Convergence Analysis for Collective Vocabulary Development *

Jun Wang
Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign
junwang4@uiuc.edu

Les Gasser
Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign
gasser@uiuc.edu

Jim Houk
Department of Physiology
Northwestern University Medical School
Chicago, IL
j-houk@northwestern.edu

ABSTRACT

We study how decentralized agents can develop shared vocabularies without global coordination. Answering this question can help us understand the emergence of many communication systems, from bacterial communication to human languages, as well as helping to design algorithms for supporting self-organizing information systems such as social tagging or ad-word systems for the web. We introduce a formal communication model in which senders and receivers can adapt their communicative behaviors through a type of *win-stay lose-shift* adaptation strategy. We find by simulations and analysis that for a given number of meanings, there exists a threshold for the number of words below which the agents can't converge to a shared vocabulary. Our finding implies that for a communication system to emerge, agents must have the capability of inventing a minimum number of words or sentences. This result also rationalizes the necessity for syntax, as a tool for generating unlimited sentences.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Multiagent systems, Languages and structures, Coherence and coordination*

Keywords

self-organizing vocabularies, win-stay lose-shift

1. INTRODUCTION

We study the question of how decentralized agents can develop a shared vocabulary without global coordination. This is a fundamental question that underpins how we understand the emergence of many communication systems arising in nature and in human society, e.g., how do different organisms develop the communication needed to coordinate their activities, how do encoding/decoding networks arise in the brain, or how did human languages emerge? The study of this question can also help us design algorithms that support self-organizing information systems such as social tagging, web ad-words, or peer-to-peer retrieval systems.

Two general frameworks, *evolutionary* and *self-organizing*, have been proposed to study the emergence of communication systems. Evolutionary frameworks assume that vocabularies of agents are inherited genetically and/or learned culturally from parents or teachers. The driving force for the emergence of shared vocabularies comes from mechanisms in which good communicators produce more offspring with similar vocabularies[2, 6, 7]. In contrast, in self-organizing frameworks decentralized agents actively develop shared vocabularies without global knowledge in a short span of time by changing their own representations via distributed learning and positive feedback loops[1, 3, 4, 8]. The work reported here can be seen as an effort in the self-organizing camp.

Most existing work on the development of shared vocabularies has been done using computer simulations and only intuitive interpretations. While this is often suggestive, to achieve a thorough understanding we need to augment simulation studies with mathematical characterizations of specific conditions and limitations under which the emergent behaviors are stable. Recent work by Baronchelli et al.[1] provides one example; they explain how a sharp transition to a shared vocabulary observed in simulations can occur as the density of certain inter-agent vocabulary hypotheses grows beyond a "tipping point."

In this paper, we focus on a formal communication model in which senders and receivers can adapt their communicative behaviors through a type of win-stay lose-shift learning strategy[5]. To simplify the characterization of the dynamics, we study a model that consists of two agents. Through simulations and analysis, we give the conditions under which the agents can develop a stable shared vocabulary.

2. WIN-STAY LOSE-SHIFT ADAPTATION MODEL

Our communication model consists of two agents, called a *sender* and a *receiver*. In our model, there are m meanings $\{x_1, \cdots, x_m\}$, and n words $\{y_1, \cdots, y_n\}$. The sender and

*Work supported by NSF Grant IIS-0340996 and UIUC Beckman Summer Fellowship.

AAMAS'06 May 8–12 2006, Hakodate, Hokkaido, Japan.

