The Fifteenth International Symposium on Artificial Life and Robotics 2010 (AROB 15th '10),
B-Con Plaza, Beppu,Oita, Japan, February 4-6, 2010

Why we talk?: Altruism and multilevel selection in the origin of language

Kana Sugiura Takaya Arita
Graduate School of Information Science, Nagoya University
Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan
E-mail: kana@alife.cs.is.nagoya-u.ac.jp, arita@nagoya-u.jp

Abstract: Talking entails costs of production and time while some of the information sent to hearers will be of value to them in general. We believe that the matter of why we talk at all is a key question for the origin of language and the answer will shed some light on the mystery of human identity. This paper focuses on the altruism in communication and aims to demonstrate evolutionary scenarios based on multilevel selection. We constructed a computational model to examine these scenarios. The evolutionary experiments showed that in case of an unstructured population a linguistic system hardly emerged due to the dynamics between interpretable utterance that imposes a penalty and correct interpretation that yields a reward, which is similar to the prey-predator dynamics. However, in case of a multi-group population a linguistic system emerged owing to the multilevel selection among the groups. In addition, the probability of success in conversation was higher in a group with a severer environmental condition. This result supports the Bickerton's hypothesis based on the ecological gap between human ancestors and other ape species.

Keywords: language evolution, altruism, multilevel selection, artificial life.

1 Introduction

When we talk we usually transfer information to hearers. Talking entails costs of production and time although some of the information will be of value to hearers. Therefore, some explanation must be given if we admit that natural selection has shaped the human linguistic behavior. We believe that the matter of why we talk at all is a key question for the origin of language [1] and the answer will shed some light on the mystery of human identity.

This paper focuses on the altruism in communication and aims to demonstrate evolutionary scenarios based on multilevel selection [2]. Multilevel selection, a sophisticated version of group selection, refers to the idea that group-level selection and within-group individual-level selection are supposed to work simultaneously and thus alleles can spread in a population because of the benefits they bestow on groups regardless of the alleles' effect on the fitness of individuals within that group.

When considering the altruism in language, we should not neglect the difference between language and nonhuman communication systems (NCSs) [3]. Most signals in NCSs are designed not to communicate information but to confer fitness on the sender. Bickerton attributes this difference between language and NCSs to the ecological gap between human ancestors and other ape species. In short, the food-poor, predator-infested environments inhabited by proto-humans placed a premium on cooperation, which might lead to the emergence of language based on altruism.

We construct a computational model to examine these scenarios. In the model, each agent has a word-meaning association matrix that explicitly stores the joint likelihood of the words and meanings. In conversation, if the receiver can successfully interpret the meaning of the sent word, the receiver receives a positive reward but the sender receives a negative reward. The matrix is evolved by the selection based on the total rewards. We conduct evolutionary experiments with unstructured and structured (multi-group) populations to investigate the effect of multilevel selection. Also, in multi-group setting, we examine the Bickerton's hypothesis by differentiating the penalties among groups.

2 Model

The population is composed of N agents and is divided into χ patches. The set of agents occupying a patch constitutes a group. Each agent has its own vocabulary represented as a Word-Meaning Association Matrix (WMAM), which is a likelihood $n \times m$ matrix that explicitly stores the joint likelihood of the n meanings and m words. The initial values of the elements are randomly set in the range of 0 to 1, and then normalized so that the sum of all elements is one.

Each agent plays one-to-one language game to send all meanings with every other member in its group in the game phase (Fig. 1). For a speaker, p_{ij}

$(= a_{ij}/\sum_{j=1}^{m} a_{ij})$ denotes the probability of using word j when sending meaning i, whereas for a hearer, q_{ij} $(= a_{ij}/\sum_{i=1}^{n} a_{ij})$ denotes the probability of interpreting it as meaning i when hearing word j. If conversation in the game is successful, in other words, the hearer can successfully interpret the meaning of the sent word, the hearer and speaker receive the payoff, α_i and β_i, respectively. The term $\sum_{j=1}^{m} p_{ij}^{(X)} q_{ij}^{(Y)}$ denotes the probability that speaker X will successfully send meaning i to hearer Y, and is then summed over all meanings, which will be the expected payoff for the agent X in one-to-one game as follows.

