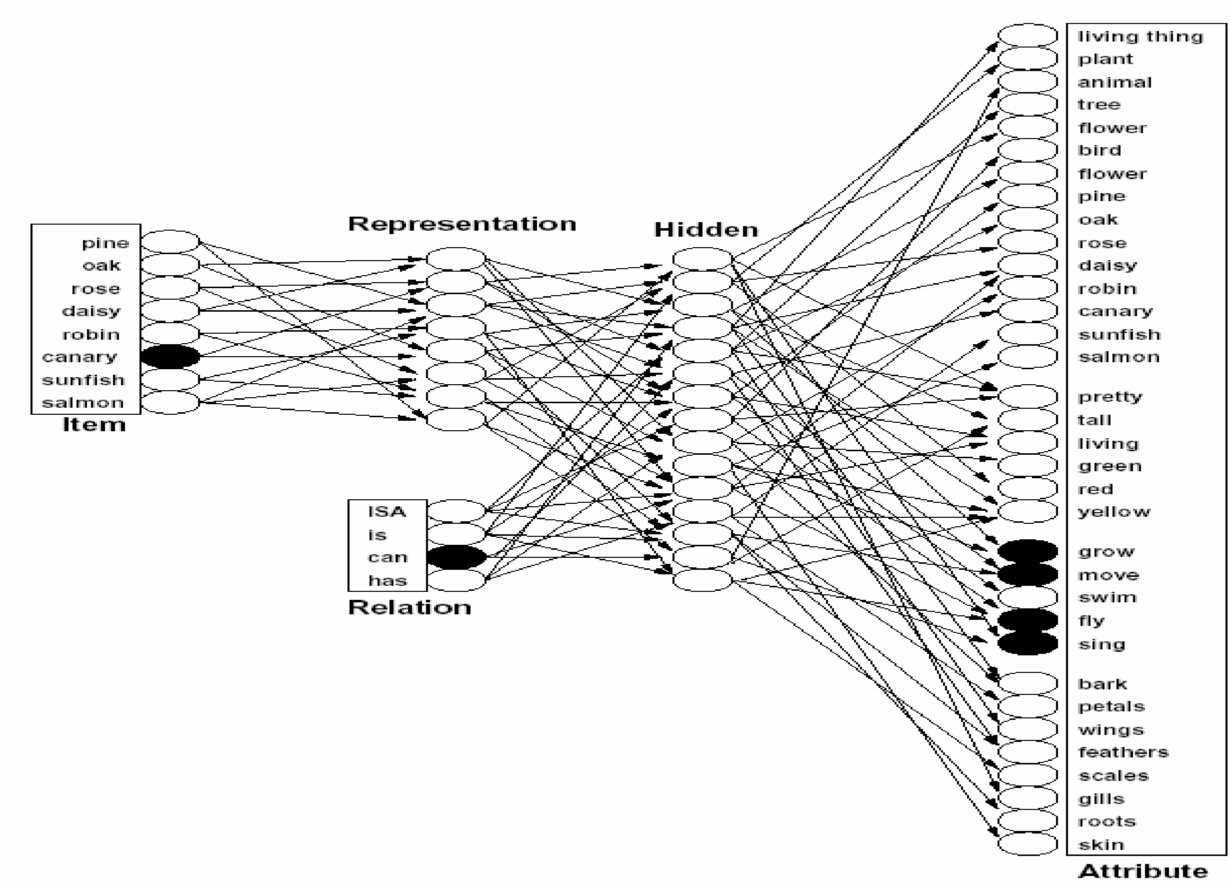
Anshuman Gupta (Y1062)

Satyam Prasad Shaw (Y1327)

Abstract

In this project we tried to model learning in a connectionist framework. We simulated learning in a Feed-forward neural network through backpropagation and tried to compare it with *child language acquisition*.

Introduction



- Figure : Rumelhart Feed-forward network(taken from http://www.cnbc.cmu.edu/ibsc/papers/RogersMcC.pdf)

Our network is based on the Rumelhart Feed-forward network Rumelhart was able to show the following things with his simple Feed-forward network

- Taxonomic hierarchy could also be captured by distributed representation acquired by backpropagation.

