Conventionalization of Linguistic Categories under Simple Communicative Constraints

Tao Gong, Andrea Puglisi, Vittorio Loreto, and William S-Y. Wang

***Abstract*—The language game approach is widely adopted to study conventionalization of linguistic knowledge. Most of contemporary models concentrate on the dynamics of language games in random or predefined social structures, but neglect the role of communicative constraints. This paper adopts one form of language games, the category game, to discuss whether some simple distance-related communicative constraint may affect the conventionalization of linguistic categories. By comparing the simulation results with those based on another form of language games, the naming game, we point out some essential differences between these two games which cause their distinct performances under the same communicative constraint. This study fills the gap between the dynamics of language games in random structures and that in complex networks, and suggests that internal properties of language games may reversely influence communicative constraints and social structures.**

I. Introduction

OUR social behaviors, such as language, have evolved primarily via contact with a limited number of other individuals, and a prerequisite for successful language use is conventional linguistic knowledge [1]. But where do these conventions come from, and how do they evolve through iterative communications? Many theoretical and practical studies have discussed these questions to better understand the evolution of human communication system (e.g., [2]–[7]), among which *the language game approach* [8] serves as an efficient method to study *conventionalization* (a process of social agreement by conforming one's language to that of the others or the community [9]) of various linguistic components during language evolution.

This approach views language as a *Complex Adaptive System* [11][12]. It usually postulates a population of agents and an interaction protocol between them, called a *language game* [10]. Using this game, agents can carry out some communicative task, such as drawing another's attention to an object in their surroundings. Agents typically establish a communication system from scratch, by inventing new forms of conceptualization and/or expression, and adjusting their available knowledge based on its utility, frequency, or social prestige. Through a number of iterated communications among agents, a set of linguistic knowledge gradually become conventional in the population, and further statistical analyses can examine the dynamics of this game and provide some quantitative understanding of language evolution.

Many forms of language games have been proposed to study conventionalization of various types of linguistic or general cognitive knowledge. For instance, *the naming game* [13] studied the emergence of coherence, in which, to draw attentions to each other, speakers and listeners create various names for a particular object, and once they successfully use a common name for that object, they will eliminate all other competing names, but leave the common one. Otherwise, the listener will acquire the speaker's name as a candidate in his/her inventory. After a number of naming games, the population converges to a common object-name pair. *The category game* [14] (which will be reviewed in Section 2.A) extended the naming game by allowing agents to categorize in a continuous semantic space and create corresponding word labels for their categories. Through iterated category games, a set of linguistic categories can be gradually shared in the population.

In addition to many successful strategies reviewed from these language games, statistical physicists further explored the dynamics of these games by putting agents into random networks, 1D/2D lattice, or other complex structures like scale-free [15] and small-world [16] networks (e.g., [14][17][18]). These studies convincingly discussed the role of social structure in converging linguistic knowledge, but most of them neglected the reverse role of language games in social structure; since in these studies, a successful or failed language game does not affect individuals' predefined social connections.

As a social phenomenon, linguistic interactions can affect not only participants' knowledge but also their social connections. Accumulative failed or successful interactions may weaken or strengthen social connections. During interactions, local communicative constraints, such as geographical or social distance, may adjust the probabilities for different agents to interact with each other, thus affecting individual or group similarities on a global scale [19][20]. These constraints, taking place much earlier than the emergence of complex social structures, could cast their influence on formation of mutual understanding and social structures. Without a careful study of these simple constraints and their roles on language evolution and social structures, it seems premature to proceed from random networks directly to complex structures, since different language games may

Manuscript received December 12, 2007. This work was supported in part by the EU under RD contract IST-1940 (ECAgents).

Tao Gong is with Physics Department, "La Sapienza" University, Rome, ITALY (phone: 39-06-49913450; fax: 39-06-4463158; e-mail: gtojty@gmail.com).

Andrea Puglisi is with Physics Department, "La Sapienza" University, Rome, ITALY (e-mail: Andrea.Puglis@roma1.infn.it)

Vittorio Loreto is with Physics Department, "La Sapienza" University, Rome, ITALY (e-mail:Vittorio.Loreto@roma1.infn.it).

William S-Y. Wang is with Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong, CHINA (e-mail: wsywang@ee.cuhk.edu.hk).

lead to different social connections, and the same structure may not always trigger similar effects on conventionalization based on different games.