$$R_{L_{(X)},L_{(Y)}} = (1/2)\sum_{i=1}^{n}\sum_{j=1}^{m}(\alpha_i p_{ij}^{(X)} q_{ij}^{(Y)} + \beta_i p_{ij}^{(Y)} q_{ij}^{(X)}). \quad (1)$$

After the game phase, each agent decides whether to migrate to a randomly-chosen patch or not by using the migration function depending on its average payoff per game. The probability of migration for agent X, $E_{mig_{(X)}}$ is defined as $\rho \times (1 - \frac{R_{(X)} - R_{Min(x)}}{R_{Max(x)} - R_{Min(x)}})$, where ρ is the parameter of the function, $R_{(X)}$ is the payoff of the agent X, and $R_{Max(x)}$ and $R_{Min(x)}$ are the maximum and minimum payoffs in the group x to which the agent X belongs, respectively.

Next, a new population is generated by the roulette wheel selection according to the average payoff per game until the population size reaches N. The selection is performed on the whole population while the population structure is maintained during this phase, in other words, each offspring is generated in the same patch as their parents. Linear scaling modifies the payoff of each agent before the selection if the least one is negative. Then a normal random number (μ, σ) is added to each element of WMAM with a probability ω as mutation, after each WMAM is normalized. Finally, the agents chosen by the migration function perform migration.

A sequence of these phases is repeated 20000 times. We will focus on the evolutionary dynamics of the vocabulary by considering the frequency of the utterance of a word from a meaning or interpretation of a meaning from a word that can be measured quantitatively as a linguistic trait.

3 Experiments

The parameters common to all experiments were set as follows. The population size N was 100, and the numbers of meanings n and words m : were equally 3. The parameters controlling mutation $\omega = 0.001$, $\mu = 0$ and $\sigma = 0.05$. The payoffs of speaker's (α_i) and hearer's (β_i) for all meanings were -1 and 1.5, respectively. This setting makes the agents with the vocabulary by which to interpret correctly the meanings of the sent words and to send the words of which the

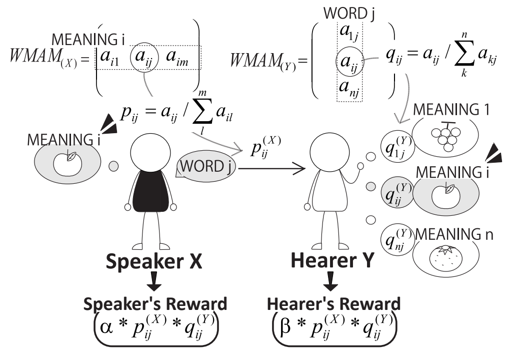


Figure 1: The language game: shows that speaker X talk about meaning i by using word j.

meanings the other agents cannot interpret correctly receive a high payoff. We shall investigate how a communication system evolves on the basis of altruism.

3.1 Unstructured population

First, we conducted the evolutionary experiments with one patch ($\chi = 1$) to grasp the basic dynamics of the communication system. Fig. 2 (top) shows the evolution of success rate in conversation. Starting from 1/3, it increased rapidly to around 0.4 within 30 generations, and then reached a peak of 0.58. After that, it gradually decreased and converged to approximately 1/3. One word was associated equally with all meanings in the converged communication system, in other words one word was used for expressing all meanings and the word was interpreted as each meaning with a probability 1/3.

Fig. 2 (bottom) shows the evolution of the six linguistic traits (expected frequencies in the population) concerning meaning 3. The trait responsible for mapping from meaning 3 to word 1 (M3W1, hereafter), M3W2 and M3W3 are expressed as "1", "2" and "3", respectively. Also, the trait responsible for mapping from word 1 to meaning 3 (W1M3, hereafter), W2M3 and W3M3 are expressed as "1' ", "2' " and "3' ", respectively. So, for example, if "1" = 1.0 and "1' " = 1.0 then success rate = 1.0 concerning meaning 3.

We see the repetition of rapid change in the dominant traits from this figure. There are at least three mechanisms that shaped this dynamics. 1) $M3W1 + M3W2 + M3W3 = 1$ by definition. Therefore, for example, M3W2 tends to decrease when M3W1 increased. 2) There is some correlation between the changes in the traits using the same words as they are based on the same elements in WMAM. 3) As conversation success gives speakers a penalty, increase in the traits responsible for interpreting a word as a meaning correctly makes decrease in a trait responsible for using the word for the meaning. Actually, we see in the graph a harmonic motion with a trait (e.g. W2M3) following another trait (e.g. M3W2), which can be seen in the classical Lotka–Volterra prey-predator model.