- The network could perform inferences that can be Quillian's hierarchial propositional network

We tried to make a general architecture which for some fixed parameters was same as Rumelharts model (for Hidden Layers=2) also we used the data set that was used by Rumelhart.

We tried to find the solutions of the following questions using our network :-

- How can a connectionist framework model child language acquisition?
    - We tried to find answer for this question by giving the network data in a way as it is given to a child.
- How good is the network in generalizing features?
    - This feature actually came from child language acquisition itself .We tried to give data to the network in such a way so as to study generalization of features taking place in the network.
- Is it better in learning an organized data?
    - We tried to make it learn both random data as well as block data. The block data was more organized.
- How consistently the system learns if the learning new representation implies modifying existing representation?
    - We tried to do this by setting a parameter which determines the collision in the representation of new input concepts with existing concepts.
- What if it is trained with data sort of like normal human beings are trained with?
    - Positive Examples
        - These were the predicates which were true for an item and a specific relation.
    - Negative Examples
        - These were the things which were not true for an item and a specific relation.
    - Don't Cares
        - These were the predicates undefined in the given context.

Procedure

The structure of the network was fixed for all experiments

- Activation function was sigmoid
- Threshold of the nodes was 0.9
- 1 Hidden layer with 18 nodes
- Representation of any input activated two nodes.
- 12 inputs,26 outputs for the neural-network

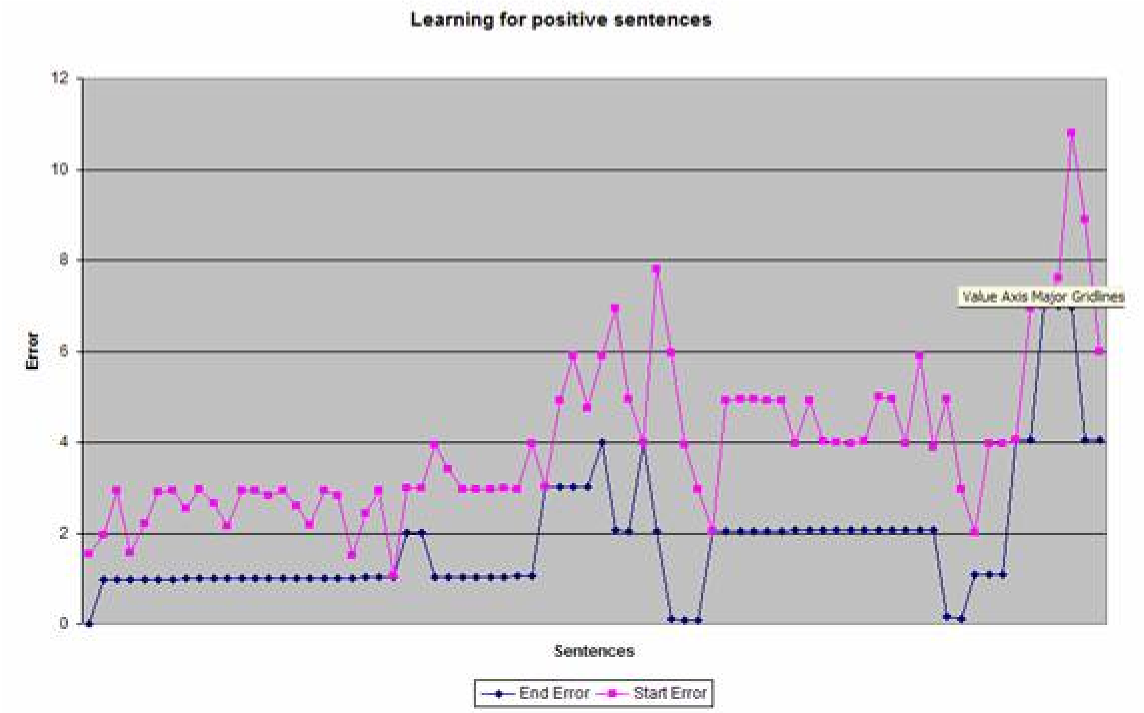
By representation we mean that any Item was represented by the activation of two nodes of the input layer.

