Games and Economic Behavior 102 (2017) 255–268

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

# The emergence of compositional grammars in artificial codes [☆]

Fuhai Hong [a], Wooyoung Lim [b,*], Xiaojian Zhao [b]

[a] *Division of Economics, Nanyang Technological University, Singapore*
[b] *Department of Economics, Hong Kong University of Science and Technology, Hong Kong*

ARTICLE INFO

*Article history:*
Received 4 November 2013
Available online 10 January 2017

*JEL classification:*
C72
C92
D03
D83

*Keywords:*
Communication games
Economics of language
Experimental semiotics
Compositional grammar

ABSTRACT

This paper experimentally explores how compositional grammars in artificial codes emerge and are sustained. In a communication game with no conflict of interest, the sender sends a message that is an arbitrary string from available symbols with no prior meaning to indicate an abstract geometrical figure to the receiver. We find strong evidence from the laboratory for the emergence of compositional grammars in the subjects' common codes that facilitate learning efficiency. Moreover, when there is a scarcity of symbols in the repertoire, a few groups in our experiments developed languages with positional compositionality, meaning the same symbol has different interpretations depending on its position in a string, whereas some other groups developed language structures that are not compositional but still efficient in communication.



## 1. Introduction

The key component of human language that enables us to efficiently transmit potentially complex information is its grammar. Despite the diversity in natural languages, economic theorists typically argue that there are common properties that guide the emergence and development of languages. Rubinstein (1996) is among the first to use the economic approach to derive linguistic properties. Rubinstein (2000) asserts that language is a mechanism of communication, and because economic theory analyzes the design of social systems, it must be relevant to linguistic issues as well. As tools of communication, languages arguably possess certain common properties that maximize efficiency in communication.[1] Although the theoretical advancement in this line of research has been well documented, the empirical test of the claimed commonality of linguistic properties is still in its infancy.

☆ We are deeply indebted to Andreas Blume for his constructive comments that improved the paper immeasurably. We would also like to thank Vincent Crawford, the advisory editor, three anonymous referees, Te Bao, Andreas Ortmann, Joseph Tao-yi Wang, Eric van Damme, Yang Yang, and Songfa Zhong, the conference participants at the 2nd Behavioral and Experimental Economics Workshop (Chengdu), the Annual Congress of the European Economic Association (Toulouse), the Asian Meeting of the Econometric Society (Taipei), the ESA World Meeting (Sydney), the Fudan University Workshop on Behavioral Science (Shanghai), the International Conference on Economics and Econometrics (Hiroshima), and seminar participants at the City University of Hong Kong, Hong Kong University of Science and Technology, Nanyang Technological University, Shandong University, Shanghai University of Finance and Economics, Singapore Management University, Southwestern University of Finance and Economics, Tsinghua University and Wuhan University for their helpful comments. This study is supported by a grant from the Research Grants Council of Hong Kong (Grant No. ECS-699613).

* Corresponding author.
*E-mail addresses:* fhhong@ntu.edu.sg (F. Hong), wooyoung@ust.hk (W. Lim), xjzhao81@gmail.com (X. Zhao).

[1] See also the follow-up survey by Lipman (2003).

http://dx.doi.org/10.1016/j.geb.2016.12.009
0899-8256/

In this paper, we explore, on an experimental basis, how the grammar-like structure of language emerges and present evidence from the laboratory. We focus on the emergence of a linguistic property called *compositionality* (Frege, 1882), which appears to be universal in natural languages. The compositionality of languages suggests that, as strings of symbols, messages should be decomposable into coordinate functions whose domains are the factors of the state space.[2] The consideration of compositionality is associated with the notion of learning efficiency, which implies that available symbols are used such that individuals are able to learn the language efficiently, i.e., from a minimal number of observations. Blume (2005) shows that compositional languages are efficient in learning.[3]

We study how compositionality arises to facilitate communication efficiency in the following simple environment. Two players, a sender and a receiver, play a communication game with no conflict of interest. The sender, who is privately informed about the underlying state of nature, sends a message to the uninformed receiver, who then identifies the realized state. If, and only if, the receiver correctly identifies the state, both receive a positive payoff. The messages are strings of symbols with arbitrary lengths, where the available symbols have no a priori meaning.