The sender's dictionary **The receiver's dictionary**

Round 0

meaning	word	wins
x_1		0
x_2		0
x_3		0

word	meaning	wins
y_1		0
y_2		0
y_3		0

Round 1 (play game on meaning x_1, successful—a luck.)

meaning	word	wins
x_1	y_1	1
x_2		0
x_3		0

word	meaning	wins
y_1	x_1	1
y_2		0
y_3		0

Round 2 (play game on meaning x_1, successful.)

meaning	word	wins
x_1	y_1	2
x_2		0
x_3		0

word	meaning	wins
y_1	x_1	2
y_2		0
y_3		0

Round 3 (play game on meaning x_2 via word y_1, failed.)

meaning	word	wins
x_1	y_1	2
x_2		0
x_3		0

word	meaning	wins
y_1	x_1	1
y_2		0
y_3		0

Round 4 (play game on meaning x_3 via word y_1, failed.)

meaning	word	wins
x_1	y_1	2
x_2		0
x_3		0

word	meaning	wins
y_1		0
y_2		0
y_3		0

Figure 1: An illustration of repeated play of communication games.

receiver each have a dictionary, consisting of a list of entries. The sender's entry is of the format $\langle meaning, word\rangle$ or $\langle x, y\rangle$. Each entry is indexed by a meaning; it indicates that for each meaning there is one and only one entry. For each entry $\langle x, y\rangle$, there is an associated number that indicates how many more times the sender has successfully communicated the meaning x via the word y than he has failed to do. Such a number is called the *wins* of the entry, and denoted by $wins_s(x)$, where the subscript $_s$ indicates the sender. We require $wins \geq 0$. When there is no win (i.e., $wins = 0$), the word y will be empty. Initially, all entries have no win.

Similarly, the format $\langle word, meaning\rangle$ or $\langle y, x\rangle$represents the receiver's entry. Each entry is indexed by a word; it indicates that for each word there is one and only one entry. For each entry $\langle y, x\rangle$, there is an associated number that indicates how many more times the receiver has successfully interpreted the word y to the meaning x than he has failed to do. Such a number is called the *wins* of the entry, and denoted by $wins_r(y)$, where the subscript $_r$ indicates the receiver. When there is no win (i.e., $wins = 0$), the meaning x will be empty. Initially, all entries have no win.

A communication event is initiated by the sender who tries to communicate a meaning to the receiver. The meaning that the sender wants to communicate is drawn according to some probability distribution. (In this short paper, we will assume it is uniform distribution.) To communicate a meaning, say, x, the sender needs to represent it as a word. The sender looks up the meaning x in his dictionary. Let the found entry be $\langle x, y\rangle$. If $wins_s(x) > 0$, then the sender will use word y to represent the meaning x. If $wins_s(x) = 0$, which means the word y is empty, then the sender will randomly choose a word to represent the meaning. (This is the idea of *win-stay lose-shift.*)

When receiving a word, say, y, the receiver looks up the word in her dictionary. Let the found entry be $\langle y, x'\rangle$. If $wins_r(y) > 0$, then the receiver will use the meaning x' to interpret the word y. If $wins_r(y) = 0$, which means the meaning x' is empty, then the receiver will randomly choose a meaning as the interpretation to the word.

If the interpreted meaning x' is correct (we suppose *successful task performance* can provide feedback on communicative success), i.e., $x' = x$, we say the communication is successful; otherwise, the communication is failed. When the communication succeeds, the sender will update the *wins* of the entry indexed by x according to the reward rule: $wins_s(x) \leftarrow wins_s(x) + 1$. If the word field is empty he will replace it with the word y. Similarly, the receiver will update the *wins* of the entry indexed by y according to the rule: $wins_r(y) \leftarrow wins_r(y) + 1$. If the meaning field is empty she will replace it with the meaning x.

When the communication fails, the sender will update the *wins* of the entry indexed by x according to the rule: $wins_s(x) \leftarrow wins_s(x) - 1$ if $wins_s(x) > 0$. If $wins_s(x) = 0$ he will replace the word field with empty. Similarly, the receiver will update the *wins* of the entry indexed by y according to the rule $wins_r(y) \leftarrow wins_r(y) - 1$ if $wins_r(y) > 0$. If $wins_r(y) = 0$ she will replace the meaning field with empty.