Experimental Setup

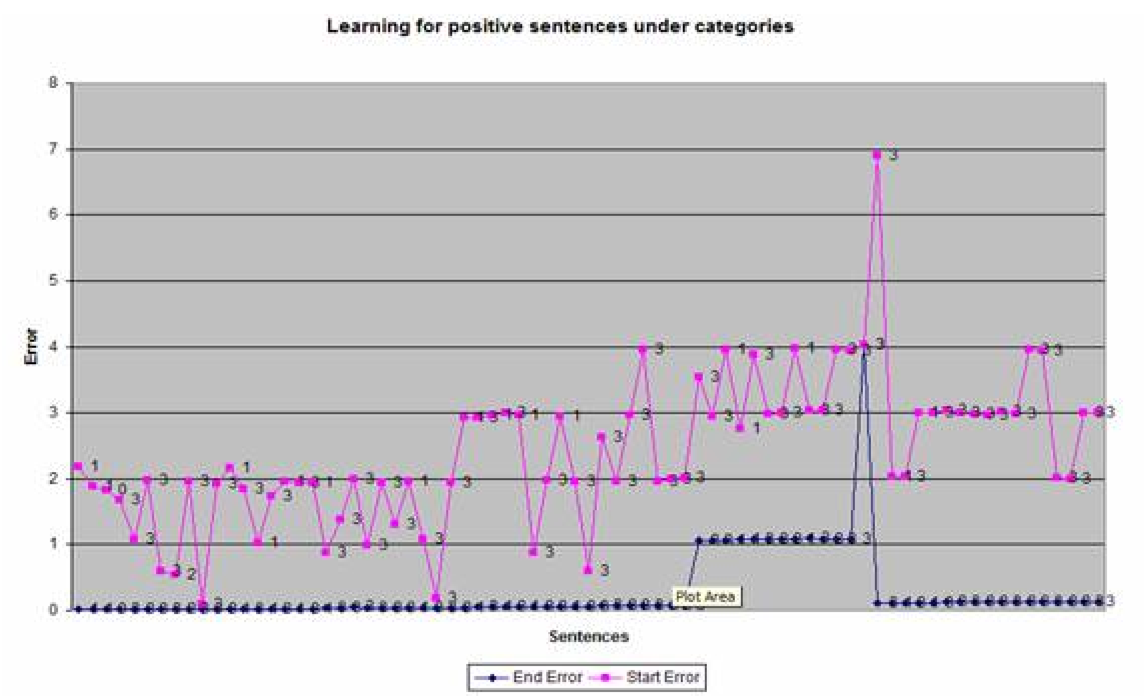
We tried to differentiate the learning in our work based on the following aspects:-

