Computational Modelling of Evolution of Language

Y N Singh (5111048)
M Tech, QIP
Department of Computer Science & Engineering
Indian Institute of Technology, Kanpur - 1, India
singhyn@cse.iitk.ac.in
19 Nov, 2005

Abstract

This paper describes the uses of computer models in studying the evolution of language. Language is a complex dynamic system that can be studied at two fold one at the level of biological evolution of some innate linguistic ability, and other at the level of cultural evolution of specific languages and language families. The dynamics of language evolution or rather language acquisition occurs because of the interaction between these two levels that makes the complex system of language in a population. To investigate this complex dynamics of language evolution computer models are indispensable tool. Using computer models it make plausible to do as many hypothetical experiments as one wants. Thus, Computer modelling allows us to investigate the long-term dynamics of this system and to perform hypothetical experiments on it by changing parameters and investigating how they influence the model's behavior. In this paper we investigate different computational models i.e., agent based models, genetic algorithm based models, and artificial life based models.

Key words: *Language, Language evolution, Computer modelling, Artificial life, Universal grammar*

1. Introduction

In recent years emergence and evolution of human language has been the focus of increasing amounts of research activity due to the involvement of computer. Computer scientists have been called upon by linguists, psychologists and cognitive scientists to investigate the most profound question of human language: what are the origins of language and how does language evolve?

So, a question comes in mind why people work on language evolution. This is because some people believe that language might not have arisen by Darwinian evolution. One such reason is that language came as the by-product of a gigantic brain. Language is the most interesting thing that evolved in the last 600 million years. Since, Language does not leave fossils, which makes it hard to reconstruct how language evolved. This paper discusses how computational modelling techniques can shed light on the mystery of language origins and evolution. Mainly modelling of language evolution can be organized in three areas, i.e., origin of language, emergence of language, and evolution of language. *Origins of language* investigate how linguistic capabilities could evolve. These may include drives to communicate, the evolution of cognitive capacities that allow language production, interpretation and acquisition. *Emergence of language* concerns for language production, interpretation and acquisition for example emergence of sound systems, meaning, vocabularies and grammar. *Evolution of language* aim is to investigate how language evolves over time across multiple generations.

In this paper it is argued that the use of computer models constitutes an advance in methodology that allow researchers to investigate the problem of language evolution. We organize this discussion in the following manner. Section 2 describes the problems of the human language evolution. The efficacy of computer modelling

2. Problems of Language Evolution

There are different questions about the evolution of language can be investigated. For example, when did language evolve? Which evolutionary pressures played a role, and what factors determined that humans ended up with language, or why did the ability of humans to use language evolve while other animals did not? How much of language evolution is the result of purely biological evolution, and how much of it is cultural? What other factors, besides biological evolution of individual humans can have played a role? What was the role of co-evolution between language and the brain? And that of co-evolution between infants' learning abilities and parenting behavior? [4] [5]. Other, subtler and more fundamental, questions emerge: what *is* language? What is actually innate, and biologically evolved, in the human capacity for language and what is simply the result of cultural learning processes over many generations?

After review some of the vast array of work aimed at answering the above questions about evolution of languages we find that there exists an interaction between human culture, its evolution and the evolution of language. Evolution is a historical process. The process of language evolution is both complex and dependent on historical coincidences. The evolution of human language is also in part the evolution of the human brain. This means that it has been influenced by coincidences of human history and environment as well.

Thus, two distinct forms of language evolution become apparent – the biological evolution of some innate linguistic ability, and the cultural evolution of specific languages and language families. Hurford distinguishes these as the evolution of language and the evolution of languages [3].

3. Utility of Computer Modelling

Language is a complex and non-linear dynamic system [4]. Complex dynamic systems can be implemented as computer models. Computers can simulate the behavior of these models, and provide insights in how they work. Using a computer model, however, one has complete control over all parameters and even over the exact dynamics. One can also run and rerun the model as often as one wants. Computer models therefore make it possible to do as many hypothetical experiments as one wants. Thus, computer models are indispensable tools for investigating natural systems such as human language. Computer modelling allows us to investigate the long-term dynamics of this system and to perform hypothetical experiments on it by changing parameters and investigating how they influence the model's behavior.

Using computer models to investigate aspects of complex biological systems has since 1989 been the domain of the field of artificial life [5]. In this field, mainly biological models are tested using computer simulations. These models can be about behavior of ecosystems but also about the growth of plants [6] or on such things as flocking in birds [7] or the emergence of ant trails [8].

In order to understand the use of computer modelling in the study of the evolution of language, we need to understand that there are two levels of language – *biological level of language* (language of individual) what Chomsky has called *performance* [9] and the *cultural level of language* (languages of population). In level of the population, language is a conventionalized communication system, with a vocabulary and a set of grammatical rules. The knowledge in the population is uniform to such an extent that users of the language can communicate meanings and intentions with it. This is the level that is related to what Chomsky has called *competence*. The interaction between these two levels of language makes the complex dynamics of language in a population.

It is often assumed that the language at the level of the population is uniform over space and time. However, it is obvious that these two levels can't disjoint. The population level is an abstraction of the collective behavior of a group of individuals. Behavior on the individual level is influenced by what individuals perceive of the language used in the population of which they are part. The interaction between these two levels is a feedback loop. Changes in behavior of an individual can change the collective behavior and this in turn can influence the behavior of individuals.

So, once one knows this information in sufficient detail, then this complex dynamic system can be implemented using computer models. Computers then can be used to simulate the behavior of these models, and provide insights in how they work. While we compare the behavior of the computer model with behavior of the real system, we can check whether the predictions of the theory correspond to what is found in reality or not.

