WWN-Text: Cortex-Like Language Acquisition with “What” and “Where”

Kajal Miyan, Juyang Weng
Department of Computer Science and Engineering
Michigan State University
East Lansing, MI, 48824 USA
Email: {miyankaj, weng} at msu dot edu

***Abstract*—Based on some recent advances in understanding and modeling cortical processing for space [1] and time [2], we propose a developmental, general purpose model for language acquisition using multiple motor areas. Through the ventral pathway, the “what” motor learns, abstracts and projects (as recurrent top-down context) information that is related to the meaning of the text. Via the dorsal pathway, the “where/how” motor learns, abstracts and projects (as top-down context) information that relates to the spatial information of text, e.g., where is the text on a page. This is a major departure from the traditional symbolic and connectionist approaches to natural language processing (NLP) — the nature of the motor areas, i.e., actions, of the developmental agent play the role of “hubs” in language acquisition and understanding. As any human communicable concept can be either verbally stated (what) or demonstrated through actions (how), this model seems to be the first general purpose developmental model for general language acquisition, although the size of experiments is still very small. Furthermore, unlike traditional NLP approaches, we do not use hand-crafted language structure but allow primary and secondary associations as seen in animal learning [3], as a general scheme for language acquisition.**

I. INTRODUCTION

It is known that languages assist in the mental processes. Much of our adult intellect is conveyed, stored and enhanced through natural languages. Natural language not only consists of sounds, symbols, syntax and semantics peculiar to human communication, but more importantly, inside the brain it corresponds to brain-organized traces of sensorimotor experience grounded in the physical world, this could include seeing an object or hearing about it or reading about it etc. as demonstrated in [4]. This perspective is supported by modern studies of language acquisition in developmental psychology. Hence, motor activities play both a preparation and an operational role in language acquisition [5]. It has also been shown that infants appear to use visual and auditory associations inherent in social contexts to learn native-language phonetic categories [6]. However, after 50 years of extensive research in the field of Natural Language Processing (NLP) in Artificial Intelligence (AI), few efforts have been spent on simulating how an agent *acquires* language merely from real-world interactions.

In fact, early acquisition of language in children, seldom includes any explicit training of language rules. Children learn to pay attention to the desirable contexts and carry out unrehearsed conversations with the co-participation of other motors [7]. Despite the lack of explicit syntactic rules, children acquire language skills interactively in ways similar to the acquisition of skills for other sensing modalities; language comes naturally to humans yet such skills are not genetically coded. In his Essay Concerning Human Understanding John Locke introduces the logic of empiricism, he argues that language and ideas are not completely innate but develop from sensorimotor experience and the experience of reasoning [8].

In the work reported in this paper, we use a simplified cortex-like model to simulate the process of language acquisition via incremental, interactive, sensorimotor interactions. For tractability at this stage of system development, we use insulated words as distributed inputs (patterns) and use distributed motor outputs (also patterns). The model is not formally taught grammatical rules governing a language but learns implicit rules on-the-fly through sensorimotor examples. It is also able to create new sentences using its learned reasoning and generalization capabilities.

II. LITERATURE SURVEY

Connectionist approach, developed in 1980s, attempts to model mental and psychological behaviors using networks with numeric, distributed internal representations. These networks have also been used to model distributed language representation. Unlike symbolic approaches, representations in such networks are emergent. These models have two main motivations. First, there is a need for parallel processing of knowledge from multiple sources in a systematic way without specifying or knowing which input component represents what meaning. Second, since the model is not symbolic the representation itself has a potential to tolerate noisy inputs, irregularities and “fuzziness” of real natural language.

Hinton, 1981, published some seminal work on distributed semantic representations [9]. McClelland and Kawamoto, 1986, [10] used distributed representation and semantic microfeatures to address the problem of case role assignment. Other early related studies that use networks include Hanson and Kegl [11] for syntactic parsing, Allen [12] on question and answering, Sharkey [13] on prepositional attachment, Lange and Dyer [14] on inference, Smolensky [15] on variable binding etc. More recently, recurrent neural networks like Elman network [16] and Jordan network [17] use temporal states in models with context units. ARTMAP [18] was based

978-1-4244-6901-7/10/$26.00 ©2010 IEEE

on the concept of similarity measure for symbolic objects and can assign class labels to the objects.