- Random data or block data
- Learning with or without negative examples

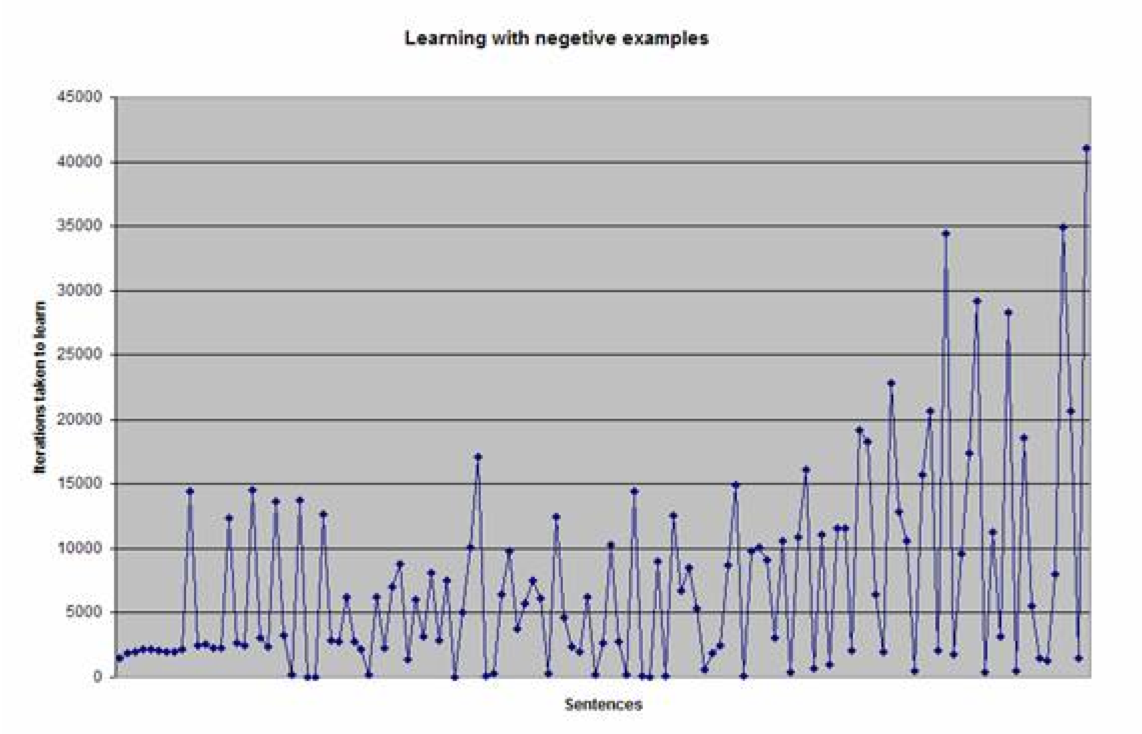
Simulation Results