With a computer model, however, one has complete control over all parameters and even over the exact dynamics. We can also run and return the model as often as we desires. Computer models therefore make it possible to do as many hypothetical experiments as we want. While interpreting the results obtained from the computer models, one should consciously map its result to the linguistic phenomenon under study so that it clearly communicate the mapping between objects in the computer model and real linguistic entities.

To make a distinction, computer modelling may be a risky technique to answering the questions of language evolution because it focused more and more on itself and rather ignored the outside community.

4. Computer Modelling Techniques

There are different modelling techniques that computationally investigate the evolution of language. Most of the techniques are agent-based. Beside that other techniques are also used that are based on optimization, genetic algorithms, and artificial life.

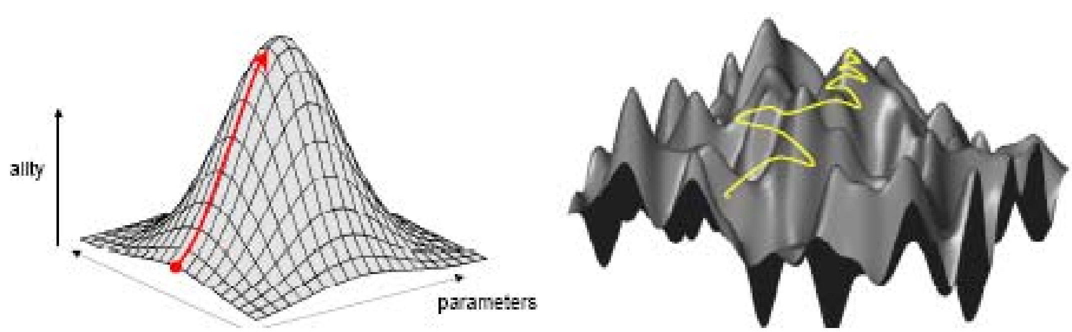
Agent-based models model (a population of) language users as simplified computer programs, and try to emulate how they use language. In this model one must be take decision the decision about what aspects of human interaction must be modeled. *Optimization techniques* define a quality function on linguistic systems and try to optimize it. *Genetic algorithms* are techniques motivated on biological evolution i.e., a population of candidate solutions, a fitness function, selection, crossover and mutation that try to evolve a good linguistic system. *Artificial life* is a scientific tool for investigation of real world where local interactions between agents give rise to an emergent and life – like behavior for example flocking in birds. Before start our discussion on agent base model let we have a quick look on optimization and genetic algorithm base modelling techniques.

4.1 Optimization Techniques

While we investigate the problem of language evolution with computer model it can be investigated the factors human language really is optimized, and how the process of optimization is brought about in human language. We can investigate the optimization criteria by generating artificial linguistic systems using different optimization criteria and comparing these systems with real human linguistic systems. For different aspects of life different optimization criteria are hypothesized. The implementation of optimization model requires,

- a clear representation of the linguistic system
- a quality measure that determines how good a given linguistic system is for investigation (formulation of a smooth quality function♣)
- Optimization algorithm

Optimization algorithms generally work by keeping track of the best solution found so far. It replaces the previous solution by a new solution with higher quality. If we find the system of high quality then we can prefer hill-climbing approach for simple ones and simulated annealing approach for complex ones [10]. Hill climbing approach gets stuck in local minima so we tries to follow the steepest path up the quality function. Simulated annealing tends to search large peaks first, and then subsequently climb up the promising peak. So, it jumps around with decreasing jumps. Fig. 1 illustrates hill climbing and simulated annealing techniques.



(a) Hill-climbing algorithm (b) Simulated Annealing algorithm

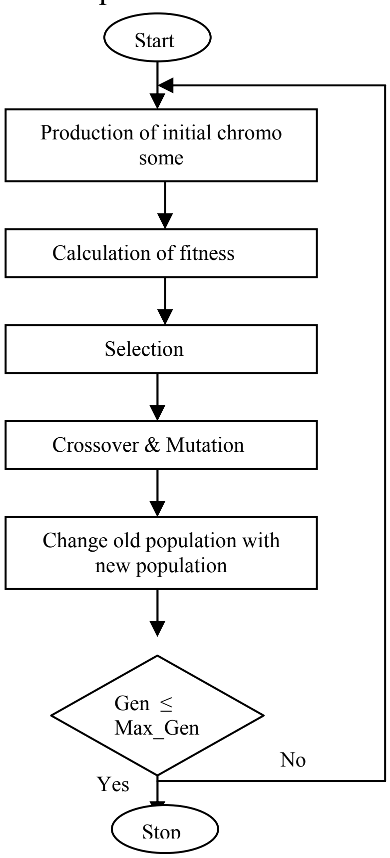
Fig. 1

♣ Smoothness of a quality function means similar systems have similar quality values. Alternatively, a function returns similar values for similar points.

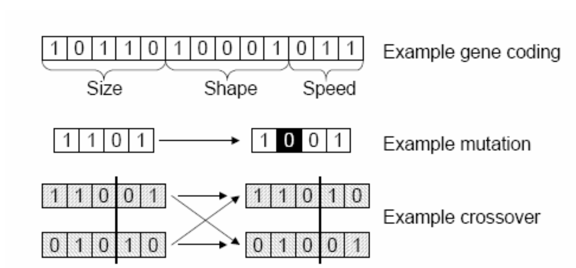
4.2 Computer modelling based on Genetic Algorithm

Genetic algorithms (GA) are powerful tools of optimizing complex systems based on *Darwinian theory*. This requires selection of an appropriate fitness function and an appropriate representation, both of the linguistic structures that are investigated, as well as their representation as artificial genes. The genetic algorithm [11] is a technique that is based on the way evolution works in nature. Instead of keeping track of only one potential solution, the algorithm has a *population* of potential solutions. GA works with two levels, at first level the solutions are evaluated and next level solutions are recombined, crossover and mutated by the genetic algorithm.