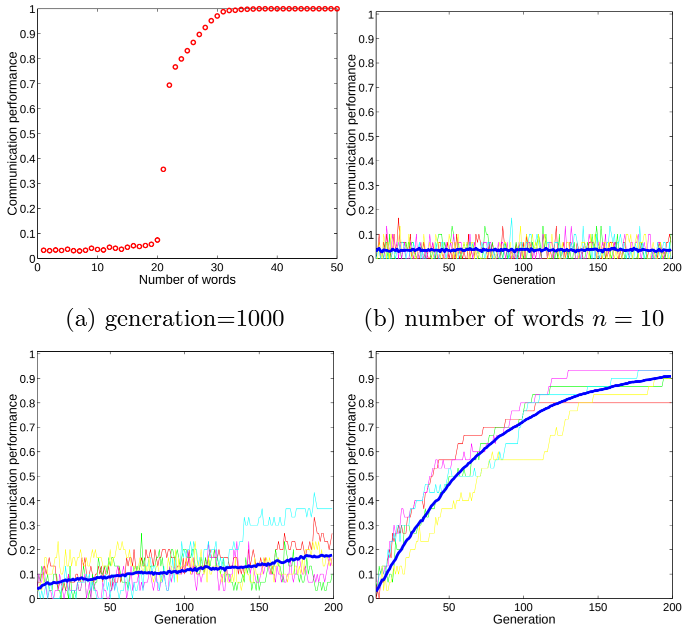
We call one communication event between the sender and receiver a *communication game*, and thus we model the vocabulary development process as repeated play of communication games (Fig. 1).

3. SIMULATIONS

In this section, we show by simulations that for a given number of meanings, there exists a threshold of the number of words below which the agents can't develop a shared vocabulary using the win-stay lose-shift adaptive communication model. In Fig. 2(a), we can see that there exists a dramatic *phase transition* when the number of words n is around a *threshold* $n^* = 21$. When n is above the threshold, the two agents can develop a communication system with converged communicative performance given by $\min\{\frac{n}{m}, 1\}$; that is, the sender and receiver can develop $\min\{n, m\}$ meaning-word agreements. However, when the number of words is below the threshold, it rapidly approaches random guess—the sender and receiver can't develop any effective communication system, or a shared vocabulary. The three graphs (c,d,e) in Fig. 2 illustrate the dynamics of communication over time for three specific settings of the number of words: $n = 10$, $n = 21$, and $n = 40$.

4. A SIMPLE ANALYSIS

In this section, we give a simple analysis aiming to understand why the above phase transition phenomenon happens. The basic idea is that soon or later the sender and receiver will form a temporary meaning-word agreement. Then, in the following rounds, this temporary agreement may be either reinforced or weakened. If the temporary agreement has a better chance to be reinforced than to be weakened, random walk theory tells us that the temporary agreement will have a positive probability of becoming a permanent agreement. We will show that the probability for a temporary agreement to become permanent depends on the relation between the number of words and the number of meanings. When the number of meanings is fixed, the more the words,



(a) generation=1000 (b) number of words $n = 10$

(c) number of words $n = 21$ (d) number of words $n = 40$

Figure 2: (a) Communication performance (success ratio) as a function of the number of words n when the number of meanings is held constant at $m = 30$. A dramatic phase transition occurs when the number of words n is around a threshold $n^* = 21$, above which the agents can develop a communication system with eventual communication performance given by $\min\{\frac{n}{m}, 1\}$. The performance numbers, marked by red circles, are obtained by averaging the results from 100 runs (after 1000 generations, where each generation contains $m = 30$ rounds of play). (b,c,d) Dynamics of adaptive communication over time (only 200 generations are shown). The thick lines in the graphs are obtained by averaging the results of 100 runs, 5 of which are shown as thin lines. All meanings are uniformly distributed.

the better the chance for a temporary agreement to become permanent. (See Fig. 1 for a better understanding of the following analysis.)

For simplicity, we will assume that all meanings are uniformly distributed. Since we have m meanings, so given a meaning the probability that it is chosen for play is $\frac{1}{m}$.