Along with all this, in recent times a lot of work has been done in the direction of binding physical grounding with language. Studies and models like that of Zwaan et al. [19] and Roy and Mavridis [20] contributed to the understanding of grounded acquisition of language skill. Weng et al. [2] recently developed a cortex-like temporal processing model for incremental learning of text-motor behaviors for natural language.

This work is unique as it models a process of language acquisition using both the dorsal (where/how) and ventral (what) pathways so that words of the language not only have their meaning in terms of "what", but also in terms of "where/how". This is the first general purpose model that is capable of dealing with multiple motor areas, including visual and auditory, for language processing. It shows how behaviors within a motor area and between different motor areas are integrated in contrast to the architecture with behavior-based robots [21] where a separate behavior arbitration module is used to determine the priority of inconsistent behaviors from different behavior modules, the behavior integration in our model is tightly integrated into the network itself.

III. Psychological grounding of the Model

Language is a complex means of human expression having varied components including lexical-semantics, phonemes, grammar and prosody, just to name a few. Acquiring language includes acquiring and developing the above mentioned skills along with many others, but it all starts with the child starting to associate words through imitation and generalizing and forming informal concept categories as he/she gains more experience. In this paper, we focus on modeling the above two phases of early language acquisition.

a) *Association*: Piaget's early work on cognition emphasized the role of active experience in development of increasingly sophisticated mental structures for early language acquisition [22]. According to [23], a language can be characterized as a continuous sequence of sounds forming structures to which our ears after a certain time get accustomed to and develop a certain amount of probability as to what word should/would follow a certain group of words. This forms a language structure that is not explicitly taught but is slowly acquired through listening to more examples or being taught and corrected by a teacher. This is called "association" of phrases and words. For e.g., after certain real life experience the sentence *Baby eats food* makes sense but *Newspaper eats house* does not because *Newspaper* is never associated with *house* through *eats*. This is also a concept of classical conditioning.

b) *Generalization*: According to Nathan Stemmer, another very powerful capability the children apply while learning languages is the ability of generalization [24]. As a child gains knowledge of the world he is able to categorize a *brown colored, Golden Retriever* and a *white colored, one eyed St. Bernard* in the *dog* category. Gomez [25] found that infants can generalize when they are presented with different samples generated by the same formal system thus being able to discern the structure if given sufficient evidence to support it. But generalization is not sufficient as it must be combined with correct discrimination [26]. Discrimination involves the organism's ability to detect differences among stimuli and respond correctly to a specific stimulus. Both generalization and discrimination together result in the complete knowledge of an object.

IV. Network Architecture

A great part of human cognitive capacity comes from the association of auditory and visual sensory modalities, this is possible due to the lexical semantic area, which in humans also responds to visual-linguistic stimulus [4] so as to be able to relate words, heard or read, to an appropriate concept. We assume the input to our network to be such visual-linguistic stimulus and hence as a major novelty of this work, we introduce where-what pathways in the network simulating the brains dorsal and ventral pathways, found by Mishkim et. al. [27] through their lesion studies. Our model is a simplified model of the lexical semantic area, and brings together important portions of the story without delving into minute details:

a) The dorsal pathway processes the "where/how" information required by its end motor — the arms. As an arm reaches a jingling toy, the location of the toy guides the action of the arm, but not directly the type of the toy. For visual-linguistics, the location of the text on a page is useful to understand the meaning or the context of the document. For e.g., text at the top of a page might mean the title. For our network, we call the where motor as "placeholder motor" to identify the spatial characteristics of a page. To draw an analogy with a webpage, some of the text is tagged, description tag, heading tag, emphasized, italicized or written in bigger font than normal.

b) The ventral pathway processes the "what" information required by its end motor — the vocal tract which helps articulating a sentence. This is like a child listening or reading a sentence and then repeating it or trying to remember the meaning, at each time frame, with/without necessarily remembering the exact sentence and other details such as the prosody.

In the proposed network, we have 4 layers (Fig. 1).

Pre-processor Layer (Purely Computational, Not Biological): Helps translate each input word into a binary encoding so that the representation takes up less memory space. The number of neurons in the layer is n if the number of unique words taught to the network is 2^n.