In most implementation of genetic algorithms, simple bit strings are used for representing genes. When needed, these genes are converted into possible solutions to the problem for example linguistic structures in the case of models of language. These solutions can then be evaluated with a fitness function. Solutions with high fitness are selected, and their genes are used to create new genes for offspring. GA creates offspring from their parent by applying the natural operators like mutation and crossover. *Mutation* generally consists of flipping one of the bits in a gene. In *crossover*, genes from two parents are combined to form offspring with the hopes to transfer good properties from parents. Like as in optimization for proper functioning of GA right fitness function and the right coding are essential. Fig. 2 shows the functioning of GA and mutation & crossover operations.



(a) Functioning of GA



(b) Crossover & Mutation operation for generating new genes

Fig 2

Although GA is a powerful search and optimization techniques but we should be very careful to correctly identify the following issues, i.e., GA's are not necessarily model of real evolution. It tends to exploit loopholes in the fitness function and converge due to loss of variation. Without crossover GA's are just random search.

4.3 Agent based Computer Modelling

In linguistic agent-based models, individual language users are modeled. These individuals are agents those are capable of some limited linguistic feat. Agents used for investigating speech sounds are able to perceive produce and learn speech sounds. Agents used for investigating syntax are able to produce, parse and learn syntactically structured utterances. For each linguistic question, specialized agents can be designed. The agents can interact usually by exchanging linguistic utterances, by observing their (shared) environment and by observing the nonlinguistic behavior of other agents. Depending on the interactions, the agents can modify their linguistic knowledge. So this model investigates the influence of the individual actions and interactions on the linguistic systems (fig. 3).

In agent-based model we would taken care of many design decisions i.e., how the agents interact and how they react to the interactions. So one must take the decisions about what aspects of human interaction must be modeled. Like, will it be on the basis of age structure of the population of agents, social structure, spatial structure, exchange of linguistic and non-linguistic information of population etc. besides these design decisions we need to further decide what linguistic utterances these agents can produce and what linguistic knowledge agent must have and how it can be learned. To model all these issues agent-based models becomes more complicated, but they are usually kept relatively simple.

Two dominant paradigms in agent-based modelling are language –game paradigm and iterated learning paradigm. Language – game paradigm is introduced by Steels [12] where large populations of agents are investigated without making any distinction between adults and children or between social classes of agents. They assume that primarily agents have no linguistic knowledge, and they interact through a language. So, this model basically investigates cultural transmission between agents. Iterated learning paradigm is introduced by Hurford and Kurby [13] divide the agents in adults and children. Adult agents produce linguistic utterances but do not learn, while children agents learn but they do not produce utterances themselves. Through, iteration children agents learn and replace the adult agents and new infants are inserted and the process repeats, thus providing a generational turnover. So, in this model we typically investigate the change in the language from one generation to next. For clear illustration of this model we investigate the problem of evolution of lexicon. That is how do word and meaning find each other.

4.3.1 Evolution of Lexicon

Here we investigate how words get associated with meaning. This is a study of the evolution of lexicons. Since words required meaning, otherwise communication is senseless. But how words get associated with their meaning? How do infants construe the meaning of thousand words so quickly? Psycholinguists believe that as words are

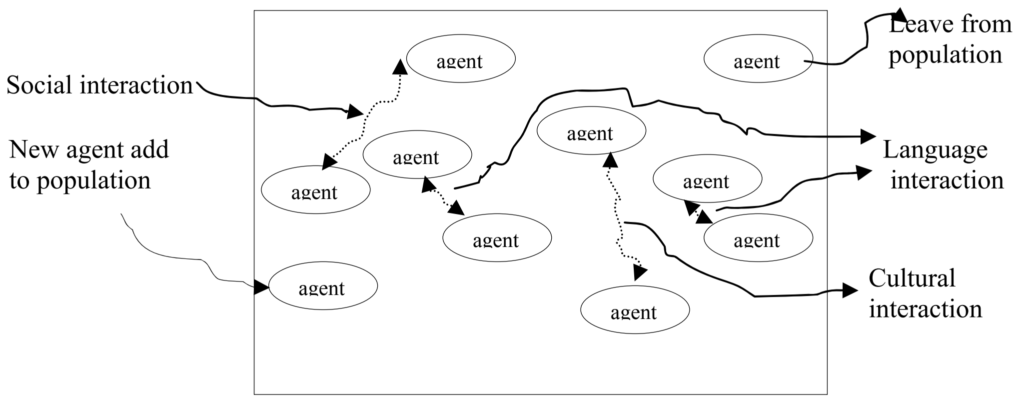


Fig. 3 Interaction between agents (population dynamics)

intricately linked to meaning, studying how children learn the meaning of words starts with studying how children represent the world. Study also show that children start to conceptualize by first relying on perceptual cues. And from start of year one on they rely more on functional information and on linguistic labels to learn and refine existing concepts [18]. How children exactly learn concepts using lexical labels is not clear. But it is clear that children start by over generalizing linguistic concepts after which they steadily refine the concept using negative and positive examples.

So, how do the word and meaning find each other agent based model build, which works on building of agents (10-10000). An agent has a lexicon, containing words. Sometimes the lexicon is fixed i.e., agent starts with fixed words. Sometimes the agents start with an empty lexicon and acquire new words as they interact with other agents. An agents need to have meanings to associate words with. Again, this meaning is either fixed or acquired as it goes long.