An increase in speaker's payoff should make it easier

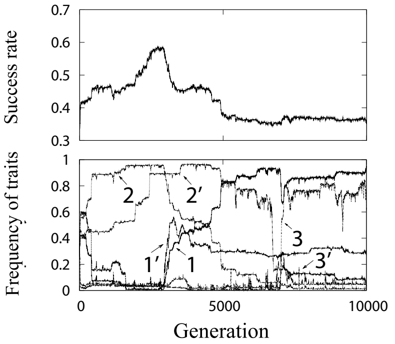


Figure 2: The success rate in conversation (top) and the frequencies of the linguistic traits concerning meaning 3 (bottom).

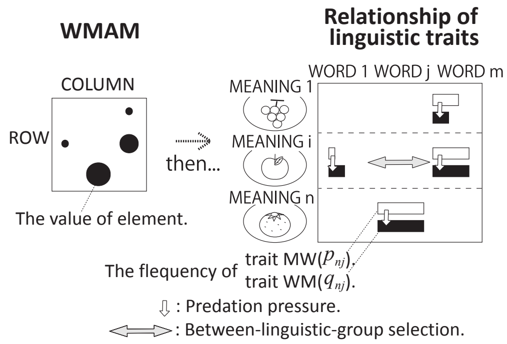


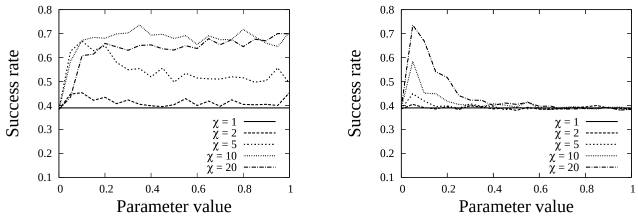
Figure 3: Multilevel selection for an implicit group.

to establish an ideal communication system in which one-to-one mapping between meanings and words. Additional experiments confirmed this tendency. It is worth noting that even the speaker's payoff was less than zero (e.g. -0.3), a high success rate (0.85) was achieved. The reason is, besides the above mechanism 2), supposed to be that agents that are the members of implicit groups in which successful conversations are held tend to gain higher payoff than the other agents who does not establish stable linguistic relations with other agents. This is a kind of multi-level selection based on the implicit linguistic groups (Fig. 3).

3.2 Structured population

Next, we examined the case with structured populations. We adopted random migration in which all agents migrate with a constant probability for comparison in addition to the method described in the previous section.

Fig. 4(a) shows the average success rate in conversation from 10000th to 20000th generations over 30 runs when varying the parameter of migration. It is shown that it was significantly greater than the one in the case of the unstructured population, especially when the number of the patch is more than 2. This suggests an effective linguistic system emerged in the structured population surely owing to multi-level selection. Also, we see that the weighted selection for poorer agents in migration works well when compared with the case with random migration (Fig. 4(b)) [4].



(a) Biased migration (b) Random migration

Figure 4: The success rate in conversation when ρ changed from 0.0 to 1.0 at intervals of 0.05. Each line corresponds to the number of patches (1, 2, 5, 10 and 20). The peak values of the lines are: (a) 0.39, 0.40, 0.49, 0.58 and 0.74, and (b) 0.39 0.45, 0.67, 0.74 and 0.70, respectively.

Here, we investigated the evolutionary scenario using the case in which the number of the patch was 5. The success rate in conversation averaged over the whole population fluctuated between 0.4 and 0.9. We found that the rate and the group size in each group were not converged but rather changing in a cyclic fashion. The mechanism can be explained as follows. As a general tendency, groups with a large conversation success rate have agents with high fitness in average. Therefore, the size of those groups tends to increase. However, at the same time, there is a tendency that agents who speaks and interprets correctly (*altruists*) have lower payoff than the agents who just interpret correctly (*free riders*). Therefore, when a group size becomes larger and larger, the average fitness has a tendency to decrease (Fig. 5).

It was also shown that extinction occurred frequently in a cyclic behavior of the group size. It is well known that the loss of variation could occur when a new population is established by a very small number of individuals (*founder effect*). This effect could accelerate the generation of cooperative groups in empty patches. Furthermore, agents from large groups tend to be consistent, in other words, speak a word for a meaning, and at the same time interpret the word as the meaning especially in the biased migration, which also facilitates the evolution of cooperation. This explanation based on multilevel selection and the founder effect is similar to the one we took in the context using the prisoner's dilemma model [4].