Sensory Input Layer (X): Could be considered to be the retina that receives visual-linguistic stimulus in the form of words read, though for simplicity we do not model it explicitly in this paper. Each word in the sentence is taken as input at a given instance of time in the order in which it appears in the sentence, this is similar to reading one word of a sentence at a time, from left to right in English or vice-versa in Persian. The network can be trained more than once on the same sentences

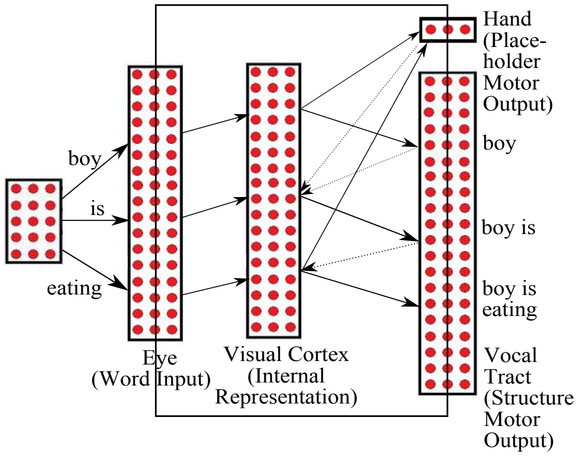

Fig. 1. Network Architecture with Dorsal and Ventral Pathways: The system boundary represents the "skull-closed" architecture. The network is taught the sentence, *boy is eating*. Each word at the 1st layer is an input to the network at a different time frame. Each input word provokes reaction from the neurons from different layers, arrows represent the synapse transferred at a single instance of time. The dotted lines represent the top-down connections while the bottom-up connections are shown with solid arrows.

as practice and review, which is like returning back to the beginning of the sentence if unable to understand. Number of neurons in this layer should be able to accommodate all the unique *n words* + ".", where "." indicates the end of each sentence/sequence, out of which we can create $n \times n$ sentences, though only a small subset of the sequence that form phrases/sentences that make complete sense will be used.
Visual Layer (Y): Neither the sensory input nor the human supervisor have a direct access to this layer, hence "skull-closed". The layer takes bottom-up input from the earlier sensory-input layer along with a supervised top-down input that could either be taught by a teacher through supervision or learnt from the past experience, to develop an internal representation of the knowledge. This involves learning a word sequence that might include the same words as the input or the word's meaning. It is important to notice that the network does not take a bag of words approach and only learns meaningful sentences/phrases. Every sentence ends with a ".", after which the network starts learning a new sequence.
Motor Cortex (Z): Consists of the motor neurons of the network that drive muscles. The placeholder motor could be the hand, reaching out to point the occurrence of a word. Similarly the structure motor could be the vocal-track helping in articulating thoughts in the form of speech (overt) or "self-talk" (intentional or covert). For its simplicity we have not modeled all the brain areas that along with the cortical connections, as described in [28], help in mapping word to articulation or mapping word to other language properties, e.g., semantics, grammar etc.
Skull-Closed Cortical Development: Before "birth" the network is not specialized in performing any particular task, it can only do so when it is trained after birth. During training, the lower layer X, receiving the input from the external world, and higher layer Z, receiving supervision signals from a human teacher, help the development of the cortical layer Y. Assuming, the input from X is $\{x_1, x_2, ..., x_n\} \in x$, representing n unique words that create the sentences taught and bottom up weights, v_x map each input word to Y. Similarly, if the output in Z is $\{z_1, z_2, ..., z_m\} \in z$, for m unique sequence of words taught. We attach top-down weights, v_z, to map each output sequence to Y. Thus, calculating the pre-action potential as,

$$y = \frac{x}{\|x\|} \cdot \frac{v_x}{\|v_x\|} + \frac{z}{\|z\|} \cdot \frac{v_z}{\|v_z\|}$$

which measures the degree of match between bottom-up and top-down inputs. The weight of the winning neuron is updated by a dually optimal Hebbian-like learning mechanism,

$$v_j = (1 - \rho(n_j))v_j + (\rho(n_j))y_j p$$

where j is the winning neuron and $v = (v_x, v_z)$ with yp being the product of pre-synaptic and post-synaptic activity of the firing neuron. $\rho(n_j)$ is the learning function that depends on n, age of the neuron. When a neuron wins its age is incremented by 1. Lateral inhibitions in the cortex allow only few top-k neurons to win. We can choose the number of winners or k, based on the amount of generalization we want the network to learn. Hence, $j = \text{top-k-max}_{i=1}^{m}(y_i)$.