Considering these, we present in this paper a simulation study exploring conventionalization of linguistic knowledge under simple communicative constraints. As an extension to our previous work [21] examining conventionalization of lexical knowledge based on the naming game, this paper concentrates on conventionalization of linguistic categories based on the category game. Through comparing the simulation results with those based on the naming game, we notice some distinct performances of these games under the same communicative constraint. The essential differences in the communicative tasks involved in these games could explain these different performances, which indicate that the internal properties of language games could influence the effect of communicative constraints and further affect the role of social structures. These findings will guide the future study on the dynamics of language games in complex networks, and contribute to the discussion on the mutual influence between communication and social structures.

The rest of the paper is organized as follows: Section 2 briefly reviews the category game, and introduces the distance-related communicative constraint; Section 3 discusses the simulation results of the two experiments under this constraint; Section 4 compares the naming game with the category game, and illustrates their different ways of conventionalization; and finally, Section 5 gives the conclusions.

II. The Category Game and Distance Constraint

A. The Category Game Review

Categories are fundamental to recognize, differentiate and understand the environment. Semantic categories, like other linguistic components, are culture-dependent conventions shared among a group of individuals [22], and the emergence of them may undergo a self-organization process via iterated interactions among individuals [23]. Different individuals may perceive or conceptualize the world differently, and establish different semantic categories. Through interactions, the boundaries of these categories tend to align, and their word labels tend to converge. Then, individuals can understand each other to a certain degree. The category game was proposed to theoretically simulate this alignment and convergence process, and demonstrate conventionalization of linguistic categories having both similar semantic ranges and common lexical labels.

In the model of the category game, N individuals are given stimuli from a single analogical perceptual space, these stimuli are represented by real-valued numbers ranging from 0.0 to 1.0. A *perceptual categorization* is a partition of the interval [0.0, 1.0] into discrete sub-intervals. Each agent has a dynamical inventory of word-meaning associations linking perceptual categories (meanings) with labels (words). All agents are initialized with a trivial perceptual category [0.0 1.0] and no words. Through pair-wise category games, the perceptual categories and their associated words coevolve dynamically among agents in the population.

In a category game, two agents are randomly selected, one as the speaker and the other as the listener (hearer). Meanwhile, M ($\geqslant$2) stimuli from the perceptual channel are presented. The speaker discriminates these stimuli, and names one of them (topic). Then, the listener tries to guess the topic from other contextual stimuli based on the word produced by the speaker and his/her own inventory of perceptual categories. An individual's discrimination ability is restricted by his/her perceptual power, which is denoted by d_{min}, i.e., the minimal numerical distance required for discriminating two stimuli. In a single game, the minimal distance between any presented stimuli is d_{min}, but in general, stimuli can take any numerical value, constrained only by the numerical precision used in the experiment. In a category game, both participants' categories and the words associated to these categories can be adjusted, which are illustrated by two examples of the category game shown in Fig. 1.

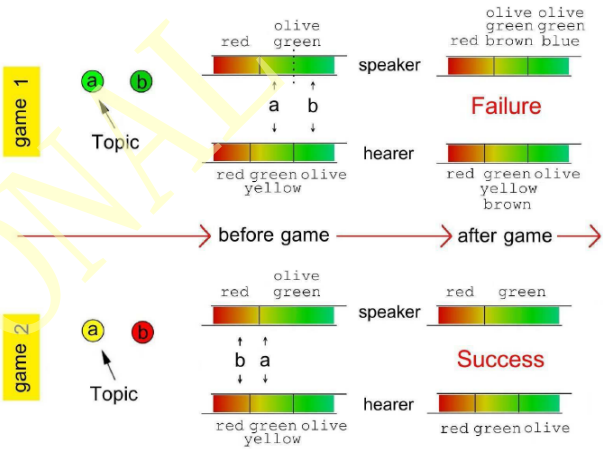

Fig. 1. Two examples of the category game (adapted from [14]). The round objects are stimuli presented in these games, among which the topics are pointed. The colorful banners represent individuals' perceptual spaces, and different individuals use different bars to partition their spaces into different perceptual categories, whose inventories of words are listed above or below.

In game 1, since the two stimuli fall into the same perceptual category, the speaker discriminates the chosen topic ("a") by creating a new boundary in his/her rightmost perceptual category at the position (a+b)/2. Then, two new categories are created, both inheriting the word-inventory ("green" and "olive") of their parent category, and a new word is invented for each of these new categories ("brown" and "blue"). Then, the speaker browses the list of words associated to the category that contains the topic. There are two cases here: if a previous successful game occurred with this category, its last winning word is chosen; otherwise, the newly created word ("brown") is chosen, and sent to the listener. Since the listener does not have this word in his/her inventory, this category game fails. Then, the speaker points at the topic, and the listener discriminates the topic, and adds the speaker's word to the inventory of his/her corresponding category.

In game 2, these two agents carry out another category game. The speaker chooses the topic "a", which is already