Language Emergence for Activities

Manas Agarwal
manas@cse.iitk.ac.in
Under the guidance of Dr. Amitabha Mukerjee

November 11, 2009

1 Introduction

Language development is an important aspect of human cognitive processes. Robotic and Artificial Intelligent models have been used to simulate language emergence situations [7]. The aim is to study the process of specification of the universal features of language and emergence of structures (symantic/syntactic) in human languages.

2 Language emergence

There are many theories and researches trying to explain the formation of structures in human languages. Some researchers assume that compositional structures emerged from exploiting regularities found in expressions based on holophrases (single word expressions) [1]. Some other researchers have assumed that the ability to use syntax has evolved as a biological adaptation. Another important line of thinking is that the emergence of language structures is based on exploiting regularities found in the interaction with natural world. This hypothesis has been investigated using multi-agent models. A Language game model have been applied to study language emergence which mainly consisted of object naming and identification [9]. We intend to apply similar model to study emergence of compositional language consisting of objects and simple actions.

3 Language Game

The scientific aim of evolutionary linguistics is to study how modern languages have evolved from a stage prior to language. This is useful, because it provides a complimentary methodology that can help researchers to develop and test hypothesis on language evolution in the simulated world. Vogt and Steels [8] developed a model which implements a scenario in which the population of one generation transmits their language to the next generation by engaging in language games. This is called a *naming game*. The game is played by two agents: the *speaker*, typically from older generation and a *hearer*, typically from a new generation. Different games can be played. The following basic steps are involved in the game:

1. *Making Contact* : Agents assume their respective roles and come close enough to have a shared context.

2. *Perception*: Each agent categorize the sensory experience of the objects and identify a distinctive feature set that distinguish the present context with any other one.

3. *Encoding*: The speaker identifies one distinctive feature set and encodes this into expression. It means finding smallest set of words that describe the feature set from present lexicons. If no word can be formed, a new word is created and added.

4. *Decoding*: Hearer looks up all the words in its lexicon and reassembles a feature set that cover all the words.

5. *Feedback*: The hearer compares the decoded feature with the distinctive feature set it was expecting. If sets matches, the game is successful and a positive feedback is given by hearer. Otherwise the game ends.

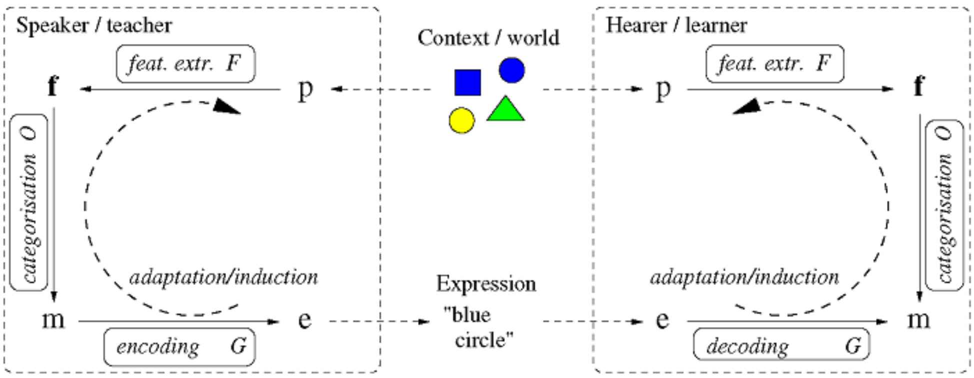

Figure 1: The Naming game processes used by Vogt. On the left are speaker's processes and on the left are hearer's processes. Between them are the feedback processes.

4 The Iterated Learning Model

The ILM is a general framework for modelling cultural transmission of language [2]. The model iterates over generations which population is divided into: adults and learners. Each generations consists a population of agents which learn language from utterances produced by the previous generations. The learner starts his life as a novice user and acquire language by observing behaviour of adults, who are assumed to have mastered the language. A number of such language games are played. At the end of each iteration, adults leave and are replaced by novel learners. This cycle is repeated. ILM is a simplifid model of population dynamics and cultural transmission of language and is also used in the *naming game*.