Layer Z is updated similarly but it has no top-down input.

So, if the network exists in time, $t-1, t, t+1, ..., t+n$, and if $V_x(t)$ and $V_z(t)$ are the weight vectors of Layer X and Z at time t respectively and f is the area function, then,

$$y(t) = f(x(t-1), z(t-1), V_x(t-1), V_z(t-1))$$
$$z(t+1) = f(y(t), V_z(t))$$

We must note that time is important but not critical for the function of the network. We expect time to become flexible after training.
Creating New Sentences Through Generalization: Words representing similar concepts tend to excite the same neurons thus creating a similar internal representation for words or phrases with similar meanings. This could be deemed similar to the concept of "partition" in set-theory, now if *W* is the set of all words and $\{p_1, p_2, ...p_n\} \in P$ be one of its partitions, where each member has similar internal representation, then to create a new sentence, let P be associated with other partition $\{z_i\} \in Z$ through a relation *R*. This is similar to *boy* being related to *banana* through *eats*. We can conclude that since, all members of *P* have similar relations with all the partitions,

$$y_{p1}(t) = f(p_1(t-1), z(t-1), V_{p1}(t-1), V_z(t-1))$$
$$\Rightarrow y_{p2}(t) = f(p_2(t-1), z(t-1), V_{p2}(t-1), V_y(t-1))$$

Hence, $y_{p1} = y_{p2} = \text{top-k-max}_{i=1}^{m}(y_i)$.

Thinking: The network takes a word and produces its corresponding equivalent in term of meaning or context. The equivalent words co-fire. In the next time step the equivalent word is taken as top-down input by layer *Y*, which is combined with the new bottom-up words (which might or might not have an equivalence class) giving a new

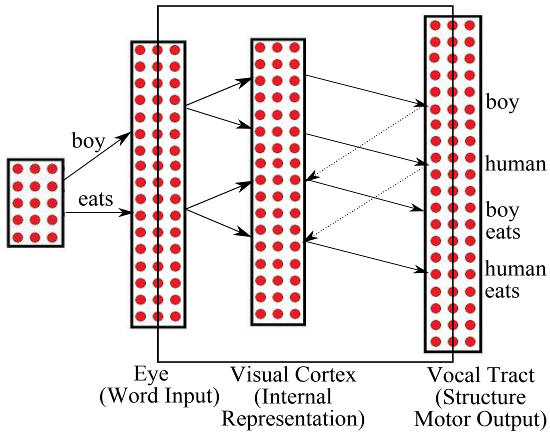

Fig. 2. Creation of New Sentences Through Generalization: The network is taught the concept *boy is a human* resulting in the neurons representing *human* and *boy* co-firing, thus associating *human* with all the concepts that are associated with *boy* like *eats*. The dotted lines represent the top-down connections while the bottom-up connections are shown with solid arrows.

sentence sequence in verbal motor. Thus both the word and its equivalent concept combine together to form new context and structural rules.

V. Probabilistic Proof of Concept

Many NLP methods are batch in the sense that all the training data are available as a batch for training. However, development is an incremental process — the agent must respond even while being trained. In general, if $x_1, x_2, ..., x_n$ are the words in a sentence that act as a sequential input to the network, then the joint probability density of this sequence will be

$$\Pr(x_1, x_2, ..., x_n) = \Pr(x_1) \prod_{i=2}^{n} \Pr(x_i | x_1, x_2, ..., x_{i-1}) \quad (1)$$