Each agent needs to have some way of associating words with meaning. This is like a graph connecting words with meanings. Finding of recent studies relied on using association matrices, in which words and meaning are associated to each other with a strength value. Higher values signify that the connection between a particular word and meaning is appropriate, and vice versa.

$$\begin{array}{c} \\ m_1 \\ m_2 \\ . \\ . \\ m_n \end{array} \begin{array}{c} \begin{array}{cccc} w_1 & w_2 & \ldots\ldots & w_m \end{array} \\ \begin{pmatrix} 0.1 & 0.6 & \ldots\ldots & 0.0 \\ 0.0 & 0.1 & \ldots\ldots & 1.0 \\ . & . & & . \\ . & . & & . \\ 0.1 & 0.0 & \ldots\ldots & 0.0 \end{pmatrix} \end{array}$$

This matrix is used to interpret a word (looking its meaning) and produce a word (looking the word associated with a meaning). While using single matrix, there is no difference between the behaviour of the agent while interpreting and producing.

Sometimes, humans' actively uses fewer words than they can understand. So their performance for interpreting and producing words is asymmetrical. So this is modeled by using two matrices, one production matrix and one interpretation matrix [15]. It is interesting to note that matrix representation adequately captures synonyms (a meaning has several words) and homonyms (a word has several meanings) properties. Other representations can also model the association between words and meaning, for example an artificial neural network might do the job as well that we will see in artificial life model example later.

Agents require a set of rules (learning rules) with which it can set the associations between words and meanings. These rules are such the agents' lexicons, meaning, and associations are all tuned to let the agent communicate with other agents. Bulks of studies are carried on the evolution of lexicons and meaning [16][17][18][19]. Since the relation of words and meaning is so complex, to study scientifically a separate field exist called semiotics. Researcher of this field searches how words, objects and objects' perception are linked together.

4.3.2 Emergence of Compositionality

(Iterative Learning Model)

One of the typical features of human language is the high level of compositionality. It means the utterances of human languages are highly structured i.e., parts of the utterances map onto parts of the whole meaning of these utterances. For example, the phrase 'paper weight', the word "paper" refers to reading/writing paper and the word "weight" to a bulky object. Conversely, in a holistic phrase like 'kick the bucket' (meaning refers to dying), no part of the utterance refers to a part of its meaning. One appealing hypothesis suggests that human languages have changed into compositional languages from initially holistic protolanguages [1], and several models have been developed to provide support this idea [20][21][22].

These models have shown that learners can learn the language from adults while observing only the part of the language. Computer models can show that this bottleneck in transmitting language from parent to children is not problematic, and even stimulates the emergence of compositionality. However, compositional languages allow a learner to produce utterances for previously unseen meanings when the learnt structures can be combined. So, language changes to become more learnable for future generations.

To evolve a model one critical assumption is that the agents are given elegant induction mechanism based on machine learning techniques. Besides semantic structure development, the syntactic structures could be find by observing the alignments in the utterance level. For example, an agent has heard the following utterance-meaning pairs, in which the alignments are underlined,

$$\underline{ab}cd \text{ ---- } \underline{1}1\underline{0}$$

and

$$\underline{ab}dc \text{ ---- } \underline{1}0\underline{0}$$

from such utterances, the agents can learn the following grammar,

$$S \rightarrow ab\ /\ 1\#\ 0\ B$$
$$B \rightarrow cd\ /\ \#\ 1\#$$
$$B \rightarrow dc\ /\ \#\ 0\ \#$$

Here *S* is the start sentence rule, # is a wild card and B is a non-terminal node. The sentence rewrites a string *abB* where B is either *cd* or *dc* with their corresponding meaning. At first glance, word-meaning pairs are created at random, but by chance alignments are found in the signal space. So the model hypothesis that, the emergence of compositional linguistic structures is based on exploiting regularities in expressions, though constrained by semantic structures and the emergence of combinatorial semantic structures is based on exploiting regularities found in the world, though constrained by compositional linguistic structures.

The approach of this model is an iterated learning modelling (ILM) based, a familiar approach taken in modelling language evolution. Typically ILM implements a vertical transmission of language, i.e., from a population of adults and learners, learners acquire the language through the interaction with adults. So, in ILM the language is transmitted from one generation to next in one pass without competition. Kirby [22] model which is an integration of ILM with language game model allows for competition between different rules and structures, but it requires more passes through the language in order for the language to be learnt sufficiently well for generation turnover.

4.4 Artificial Life Modelling

Artificial life (ALife) is an advanced modelling technique, where the behavior of whole population can be observed [5]. Although the rules for how population might behave are not known explicitly detailed within the model – but behavior of individual within population are detailed. So through computational process repeated individual interaction over time results the population level observation. ALife is a scientific tool for investigating of real world. The basic element of ALife model is the agent. Agent is a single individual in some simulated population. Agent may be very simple and abstract and interacting with one other within simulated population. Rules govern the behavior of agents during interactions. Thus, simulated population may consist of many thousand of agents interacting according to simple rules.

ALife is the system in which local interactions between agents give rise to an emergent and life-like behavior.

An interesting example of ALife is "Biods", a model of flocking birds [7]. Individual biods follow simple rules to avoid collisions, match velocity with other biods, and try to stay near the center of the flock. So, from the interactions of many biods following these rules an emergent and life –like flocking behavior emerges.