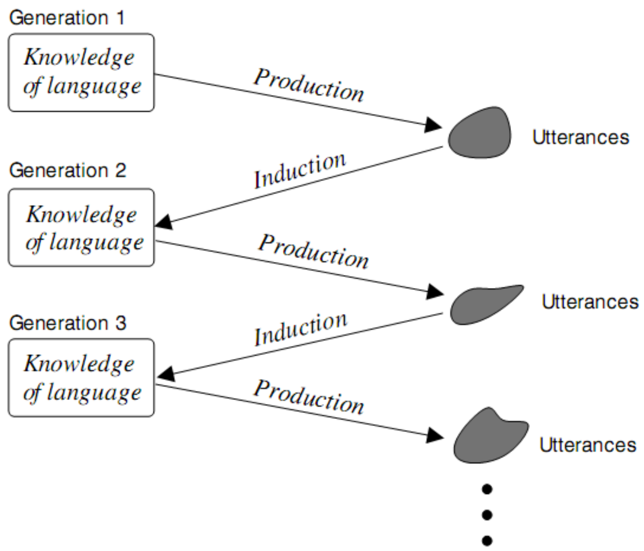

Figure 2: The schematic view of the iterative learning model. The utterances are used by the agents in the next generations to induce the knowledge of language.

5 New model: Learning for Activity

Notably, Vogt and Steel's work focussed on learning language structure only for objects, i.e. common nouns. The objects are described by color and shapes. Other works have also studied for proper nouns. But language emergence for activities have largely been untouched. In the formative years of human life, the concepts are largely acquired from dynamic activities rather than static sceneries [4]. Thus, study of formation of verbs during language evolution can provide a useful insight to the actual human language learning.

Our model uses the Vogt model of *naming game* to simulate learning of language composition when their are some agents and an activity is being performed by them. The objective of our game is to find a distinctive category for this context. Categorization means providing identifying a distinctive feature set to the two objects and the underlying action between them.

Work on the construction of an activity template from visual inputs is mostly centered on single agent activities and that too require visual priors for recognition. However, Satish, Mukerjee[5] has provided an unsupervised learning technique, which can cluster the activities into basic groups using attentive focus.It claims that semantics of certain actions may be learned prior to language, that is, we can form a crude linguistic description of a scene just by the visual inputs only. These actions were prominently change of positions of agents in a multi-agent setups, and the result were semantics describing the relative motions amongst them. They also incorporated centre of attention feature since they claim that a commentator is more likely to talk about things that are in attentive focus. The results of the work [Fig. 3] show that by restricting the constituents participating in an action using computational model of visual attention, they have achieved better correlation with the human commentary provided.

	C_1	C_2	C_3	C_4	Total	%	TCA
CC	399	6	10	29	444	90	
MA	16	311	5	48	380	82	84
Chase	21	59	149	154	383	79	

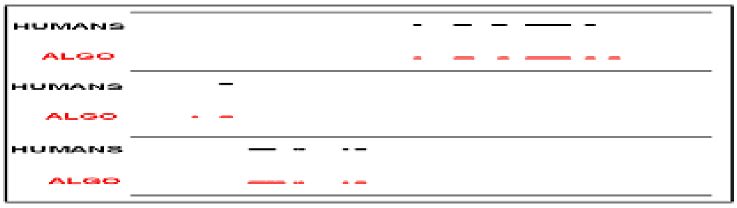

Figure 3: Clustering by Merge Neural Gas. The i^{th} row, j^{th} column gives the number of i^{th} action labels in j^{th} cluster. Below given is comparison of human and algorithm labelling.

Our model cosists of following major steps:

- Our world (The game) consist only of two agents (distinct) in a 2d space. We would like to simulate basic movements that can be categorized later. [3] describe some trajectories for such type of interaction that are shown to be classified as same action universally during learning by infants. These trajectories [Fig. 4] are used for action generations.