However, estimation of this joint probability is expensive, and it does not lead to generalization required abstraction. Hence we introduce the concept of equivalent classes. Two sentences belong to the same equivalent class if they have the same meaning.
Now we can write (1) as,

$$\Pr(x_1, x_2, ..., x_n) = \Pr(x_1) \prod_{i=2}^{n} \Pr(x_i | \phi(x_1, x_2, ..., x_{i-1}))$$

where $\phi(x_1, x_2, ..., x_{i-1})$ is the equivalence class for $x_1, x_2, ..., x_{i-1}$. Traditionally, the above has been used for NLP. However, according to our above discussion, the purpose of cognition is to generate desired action, z_n, at each time. Thus, our formulation of a developmental agent is to focus on $\Pr(z_n | \phi(x_1, x_2, ..., x_{i-1}))$ instead of the sensory distribution $\Pr(x_i | \phi(x_1, x_2, ..., x_{i-1}))$. This is critical for "skull-closed" development because the teacher does not manipulate internal "brain" representation directly. Symbolic representation, on the other hand, corresponds to a "skull-open" approach as it is handcrafted. Furthermore, z_n is also general and flexible as it can correspond to any property of the input context. For example, the action can be directly related to the sensory class (e.g., state the name of input) or to other property of the sensory input (e.g., its location for correct arm reaching).

Lastly, the agent learns z_n recursively as the context that it needs to attend at the n-th time frame from any point in the past. The intractable problem of estimating very long temporal joint distribution above is converted into a single frame problem:

$$\text{top-k-max}_{z_n \in Z} \Pr(z_n | x_1, x_2, ..., x_{i-1}) \approx \text{top-k-max}_{z_n \in Z} \Pr(z_n | z_{n-1}, x_n)$$

where $z_{n-1} = \phi(x_1, x_2, ..., x_{n-1})$ and top-k means top-k actions to top matched probabilities.

VI. Language Association

The network uses the concept of "primary and secondary association" or classical conditioning [29] to learn new concepts about objects. The occurrence of conditioned stimuli, CS_2 followed by CS_1, followed by a conditioned response (CR) trains a subject to correlate the occurrence of CS_2 to an otherwise unrelated CR.

$$CS_2 \xrightarrow{p} CS_1 \xrightarrow{p} CR \Longrightarrow CS_2 \xrightarrow{s} CR$$

($\xrightarrow{p}$, $\xrightarrow{s}$ means "primary" and "secondary" associations respectively, $\Longrightarrow$ means "results in"). In our model, firing of neurons is equivalent to a stimulus or an event hence sequential firing of neurons could result in the emergence of a new patterns and concepts.
We here present two of the many such new patterns:
Parent-Child Association: If $\{C_1, C_2\} \in C$, $\{C_{11}, C_{12}\} \in C_1$, $\{C_{21}, C_{22}\} \in C_2$, where C, C_1, C_2 are all conceptual objects and neurons representing a "child" concept and its corresponding "parent" class co-fire. For e.g., every *girl* is a *human* so *girl* is a child concept of parent class *human*. Tracing the progression of time, we can see that,

$$C_{11} \xrightarrow{p} C_1 \longrightarrow C_1 \xrightarrow{p} C \Longrightarrow C_{11} \xrightarrow{s} C$$

($\longrightarrow$ means "follows in time").

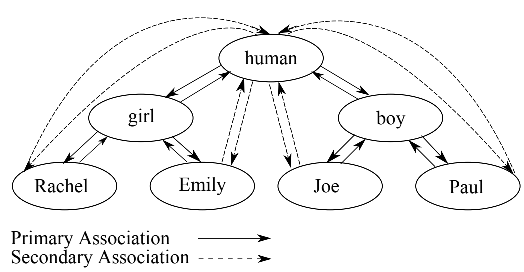

Fig. 3. Parent-Child Association: A concept "human" is defined along with the concept of "girl" and "boy", the left branch of the tree is explained thus, *girl is a human* and *Rachel and Emily are two girls*, the network is then expected to figure out that *Rachel and Emily are both humans* through association.

Sibling Association: Network is taught to identify members of the same "partition", as defined earlier, and apply the property of one member to the other, while not confusing the members of separate partitions to be similar.

$$C_{11} \xrightarrow{p} C_1 \longrightarrow C_1 \xrightarrow{p} C \Longrightarrow C_{11} \xrightarrow{s} C \; ; C_{12} \equiv C_{11}$$
$$\Rightarrow C_{12} \xrightarrow{p} C_1 \longrightarrow C_1 \xrightarrow{p} C \Longrightarrow C_{12} \xrightarrow{s} C$$

But since $C_1 \neq C_2$, hence though $(C_{21} \equiv C_{22}) \xrightarrow{s} C$, yet $C_{12} \neq C_{22}$, i.e., though *Kiwi* and *Sparrow* are both birds yet the network understands that they have different flight capabilities.