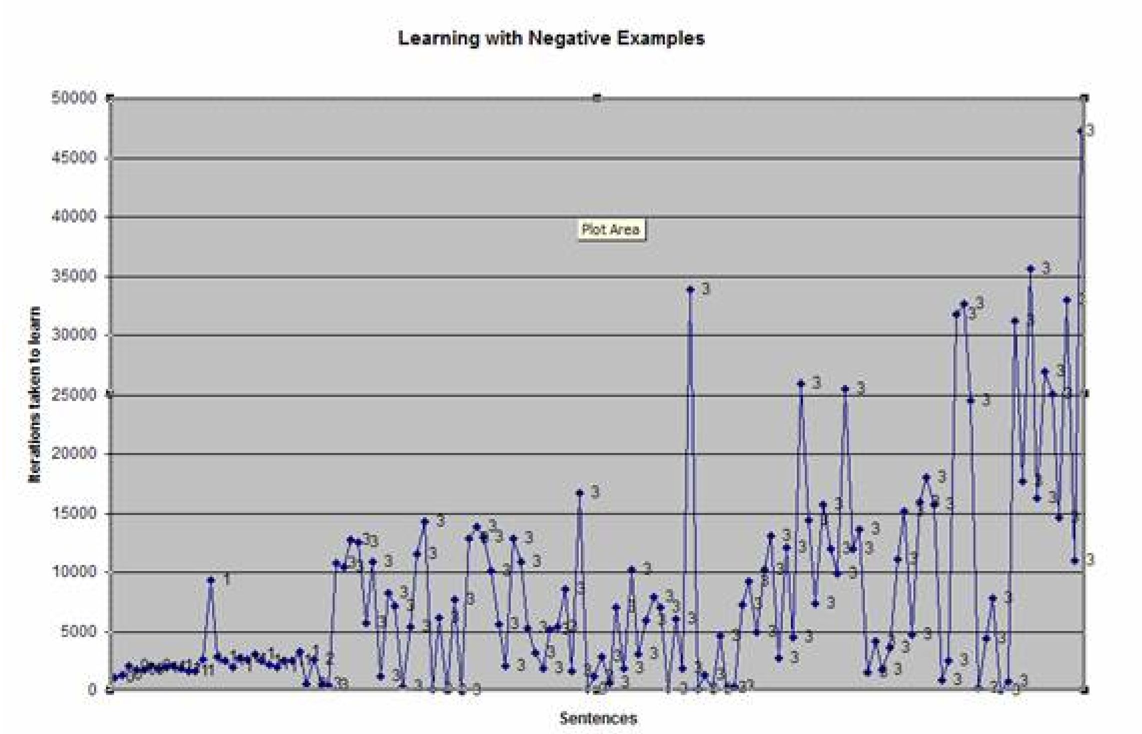
Graph 1



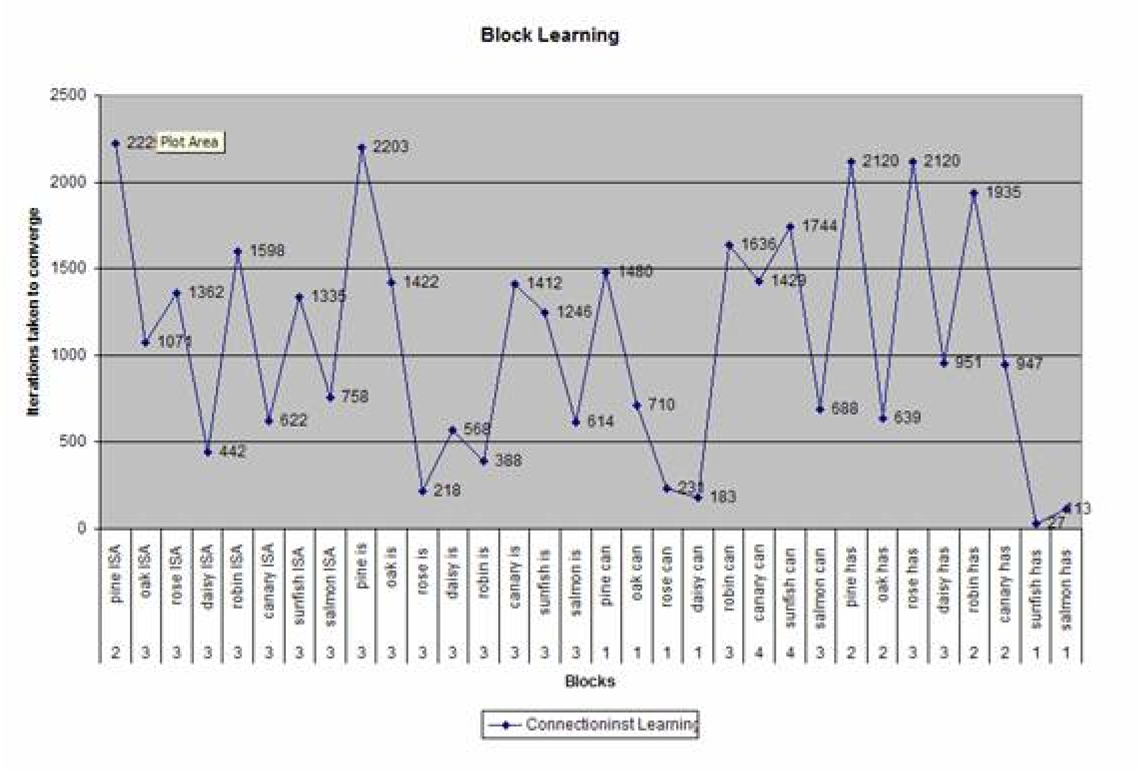
Graph 2 – The numbers on graph show categories