Without loss of generality, let a temporary agreement be on the meaning-word pair (x_1, y_1). With probability $\frac{1}{m}$, at the next round the sender will play game on the meaning x_1. Clearly the communication on meaning x_1 will succeed, so the *wins* value of the sender and receiver, $wins_s(x_1)$ and $wins_r(y_1)$, will increase by 1. Therefore we have a *reinforcement probability* of $\frac{1}{m}$ that the temporary agreement will be reinforced.

With probability $1 - \frac{1}{m}$, at the next round the sender will play game on a meaning other than x_1, say, x_2. If the *wins* value of x_2 in the sender's dictionary is zero, i.e., $wins_s(x_2) = 0$, then the sender will randomly choose a word from the n words to represent the meaning x_2, and so there is a probability of $\frac{1}{n}$ that word y_1 will be used to represent the meaning x_2. If this happens, the communication will fail, because the receiver will interpret word y_1 as meaning x_1. And then, the *wins* of the word y_1 in the receiver's dictionary, $wins_r(y_1)$, will decrease by 1. In summary, we have a *weakening probability* of $(1-\frac{1}{m})\frac{1}{n}$ that the temporary agreement will be weakened.

If the reinforcement probability, $\frac{1}{m}$, is larger than the weakening probability, $(1 - \frac{1}{m})\frac{1}{n}$, the temporary agreement will become stable. Therefore, by solving the following inequality

$$\frac{1}{m} > (1 - \frac{1}{m})\frac{1}{n}$$

we can see that when $n > m-1$, the sender and receiver can guarantee to establish a permanent agreement. Note that if in the above analysis we consider the case of $wins_s(x_2) > 0$, then the weakening probability will be even lower, and thus the condition $n > m-1$ can be relaxed. This partly explains why in the above simulations the threshold value is $n^* = 21$ rather than $n^* = m - 1 = 29$.

Clearly, the two agents can develop as many as $\min\{m, n\}$ stable agreements, since there are at most $\min\{m, n\}$ distinct pairs of meaning-word. And thus, the maximum converged communicative performance should be $\min\{m, n\}\frac{1}{m}$ or equivalently $\min\{\frac{n}{m}, 1\}$.

5. CONCLUSION

We have presented an adaptive communication model in which a sender and a receiver can adapt their communicative behaviors through a simple learning strategy—win-stay lose-shift. By computer simulations and mathematical analysis, we find that for a given number of meanings, there exists a threshold for the number of words below which the agents can't develop a shared vocabulary. Our finding shows that for a communication system to emerge, agents must be able to invent a minimum number of words or sentences. This result also rationalizes the necessity for syntax, as a tool for generating unlimited sentences.

Though this analysis treats just two communicating agents, we believe it can be extended to the case of multiple agents—one sender and many receivers, many senders and one receiver, or many senders and many receivers.

6. REFERENCES

[1] A. Baronchelli, M. Felici, E. Caglioti, V. Loreto, and L. Steels. Sharp transition towards shared vocabularies in multi-agent systems. http://arxiv.org/abs/physics/0509075, 2005.

[2] J. Hurford. Biological evolution of the saussurean sign as a component of the language acquisition device. *Lingua*, 77(2):187–222, 1989.

[3] F. Kaplan. Semiotic schemata: Selection units for linguistic cultural evolution. In M. Bedau and et al., editors, *Artificial Life VII*. The MIT Press, 2000.

[4] T. Lenaerts, B. Jansen, K. Tuyls, and B. de Vylder. The evolutionary language game: An orthogonal approach. *Journal of Theoretical Biology*, 235(4):566–582, August 2005.

[5] F. A. Matsen and M. A. Nowak. Win-stay, lose-shift in language learning from peers. *PNAS*, 101(52):18053–18057, December 2004.

[6] M. A. Nowak, J. B. Plotkin, and D. C. Krakauer. The evolutionary language game. *Journal of Theoretical Biology*, 200:147–162, 1999.

[7] K. Smith. The evolution of vocabulary. *Journal of Theoretical Biology*, 228(1):127–142, 2004.

[8] L. Steels. Self-organizing vocabularies. In C. Langton and T. Shimohara, editors, *Artificial Life V*, Nara, Japan, 1996.