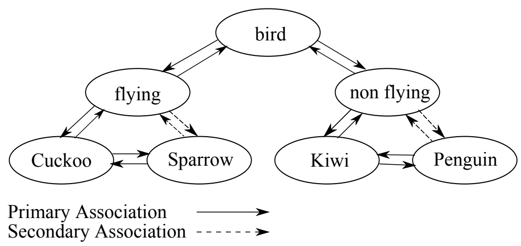

Fig. 4. Sibling Association: The concept of "bird" is taught, along with the fact that a bird could be a "flight bird" or a "non-flight bird", explaining the left branch of the tree, *Cuckoo is a flying bird* and *Cuckoo has the same properties as a Sparrow*, the network figures out the that *Sparrow is a flying bird too*.

The process involved in both the processes is association and not logical reasoning. It should also be noted that in the current network we have to explicitly teach the network if the relation between the concepts is bidirectional or unidirectional.

VII. EXPERIMENTS

The data consists of 40, simple English sentences, e.g., *Water flows* and *Boy sees a bird* in active voice. The network is trained on the same sentences multiple number of times. An "is-a" relationship is defined in some sentences, to define an object partition relationship, e.g., *Sparrow is a bird*, relates *Sparrow* to more general class *bird*. Needless to say objects in the same partitions are equivalent. It should be noted that the concept of "class" or "member object" is not hard-coded in the network. The "where" information is an "associated property" for each input. In the experiments the associated property is a tag for each word, the 4 tags are title, italic, bold and normal word, analogous to that of a webpage. We keep in mind that tags are helpful in identifying important words and hence can draw a reader's attention. Learning rate = 0.7. The network was introduced to 6 general classes with 2 subclasses each having several members. The following experiments show the 5 major capabilities of the network.

Parroting/Repetition: The network was expected to learn and repeat all the word sequences constituting a phrase/sentence along with the corresponding tags/"associated property" taught during training. Example of a word sequence would be, in a 4 word sentence, $s = ABCD$, where A, B, C and D are the words that make up the sentence, the word sequences learnt will be $s_1 = A$, $s_2 = AB$, $s_3 = ABC$ and $s_4 = ABCD = s$. All words have one associated property/tag.

Perceptual Generalization: The network is introduced to some objects belonging to a more general class, e.g., *Sparrow is a bird* and is expected to apply the object's, *Sparrow*, properties, *flight*, to its respective parent class, *bird*.

Unseen members of a class: If a certain fact/property of a certain member of a class is known to the network it is also applied to the equivalent members of the same class as an equivalent relationship. E.g., the network is taught *Sparrow can fly*, and $\{sparrow, cuckoo\} \in bird$. The network should learn *Cuckoo can fly* on its own. Here the network employs the concept of *Thinking* as described earlier.

Parent-Child and Sibling Association: The processes are as explained in section VI. The network was introduced to 6 general classes with 2 to 3 subclasses each and was expected to create logically correct sentences using the two concepts.

The results of the experiments are shown in Fig. 5.

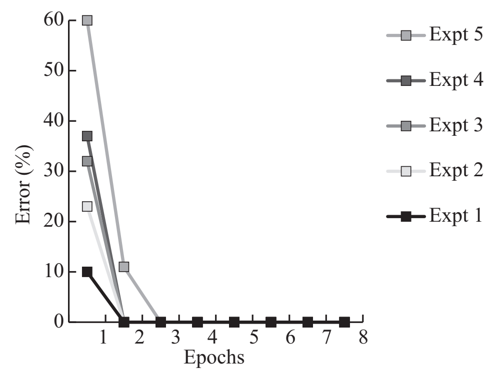

Fig. 5. Results of Experiments 1 to 5: The network is able to reach 100% recognition rate within 2 epochs of training in all the experiments. It should be noted that the network does not apply logical reasoning but learns association between objects and concepts.