4.4.1 Artificial Life and Evolution of Language

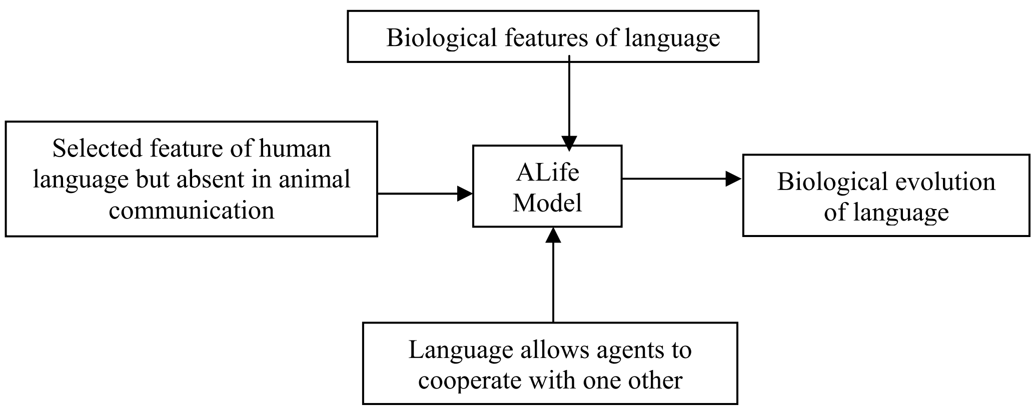
Since language is a complex dynamic system that can be studied at the level of the individual and at the level of the population. Thus, language evolution occurs as two distinct evolutionary processes. First, the ability to use language – is clearly the result of biological evolution i.e. biological evolution of some innate linguistic ability. Second, the changes that occur overtime to all spoken languages are the part of a process of cultural evolution, i.e. cultural evolution of specific languages and language families. ALife models for these two processes are different.

4.4.2 ALife and Biological Evolution of Language

By assuming that language is an organ and innovations are influenced by biological mutation and selection Hurford [16] presents six principles for evolutionary explanation of language.

1. *Universality* refers to investigation of language feature.
2. *Innateness,* the feature should be innate and evidence provided to this.
3. *Contingency,* a range of hypothetical alternatives should be presented and tested.
4. *Genetic Expression,* (language – gene mapping) some relation of language feature to genes has to be made.
5. *Adaptive Value,* (Language – advantage mapping) the possibility space of alternatives should be related to fitness and/or reproduction.
6. *Demonstration,* an argument should be presented that given 4 and 5 the feature will either necessarily/ probably emerges as the only survivor out of the possible alternatives 3.

Based on these guidelines a model is built for genetic transmission of language, which is shown in fig. 4 which successfully using language allows agents to some how cooperate with one other. Such cooperation is reward by the model.



(Fig. 4 genetic transmission of language)

A simulation is run and results gathered which show how language may evolve to support cooperation. Sooner or later some agents will evolve the 'linguistic ability' and will start to cooperate. This is rewarded, and language evolution succeeds.

4.4.3 ALife and Cultural Evolution of Language

A different set of guidelines will apply to model the historical or cultural evolution of language. Some ALife models possibly using the, "*meme*" paradigm [23], may use hereditary signaling systems to represent language, where means of transmission incorporates learning from other agents.

Another model developed by Arita and Koyama [24], although not a memetic model, uses a hereditary signaling system although it has been developed to investigate the evolution of dialects in language. The finding of this model is more closely to the evolution of cooperation rather than the historical evolution of language, i.e., common dialects allowing cooperation exist where resources are plentiful; non-compatible dialects preventing cooperation exist where resources are scare. Such findings obviously show how different environmental conditions can affect the evolution of cooperation in species but model fails to allow for non – cooperation between agents that share a common dialect.

So, an ALife model of the historical evolution of languages should avoid the genetic transmission of languages, and instead incorporate more realistic means of cultural transmission. An explicit learning mechanism is a better alternative – agents can learn language from a selection of other agents instead of inheriting it from two parents.

4.4.4 An ANN based ALife Model Example

For investigating the evolution of language any computational model should capture relevant features of language. The key features of human language are that language is, communicative, arbitrarily symbolic, regularly structured and structured at multiple levels, generative and productive, dynamic, and transmitted by learning rather than hereditary means. We use ANN based language agent. The principle advantage of an ANN implementation is that it is relatively easy to generate individuals with differing network structures, representing differing innate linguistic abilities. A suitable learning rule will allow the development of language by individuals. It is easy for an ANN to learn uni-directional mapping, i.e., from some nominal 'meaning' to produce a 'signal' A learning rule was chosen which would allow 'signals' to be fed backwards to produce meaning like bi-directional associative memory.

For the purpose of investigating the evolution of language, a simpler 2-layer (internal state and signal) model can be used with a requirement that agents are able to learn signal-meaning mappings. While agents no longer have differing internal representation, the function performed by the weights is same i.e., learning a mapping to generate common signal for common input states [25]. The agents maps an environment, *E* to an internal state, *I* to a signal *S* i.e.,

$$S = f_2(I) = f_2(f_1(E)) \\ = f(E)$$

Here assumption is that function mapping from *E* to *I* is given and the agents have a common representation for different meanings. The results of this study suggest that an algorithm using the transmission behaviour of the population to train language reception and the reception behaviour to train language production performs optimal learning. Therefore, some kind of inverted learning algorithm is required. In these models, networks posses feedback generative weights and feed forward recognition weights. So, learning algorithm should try to adapt the weights i.e., for any given meaning-signal pair the signal should produce correct meaning when feedback through ANN.

The operation of the learning algorithm is then as follows. A learner is presented with a meaning-signal pair. The signal is presented at the output layer of the learner and feedback to produce a generated meaning. This is compared with the original meaning, if any error then updates the current weights of the network (fig. 5).

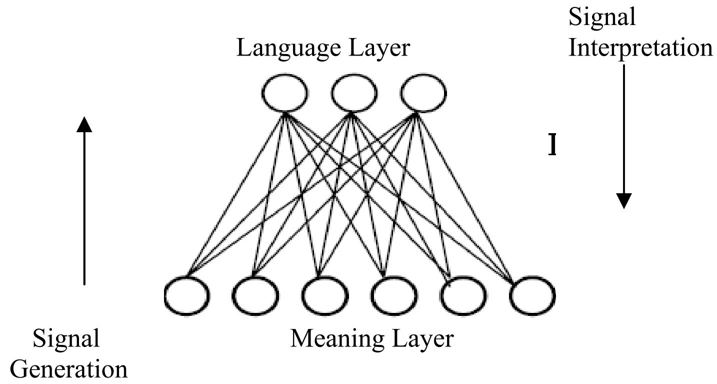


Fig. 5 A language agent ANN

Signal Production, assume a language agent ANN with layers containing M and N nodes respectively, and meaning is modeled as a bipolar (±1) vector of length M which is presented at the inputs of a language agent, representing the agents' internal state. Signals are represented at the language layer as arbitrary bipolar vectors of length N.