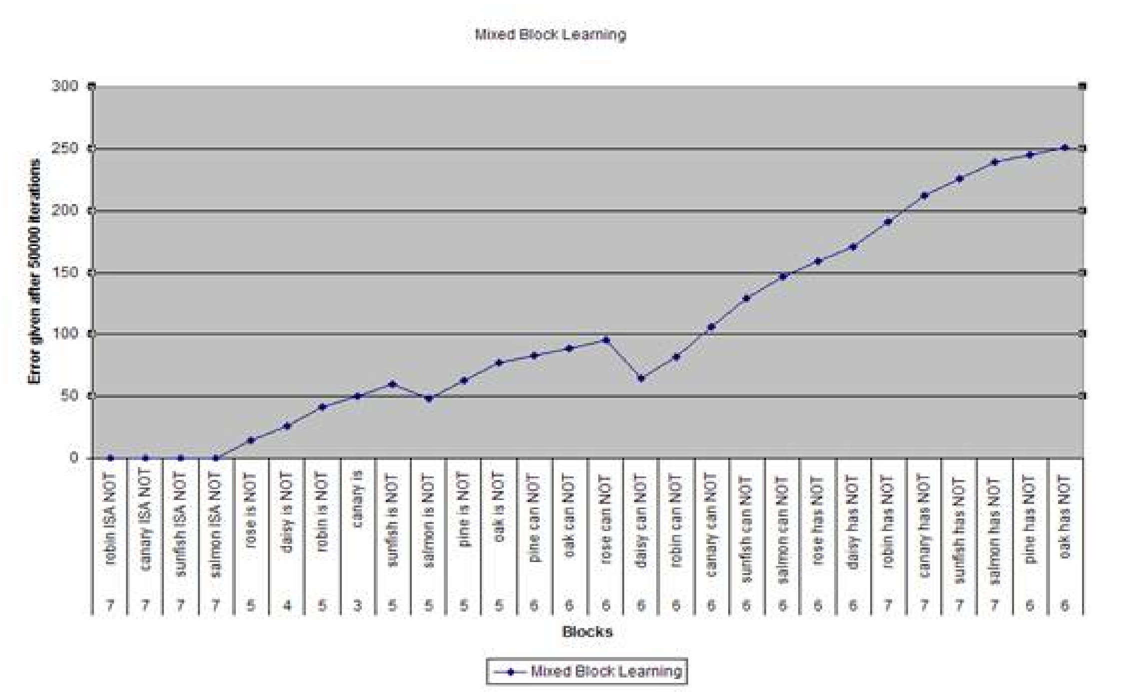
Graph 3



Graph 4 – The numbers on graph show categories



Graph 5 – The bottom line shows size of blocks



Graph 6 – The bottom line shows size of blocks

Observations

Learning with positive examples with single sentence learnt at a time

The first observation is the number of iterations required is huge (>50000). But we need to see the number of iterations required in other cases to comment upon this.

The error size obtained after 50,000 iterations is increasing with increase in Learning Set Size. In other words, it is becoming more and more difficult to learn new concepts and predicates with increasing base knowledge. This is quite unlike what happens in case of child language acquisition, as in case of children with increase of knowledge the ability to grasp new concepts and predicates increases, signifying learning becomes easier.

Another observation that can be made is that the start error is the same as end error in certain cases. This happens when neural net is stuck in a loop. Like for example, in an iteration an edge weight is changed from w_1 to w_2, but in the next iteration it is changed back to w_1 and so on.

Learning with positive examples with single sentence learnt at a time looked category wise

The categorization of the sentences in Learning Set does not reveal much as the error received are similar for all after 50000 iterations. Thus the simulation failed to distinguish between the categories which were –

0 – Both concept and predicated are used for the first time in the Learning Set.
1 – The concept is new but the predicate has already been encountered in the Learning Set.

2 – The predicate is new but the concept has already been encountered in the Learning Set.

3 – Both the predicate and concept have already been introduced.

Learning with positive as well as negative examples with single sentence learnt at a time

The first thing which can be observed is that the number of iterations required to learn have come down as compared to using only the positive examples. This is a clear piece of evidence pointing at the claim that the negative examples serve very important purpose in language acquisition as they make learning much easier and faster.