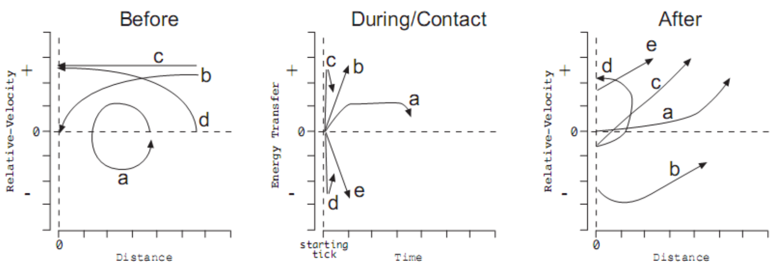

Figure 4: The labeled trajectories characterize the component phases of seven interaction types, as described by the verbs: push, shove, hit, harass, bounce, counter-shove and chase.

- The feature vector set for the context of the game will include 3 data: color of ball A, color of ball B, $f_1 = (\vec{x_B} - \vec{x_A}).(\vec{v_B} - \vec{v_A})$, $f_2 = (\vec{x_B} - \vec{x_A}).(\vec{v_B} + \vec{v_A})$ and time t of frame.

- Using these feature vectors, we would like to categorize the relationship between different balls as moving towards each other, moving away(both or one-static),one approaching another(other also moving or static), chase, push, etc. In the above mentioned work [5], Merge Neural Gas Algorithm is used. Neural Gas Algorithm learns important topological relations in a given set of input vectors in an unsupervised manner by means of simple Hebb-like learning rule. Since input features are arriving from temporally connected data(from different time frames), a *context* information is also added to pass on temporal information. This changed algorithm is the Merge Neural Gas Algorithm.

- A study of the effect of attentive focus will also be done.[6]
- The *Naming game* continues after categorization. The language consist of 3 words. Two words are for agents and the third for activity. There can be six possible ways for an agent to form sentence using same categorization. The iterative model will try to find the final language describing the activities and the syntax emerged during learning.

6 Work to be done

This work focussed on studying various language games and developing and proposing a new game for language learning for actions. Later work will include:

- Impementing the game and finding the result.
- How the game can be improved to get better learning.
- Increasing objects and introducing topic selection in the game.

References

[1] Brighton, H. Compositional syntax from cultural transmission. *Artificial Life 8*, 1 (2002), 25–54.

[2] Brighton, H., Kirby, S., and Smith, K. Situated cognition and the role of multi-agent models in explaining language structure. In *Adaptive Agents and Multi-Agent Systems: Adaptation and Multi-Agent Learning* (2003), D. Kudenko, E. Alonso, and D. Kazakov, Eds., Springer, pp. 88–109.

[3] Cohen, P. R., Morrison, C. T., and Cannon, E. Maps for verbs: The relation between interaction dynamics and verb use. In *Proc. IJCAI-05* (2005), pp. 1022–1027.

[4] Mandler, J. Actions organize the infant's world. In *Action Meets Word: How Children Learn Verbs (Hirsh-Pasek and Golinkoff, 2006)*. Oxford University Press (2006), 2006, p. 111133.

[5] Satish, G., and Mukerjee, A. Acquiring linguistic argument structure from multimodal input using attentive focus. In *Development and Learning, 2008. ICDL 2008. 7th IEEE International Conference on* (2008), pp. 43–48.

[6] Singh, V. K., Maji, S., and Mukerjee, A. Confidence based updation of motion conspicuity in dynamic scenes. In *CRV '06: Proceedings of the The 3rd Canadian Conference on Computer and Robot Vision* (Washington, DC, USA, 2006), IEEE Computer Society, p. 13.

[7] Steels, L. A case study in the behaviour-oreiented design of autonomous agents. In *3rd simulation of adaptive behaviour Conference, The MIT Press* (1994).

[8] Steels, L., and Vogt, P. Grounding adaptive language games in robotic agents. In *Proceedings of the 4th European Conference on Artificial Life* (Cambridge, MA, 1997), I. Harvey and P. Husbands, Eds., The MIT Press, pp. 474–482.

[9] Vogt, P. The emergence of compositional structures in perceptually grounded language games. *Artificial Intelligence 167* (2005), 206–242.