A meaning vector can be fed forward through a ANN to determine an agents' signal for that meaning, and each output is being threshold to a bipolar value (±1) as,

$$Y_j = \sum_{i=1}^{M} X_i W_{ij}$$

So, if $Y_j > 0$, then $Y_j' = 1$ otherwise $Y_j = -1$ where the vector $Y = (Y_1', Y_2', \ldots, Y_N')$ is the word generated for meaning vector X. A sparse coding of the meaning is used with only one bit in the vector having value +1, and all others are –1.

Signal Interpretation, to interpret a signal vector the signal can be fed back to generate a meaning vector. Competition can be applied to set only one bit of the vector to +1, and rest of bits to –1. So, for N language neurons, there are 2^N possible signals and for M meaning neurons there are M possible meanings. Competition exists between neurons in the meaning layer i.e., any signal fed back from the language layer only has one corresponding meaning. Using following equation activation value of each neuron of the meaning layer can be determined i.e., single neuron with greatest activation value is set to +1, and reminder to –1.

$$X_i' = \sum_{i=1}^{M} Y_j W_{ij}$$

$$X_i' = 1 \text{ for } i = \arg\max X_k \text{ otherwise } X_j' = -1 \text{ for } j \neq i$$

Learning, during learning an agent will be presented with a meaning-signal pair. The receiver agent uses the error between the actual meaning X and the generated meaning X′ for learning. So, the rule to update the weights will be,

$$\Delta W_{ij} = \eta (X_i - X_i') Y_j$$

This learning algorithm (learning factor η is in [0,1]) only update weights when a word is misclassified and when word is correctly classified the receiving agent performs no learning.

5. Evolution of Universal Grammar

In this section we will discuss the computational paradigm of the evolution of universal grammar. In continuation of previous discussion here we look insight and formulate the evolution theory for how language changes over time and describe the emergence of the basic design features of human language i.e., arbitrary signs, words, syntactic signals and grammar. Grammar is the computational system of language. Grammar consists of rules that associates phonetic forms and semantic forms

Phonological Rules ⟷ Syntactic Rules ⟷ Conceptual Rules

Hearing & Speaking   Perception & Action

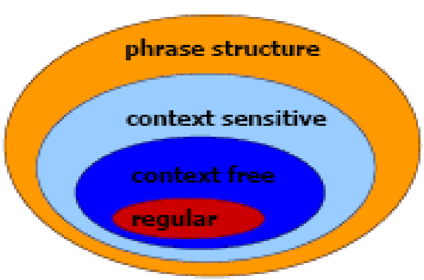
Children acquire the grammar of their native language by hearing sentences from the environment (*grammar acquisition*). This information doesn't uniquely determine the underlying grammatical rules (*poverty of stimulus*). Children could not guess the correct grammar if they had no performed, innate exception. This innate exception is Universal Grammar (UG) [26].

Formal language theory defines grammar, which consists of a finite set of rules that specifies a language, i.e.,

$S \rightarrow xSy$
$S \rightarrow \epsilon$

$L = x^n y^n$

There is equivalence between languages, grammars, and machines. To show this equivalence Chomsky constructs a hierarchy, which is known Chomsky hierarchy fig 6(a).



$A\ (T) \rightarrow L$

(a) Chomsky hierarchy   (b) algorithm for language learning

Fig. 6

Learning Theory a text T, of language L is a list of sentences S_1, S_2, ….. which contains each sentence of L at least once. Let T_n denotes the first n sentences of T. An *algorithm for* language learning is a mapping from text to language. It receives text as input and specifies a language as output (fig 6(b)). Similarly, a language L is learnable by an algorithm A, if for all texts of L, the algorithm will specify the correct language, i.e.,

$$A\,(T_n)\ \rightarrow\ L\ \text{ for }\ n \rightarrow \infty$$

A set of languages, $L = \{L_1, L_2, \ldots.\}$ is learnable by an algorithm A, if each language is learnable by this algorithm.

Using *statistical (probabilistic) learning theory*, the algorithm must converge with high probability to a language that is close to the correct language. It shows positive and negative evidences. It has computational complexity. Since the set of all regular languages are not learnable so set of all finite languages is not learnable. To formalize this theory over *human language learning*, assume that human brain contains an algorithm, A_H that can learn language. Now the question arises what is the set L_H that can be learned by this algorithm? Fig 7 shows a shared language L_H .

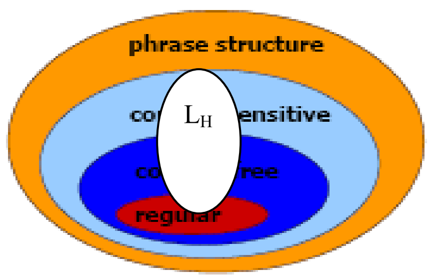
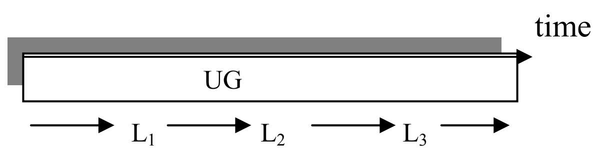
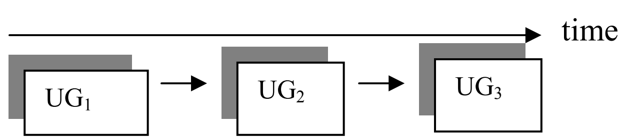