VIII. DISCUSSION

A few speculations could be made about the context and attention; context helps attention and attention helps balancing generalization and discrimination. The ventral motor pathway could be used to develop context such that if the network is taught *dog is an animal* and *sparrow is bird*, then "bird", "animal" and "not bird", "not animal" could be higher level concepts. The multiple levels of generalization could help the network to focus on the correct class. Generalization could be further used to fine tune our results through multiple muxel priming, as noted by [2]. We can introduce a Pre-TM layer, between the V2 and TM layers, which is a part of the motor hierarchy. The Pre-TM helps generalization when more than one motor neuron primes on at one time, both the object as well as its class fire together each time any of the object primes in Pre-TM, so that the network develops a concept of the class. Hence the class and object do not compete at the Pre-TM layer and are counted as the same, but they are perceived as individuals in the motor layer. Thus we are able to strike a balance between generalization and discrimination.

But there are a few questions that still need to be answered. How fine or coarse should the generalization be so that the network is able to create correct classification model? Though for now the teacher decides the value of k i.e., how many neurons should win in Layer Y, but ideally after ample experience the network should be able to decide it on its own.

IX. Conclusion

Although there are symbolic systems that model language acquisition [30], our system appears to be the first recurrent connectionist model that does it without using handcrafted internal representation. For e.g., traditional NLP systems require the human programmer to handcraft a static vocabulary and hand designate a word to each Hidden Markov Model (HMM). Further how such HMMs link with others are also handcrafted. In contrast, our network fully automatically develops all such wirings and strengths through weight adaptation. In this sense, this seems the first truly “autonomous” developer for language acquisition in the sense that internal self-organization is fully autonomous after the “birth”.

Comparing the open-style connectionist language networks (e.g., those used by Rogers and McClelland [31]) and many open-style symbolic networks (e.g., [30]), the most obvious characteristic of our architecture is that the network is highly recurrent between the internal layer and the motor layers. While some modelers turned off recurrence during learning of their recurrent networks [32], [33], the major reason for us to succeed in dealing with such a high degree recurrence during learning was because of the series of cortex-like mechanisms of LCA [1]. The network is still at a very nascent stage. It is needless to say that it is far from reaching its potential in terms of richness, complexity and scale yet it does try to open new avenues by modeling a cortex like robust and efficient network and giving acceptable results. Language acquisition is not a trivial task and there are a lot of psychological motifs behind it, by studying language acquisition more we might finally gain deeper understanding of intelligence and thoughts.

References

[1] Ji Z., Weng J., Prokhorov D., ’Where-What Network 1: “Where” and “What” Assist Each Other Through Top-down Connections’. Proc. IEEE International Conference on Development and Learning, Monterey, CA, p.61-63, Aug. 9-12, 2008.

[2] Weng J., Zhang Q., Chi M., Xue X., ’Complex Text Processing by the Temporal Context Machines’. Proceedings of the 10th Annual Conference of the Cognitive Science Society, Montreal (1988).IEEE 8th International Conference on Development and Learning, Shanghai, China, June 4-7, 2009.

[3] Zhang Y., Weng J., ’Task Transfer by a Developmental Robot’. IEEE Transactions on Evolutionary Computation, Vol. 11, No. 2, p.226-248, April, 2007.

[4] Barsalou LW, Simmons WK, Barbey AK, Wilson CD. Grounding conceptual knowledge in modality-specific systems. Trends Cogn Sci. 7:8491, 2003.

[5] Iverson J.M.,’Developing language in a developing body: the relationship between motor development and language development’. In Journal of child language. Vol. 37, No. 2, p.229-261, 2010.

[6] Yeung H.H., Werker J.F., ’Learning words’ sounds before learning how words sound: 9-month-olds use distinct objects as cues to categorize speech information’. Cognition, Vol. 113, No. 2, p.234-243, 2009.

[7] Piaget J., ’The Origins of Intelligence in Children’. International Universities Press, Madison, New York, 1952.

[8] John Locke, ’Essay Concerning Human Understanding’, 1689.

[9] Hinton G.E., ’Implementing semantic netowrks in parallel hardware’. In G.E. Hintons and J.A. Anderson (Eds) Parallel Models of Associative Memory, Hillsdale, NJ: Lawrence Erlbaum, 1981.