Highly erratic pattern is observed in the graph for iterations required v/s sentence learned. Nothing conclusive can be said about the learning pattern. But we do notice one thing that with increasing size of learning set, the peaks get higher and higher. This again underlines what was stated in the case of positive examples learnt only. The learning might not be a strictly connectionist learning, in the sense that one might not check the consistency of every sentence known whenever he learns a new sentence, but maintains an abridged structure which is checked and modified. This may take considerably less amount of time as compared to a full check. The modification made to a specific concept's or predicate's knowledge is based on the modification in the abridged representation.

Learning with positive as well as negative examples with single sentence learnt at a time looked category wise

Unlike the case of positive only examples, we can observe here that in case of category 0 and 1, the number of iterations required is roughly the same, while it is highly erratic in the case of category 3. This makes sense as in case of categories 0, 1 and 2 something new is learnt, which makes it equally difficult in all cases. While in the case of category 3, the iterations required depend on whether the predicate being learnt has already been learnt with a concept which is related to the concept being learnt or not.

Directly Related concept – When 2 concepts have been learnt, with same predicate. It is easiest to learn these sentences.

E.g. Daisy has petals ; Rose has petals

Daisy and Rose are directly related.

Indirectly Related concept – When 2 concepts are not Directly Related but there is a another concept which is directly or indirectly related to both. It is slightly difficult to learn and might depend on relation distance i.e. how many concepts occur in the relation chain.

E.g. Daisy has petals ; Rose has petals

Daisy is a plant ; Cotton is a plant

Rose and Cotton are indirectly related

Unrelated concept – When 2 concepts are not related directly or indirectly. These are the most difficult to learn amongst category 3 sentences.

Learning with positive examples with whole block learnt at a time

As can be seen, Block learning required significantly less number of iterations compared to learning one sentence at a time. This gives an indication that it is easier to learn a concept when a whole picture of the thing is presented in one go rather then one sentence about it at a time.

Whenever a new relation is introduced, iterations required are more. This is in line with the expectations as a new relation means a new kind of understanding, which is comparatively more difficult.

There is not a high increase in iterations required with increasing Learning Set Size, unlike single sentence learning, meaning much less modifications required. Thus it strongly supports the claim that batch learning is preferred when teaching a child.

Zig Zag curves are observed in the graphs. The reason is that we are learning related concepts in succession like -

Pine-oak , rose-daisy , robin-canary , sunfish-salmon

Both concepts in the pair have similar predicates associated. So once you learn one, it is easier to learn the other.

Less number of iterations was required for small blocks (true for block size 1, though not visible in case of block size 2). This is obvious as complicated concepts take more effort to be learnt.

Learning with positive as well as negative examples with whole block learnt at a time

There is an increase in the number of iterations required with increasing Learning Set size. The number of iterations is very high. This is due to very large size of Learning Set.

The end error keeps increasing almost monotonically, and so does the start error. We can attribute this to increasing Learning Set Size, which dominates all other features appearing in the previous case. But, this is the first time the increase is so uniform. We have not been able to understand this difference.

Conclusion

Based on the results of experiments we concluded the following –

1. Learning time increases with increasing Learning Set in connectionist system. Since, this is not observed in reality, it might happen that concepts are stored in an abridged manner and this structure is modified during learning. Changes to concepts are made based on changes in this structure.
2. The learning time reduces considerably when negative examples are mixed with positive ones, underlining the requirement of negative examples in Child Language Acquisition.
3. Learning of new Concepts, Predicates in a sentence takes large and similar amounts of time. Though with known concept and known predicate, learning time is reduced only if the concept being learnt and concept connected to the predicate being learnt are related, directly or indirectly.
4. For block learning the efforts required are considerably lesser then in single sentence learning. Thus, pointing out towards the preference for batch learning.
5. Complex concepts difficult to learn compared to simpler ones. Complexity is proportional to the predicates linked with the concept.
6. When one concept has been learnt, another similar concept can be learned very easily with considerably less effort.

Thus, the simulations show preference for batch learning and also for having a mixed set of positive and negative examples for Child Language Acquisition. Also, children can learn simple concepts easily, but have problems in learning complex concepts.

References

http://www.cnbc.cmu.edu/ibsc/papers/RogersMcC.pdf