Fig . 7 The theory for the set L_H is UG

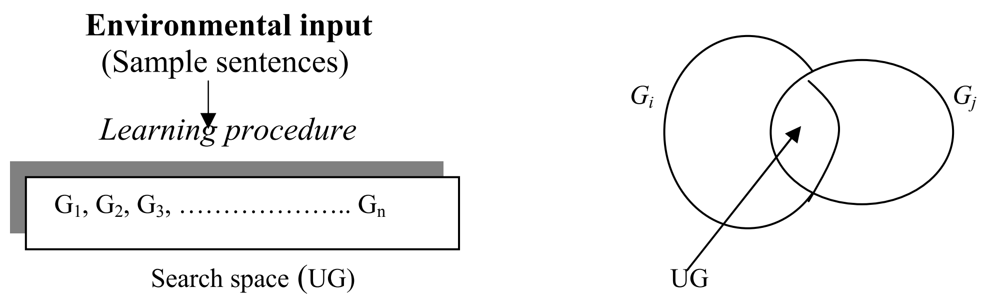
As we have earlier discussed that there are two aspects of language evolution (1) cultural evolution of language within the same UG and (2) biological evolution of UG.



In first, change in languages occur due to several factors like, neutral evolution (randomly), by product of other process (cultural, civilization, military success), and adaptively (selection for acquisition and communication). In second, UG changes randomly (neutral evolution), as by product of adaptation of other cognitive function, and adaptively.



So, it is argued that at some in the past around five million years, a UG arose that allowed recursion, discrete infinity, and making infinite use of finite means. To investigate the evolution of UG we probably investigate what criteria does UG have to fulfill to induce linguistic coherence in the population, to allow language adaptation, and to admit localization in language space. Fig. 8(a) shows the process of grammar acquisition. Fig 8(b) describes the compatibility of two grammars such that probability P_{ij}



(a) Grammar Acquisition (b) Compatibility of two grammars

Fig. 8

that a speaker of one grammar (G_i) speak a sentence that is compatible with another grammar (G_j) where,

$$P_{ij} = \mu_i\,(G_i \cap G_j)$$

Payoff for successful communication

$$F\,(G_i,\, G_j) = 1/2(P_{ij} + P_{ji})$$

Language equation

$$\chi_i = \Sigma\, \chi_j\, f_j(\chi)\, Q_{ji} \;-\; \Phi(\chi)\, \chi_i \qquad \text{(for } j = 1 \text{ to n)}$$

where,

- χ_i, fequency of G_i i.e., $\Sigma\, \chi_i = 1$ (for $i = 1$ to n)
- Fitness of G_i : $f_i(\chi) = \Sigma\, x_j\, F\,(G_i,\, G_j)$ (for $j = 1$ to n)
- Q_{ij}, probability that a learner will acquire G_j from adult with G_i.
- $\Phi(\chi) = \Sigma\, \chi_i\, f_i(\chi)$, average fitness and grammatical coherence

The language equation can be a result of a constant fitness, which is a quasispecies equation, and perfect learning, which is a replicator equation.

$$\chi_i = \Sigma\, \chi_j\, f_j(\chi)\, Q_{ji} \;-\; \Phi(\chi)\, \chi_i$$

Constant fitness

$$\Sigma\, \chi_j\, f_j\, Q_{ji} \;-\; \Phi(\chi)\, \chi_i$$

Perfect Learning

$$\chi_i = \chi_i\, [f_i(\chi) \;-\; \Phi(\chi)]$$

Assume a very symmetrical case where all grammars are all equally good. In general P_{ij} is a random number from [0, 1] and P_{ii} =1. Also they are equidistant apart. Let q is the accuracy of grammar acquisition, then

$$P = \begin{pmatrix} 1 & p & p & p & p \\ p & 1 & p & p & p \\ p & p & 1 & p & p \\ p & p & p & 1 & p \\ p & p & p & p & 1 \end{pmatrix} \qquad Q = \begin{pmatrix} q & p & p & p & p \\ p & q & p & p & p \\ p & p & q & p & p \\ p & p & p & q & p \\ p & p & p & p & p \end{pmatrix}$$

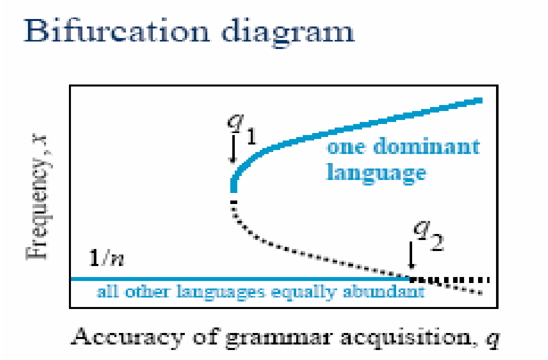
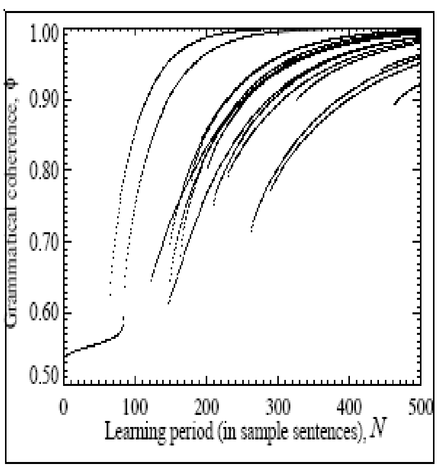
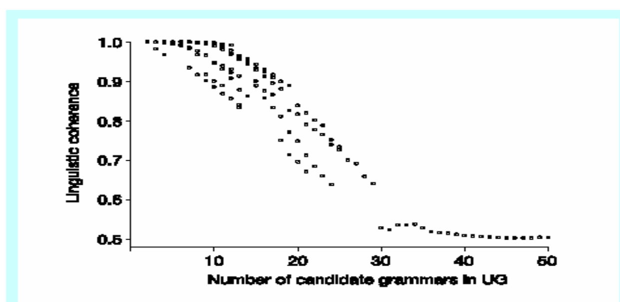


Fig 9

Fig 9 shows the bifurcation diagram. If $q < q_1$ (independent of n), then the universal grammar can induce coherent communication in a population.