[10] McClleland J.L., Kawamoto A.H., ’Mechanisms of sentence processing: assigning roles to constituents’. In J.L. McClleland and D.E. Rumelhart, Parallel Distributed Processing, Cambridge, MA: MIT Press, 1986.

[11] Benello J., Machie A.W., Anderson J.A., ’Syntactic category disambiguation with neural networks’. Computer Speech and Language, p.203-217, 1989.

[12] Allen R.B., ’Sequential connectionist networks for answering simple questions about a microworld’. Proceedings of the 10th Annual Conference of the Cognitive Science Society, Montreal, 1988.

[13] Sharkey N.E., ’Implementing soft preferences for structural disambiguation’. KONNAI. Journal of Psycholinguistic Research, Vol. 23, No. 4, p.295-322, 1990.

[14] Lange T.E., Dyer M.G., ’High level inferencing in a connectionist network’. Connection Science, p.181-217, 1989.

[15] Smolensky P, ’On variable binding and the representation of symbolic structures in connectionist systems’. Tech Report CU-CS-355-87, Dept of Computer Science, University of Colorado, Boulder, CO, 1987.

[16] Elman, J.L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., and Plunkett, K. ’Rethinking innateness: A connectionist perspective on development’. MIT Press, Cambridge MA, 1996.

[17] Cruse H., ’Neural Networks as Cybernetic Systems’. Brains, Minds, Vol. 2, bmm615, 2006.

[18] Olga Parsons , Gail A. Carpenter, ’ARTMAP neural networks for information fusion and data mining: map production and target recognition methodologies’, Neural Networks, v.16 n.7, p.1075-1089, 2003.

[19] Zwaan R.A., Radvansky G.A., ’Situation Models in Language Comprehension and Memory’. Psychological Bulletin, Vol 123, No. 2, p.162-185, 1998.

[20] Mavridis N., Roy D., ’Grounded Situation Models for Robots: Where words and percepts meet’, IEEE IROS, 2006.

[21] Arkin R.C., ’Behavior-Based Robotics’. MIT Press, Cambridge, Massachusetts, 1998.

[22] Roy Evans and Deborah Jones, ’Metacognitive Approaches to developing Oracy, Developing Speaking and Listening with young children’ 2009.

[23] John Harris. ’Early Language Development, Implications for Clinical and Educational Practice’. Routledge, London, 1990.

[24] Stemmer, Nathan. ’The Role of Innate and Acquired Generalization Classes in Language Acquisition’. Paper presented at the Interdisciplinary Conference, ”Perspectives on Language”, University of Louisville, May 6-8, 1976.

[25] Gomez, R. L., ’Variability and detection of invariant structure’. Psychological Science, 13(5), p.431436, 2002.

[26] Blanton, Richard L., ’Sensory Discrimination, Generalization and Language Training of Autistic Children’, 1984.

[27] Mishkin M., Unterleider L.G., Macko K.A., ’Object Vision and Space Vision: Two Cortical Pathways’. Trends in Neuroscicence, Vol. 6, p. 414-417, 1983.

[28] Joshua Knobe., Intentional action and side effects in ordinary language. Analysis, 63:190193, 2003.

[29] Domjan M., The Principles of Learning and Behavior, 4th ed. Belmont, CA: Brooks/Cole, 1998.

[30] Levinson S.E., Liu Q., Dodsworth C., Lin R., Zhu W., Kleffner M., ’The Role of Sensorimotor Function, Associative Memory and Reinforcement Learning in Automatic Acquisition of Spoken Language by an Autonomous Robot’. Proc. of Workshop on Development and Learning, East Lansing, Michigan, p.95-100, April 5-7, 1999.

[31] Rogers T.T., McClelland J.L., ’Precis of Semantic Cognition: A Parallel Distributed Processing Approach’. Behavioral and Brain Sciences, Vol. 31, p.689-749, 2008.

[32] Deco G., Rolls E.T., ’A Neurodynamical cortical model of visual attention and invariant object recognition’. Vision Research, Vol. 40, p.2845-2859, 2004.

[33] Roelfsema P.R., van Ooyen A., ’Attention-Gated Reinforcement Learning of Internal Representations for Classification’. Neural Computation, Vol. 17, p.2176-2214, 2005.