3.3 Effects of environmental variation

In short, nonhuman communication systems (NCSs) primarily benefit the speaker, while human language benefits the hearer. In this sense, the setup of the payoffs in the model is directed not to NCSs but

Number of patches	0	2	4	6	8	10
Success rate (biased migration)	0.509	0.518	0.523	0.525	0.525	0.527
Number of agents	41.6	24.3	15.2	10.8	8.3	6.9
Success rate (random migration)	0.437	0.441	0.442	0.445	0.447	0.449
Number of agents	75.9	7.6	4.9	4.4	4.3	4.8

Table 1: The success rate in conversation and the group size in each patch (averaged over generations and over 30 runs) with the parameters ρ=0.2 and θ=0.005 (biased migration) and ρ=0.05 and θ=0.02 (random migration).

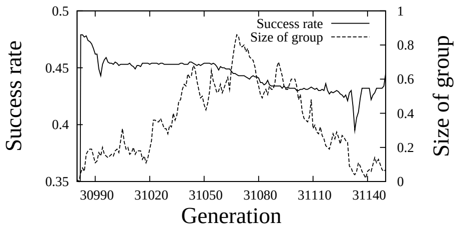


Figure 5: The conversation success rate and the population size in a group with ρ=0.1 (biased migration).

to language. We examined the effect of ecological difference on the emergence of a reliable communication system for which altruism is required, according to the discussion by Bickerton [3]. We represented the ecological gap simply as a difference in the default penalty for each patch. In this experiment, the average payoff of each agent in patch x was decreased by $x \times \theta$ as environmental severity ($0 \leq x \leq 10$).

From this experiment, we found a tendency that in severer patches the success rate was higher and the group size was smaller as the typical case in each setting is shown in Table 1. This tendency was slightly weaker when using random migration. Table 1 shows the differences in the success rate between group 0 and group 10 were 0.0179 (biased migration) and 0.0128 (random migration). The reason for this is supposed to be that severe environment requires cooperative communication and produces a larger founder effect caused by frequent extinction.

4 Conclusion

Altruistic traits that reduce the actors' fitness but increase the fitness of recipients of the act are selectively favored under positive assortment between cooperators and noncooperators [6]. Many researchers have reported that positive assortment is facilitated by such mechanisms as kin recognition, limited dispersal or behavioral bookkeeping.

This paper focused on the altruism concerning the traits responsible for speaking in the origin and evolution of vocabulary. Computational experiments showed that in an unstructured population a stable communication system hardly emerged and complex dynamics including prey-predator dynamics was observed. The experiments with a structured population demonstrated the key role of a realization of limited dispersal, that is multi-population accompanied with migration as environmental response, in the evolution of vocabulary. Also, the results supported Bickerton's claim on the significant role of ecological difference.

It should be noted that we observed that altruistic communication could evolve even in a single population. This derives from the fact that speakers tend to be correctly interpreted by the hearers having similar genetic information in this model (as both traits responsible for speaking and hearing share an identical association matrix and correlate each other) and very likely in general. We believe that this language-specific property facilitating positive assortment played a key role in the evolution of human identity.

Acknowledgements

We are grateful to R. Suzuki for fruitful discussions.

References

[1] T. C. Scott-Phillips, "The Social Evolution of Language and the Language of Social Evolution", *Evolutionary Psychology*, 5(4), 740-753, 2007.

[2] E. Sober and D. S. Wilson, *Unto Others, The Evolution and Psychology of Unselfish Behavior*, Harvard University Press, 1998.

[3] D. Bickerton, "Two neglected factors in language evolution", *The Evolution of Language*, World Scientific, 26-33, 2008.

[4] G. Ichinose and T. Arita, "The Role of Migration and Founder Effect for the Evolution of Cooperation in a Multilevel Selection Context", *Ecological Modelling*, 210(3), 221-230, 2008.

[5] M. A. Nowak, "Evolutionary biology of language", *Philosophical Transactions, Biological Sciences The Royal Society*, 355(1403), 1615-1622, 2000.

[6] D. S. Wilson, "Structured Demes and the Evolution of Group Advantageous Traits", *American Naturalist,* 111(977), 157-185, 1977.