Under *memory less Learner* case we start with randomly chosen grammar and stay with current grammar as long as sentences are compatible. Study the compatibility that if a sentence is not compatible with chosen grammar then select another grammar and study the result up to N sentences. For *batch learner* case memorize N sentences then decide which grammar is most consistent with all those sentences.





(a) (b)

Fig 10 (a) (b) shows equilibrium solutions of the language dynamical equation. Linguistic coherence – probability that one individual says a sentences that is understood by another individual shows on x-axis. UG specifies n candidate grammars on y-axis.

The language acquisition device is a memory less learner receiving N = 100 example sentences. For n > 30, all candidate grammars are represented in the population with similar frequencies, with linguistic coherence is about 0.5, which means a complete randomness. For n < 30 the equilibrium is dominated by single grammar. For each value of n there can be multiple equilibrium dominated by different parameters. Coherent is required for adaptation of language and selection of UG

6. Wrapping up

Computational modelling is a form of synthetic science. Using computer model we feel an abstraction of reality. Computer models are additional tools for studying the evolution of language. They provide insight to factors that play a role in making language evolution more understandable with simulation of complex dynamics of language in population. Since language is seen as much more variable and dynamics thing, so to model them we should have not only the proper computational equipment but also linguistic frame of mind. The success of computational models will depend upon that, up to what extent a researcher carefully states the assumptions and abstractions. The investigator should be careful to combine the computer modelling with careful analysis of the available data and with knowledge and understanding of the sampled linguistic data.

In this paper we investigated different computational techniques for the modelling of human language. As we saw in above discussed examples these different techniques can be applied successfully, depending on what it is exactly a researcher wants to investigate. While simulations, it is also needs to be decided what simplification and abstractions to make. Another aspect of modelling is finding of right measures to describe the performance of a model.

Most of the computational modelling techniques discussed so far are based on simple language games where some aspect of visible entity is communicated in one direction to investigate learning techniques. However, human language use is much more based on dialogues. So, to investigate the language evolution problem clearly, futures models should have the capability to investigate how dialogues can be aid in evolving language.

7. Acknowledgements

This term paper is a part of our course work in the subject *Special Topics in Language Acquisition, Evolution and Origins (CS-789)* run in the department of computer science and engineering at IIT Kanpur in session 2005-06. I am thankful to our instructors Prof H Karnik and Prof Raina for their valuable guidance and support.

References

[1] Wray, A. *The transition to language*. Oxford: Oxford University Press, 2002
[2] Hurford, J. R, Studdert-Kennedy, M. and Knight, C. *Approaches to the evolution of languages: Social and cognitive bases*. Cambridge: Cambridge University Press, 1998
[3] Hurford, J. R. *The Evolution of Language and Languages*, Edinburgh University Press, 1999
[4] Steels, L. *The synthetic modelling of language origins*, *Evolution of Communication*, 1998
[5] Langton, C. G, *Artificial Life*, Reading, MA Addison-Wesley, 1989
[6] Prusinkiewicz and LindenMayer, 1990
[7] Reynolds, C. W., Flocks, Herds and Schools: A Distributed Behavioral Model, 1987
[8] Colorni, Dorigo and Maniezzo, *Distributed optimization by ant* colonies, Elsevire Publishing, 1991
[9] Chomsky N., *Aspects of Theory of syntax*. Cambridge, MIT Press 1965,
[10] Kirkpatrik, S. et al., *Optimization by simulated Annealing* Science, 1983
[11] Goldberg, D E, *Genetic Algorithms in search, optimization, and machine learning*, Addison Wesley. 1998

[12] Steels L., *The Synthetic modelling of language origins*, Evolution of Communication, 1997
[13] Hurford et al., *Approaches to the evolution of languages,* Social and Cognitive bases, 1999
[14] Booth and Waxman, *Object names and object functions serves as cues to categories for infants*. Developmental Psychology, 2002
[15] Oliphant M., *The dilemma of Saussurean communication*. BioSystems, 1996
[16] Hurford J. R., *Biological evolution of the Saussurean sign as a component of the language acquisition device*. Lingua, 1989

[17] Smith K., *The evolution of vocabulary*. Journal of Theoretical Biology, 2004
[18] Steels L.& Kaplan. F., *Bootstrapping grounded word semantics*. In T. Briscoe, editor, *Linguistic evolution through language acquisition: formal and computational models*. Cambridge University Press, Cambridge. 1999
[19] Belpaeme T. and Bleys J., *Explaining universal colour categories through a constrained acquisition process*. Adaptive Behavior. To appear 2005
[20] Smith et al., *Intelligent meaning creation in a clumpy world helps communication.* Artificial Life, 2003
[21] Vogt, P., *The emergence of compositional structures in perceptually grounded language games*. Artificial Intelligence. In press.
[22] Kirby S., *Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity*. IEEE Transactions on Evolutionary Computation, 2002
[23] Dawkins R., *The Selfish Gene*. Oxford University Press, Oxford. 1976,
[24] Arita T. and Koyama Y., *Evolution of Linguistic Diversity in a simple Communication System*. In C. Adami, R. K. Belew, H. Kitano and C. E. Taylor (Eds.), Artificial Life, MIT Press, 1998
[25] Fyfe C. and Livingstone D., *Developing a Community Language*. In 4th European Conference on Artificial Life, ECAL 97, Brighton, UK. 1997
[26] Chomosky N., Rules *and representations*. Behavioral and Brain Sciences, 1980