In our laboratory implementation, we considered a state space with a $3 \times 2 \times 3$ product structure. We used abstract geometrical figures to denote the states of nature and allowed individuals to communicate using abstract symbols with no prior meaning to minimize the impact of any existing communication convention among the participants. Specifically, we used shapes such as ◯, △ and □ to denote objects and used vertical and horizontal orderings of the shapes to denote the spatial relations between objects.[4] The number of available symbols in a message string was our main treatment variable. In our first set of treatments, senders were allowed to send a message consisting of a string with arbitrary length from five available symbols !, @, #, $ and %. In our second set of treatments, there were three available symbols !, $ and % only.[5]

With the above symbols, however, there may be focal associations between certain symbols and certain shapes, e.g., @ may be focal to be used to represent ◯. These focal associations may affect the speed of learning. To address this issue, we designed robustness treatments in which the sender's message was privately translated into a message with English letters, following some randomly determined rule, before being sent to the receiver. This *privatized translation of message* protocol is initially developed by Blume et al. (1998) to remove all conceivable language conventions that may be shared among players.

We obtained strong evidence for the emergence of compositional grammars. Most groups in our first set of treatments (with five symbols) and more than half of the groups in our second set of treatments (with three symbols) developed compositional languages. We also observed that some symbols were never used in some groups because compositionality saved the use of additional symbols even if using them is costless. Interestingly, three groups in the second set of treatments developed languages with the property of "positional" compositionality, meaning that the same symbol has different interpretations depending on its position in a string. More importantly, compositionality indeed facilitated learning efficiency. Most groups learned the entire language based on a few observations in the beginning rounds. That is, participants were indeed able to rely on the structure of language to successfully communicate states they have not previously communicated. Moreover, the groups that did not develop compositional languages in the second set of treatments developed certain language structures as well that also facilitate communication. We found that most of our qualitative findings are robust to the introduction of the privatized translation of message.

The paper is organized as follows. The rest of this section reviews the related literature on language and economics. In Section 2, we discuss preliminaries that define the compositionality of languages and link it to the notion of learning efficiency. Section 3 introduces the experimental design and procedure. Experimental results are reported in Section 4. Concluding remarks are presented in the last section.

### 1.1. Related literature

There is a small but growing literature in experimental economics on the emergence and evolution of language. Among others, the studies by Blume et al. (1998) and Selten and Warglien (2007) are most closely related to ours.[6]

Blume et al. (1998) experimentally explore the endogenous emergence of meanings in various sender–receiver games with two types considered by Lewis (1969). The authors are pioneers in providing a fully controlled laboratory environment

[2] This is the definition of compositionality considered in the economics literature (refer to, e.g., Blume, 2005; Selten and Warglien, 2007). We discuss broader definitions of compositionality available in the linguistics literature in Footnote 12 and their implications for our experimental results in Footnote 27.

[3] Blume (2000) shows that grammatical structures facilitate coordination and learning in an environment in which novel meanings should be expressed. In philosophy and psychological linguistics, Franke (2014, 2016) and Steinert-Threlkeld (2016) consider the simple sender–receiver games proposed by Lewis (1969) to theoretically show that the compositionality of languages can emerge for unsophisticated agents from a variant of reinforcement learning.

[4] In our experiment, we focused only on spatial relations, although in principle, relations can be more broadly demonstrated. Similarly, Kirby et al. (2008) introduce a variety of spatial relations and motions and study the origins of structure in human language. To study the non-intentional evolution of language structure, the authors focus on transmission errors in an iterated learning process via a diffusion chain, a standard methodology for studying information and culture transmission in experimental psychology.

[5] Due to technical difficulties, we were only able to implement arbitrary strings from the available symbols with any length of up to 20 characters in the laboratory. However, we argue that a length of up to 20 characters is sufficiently long because the number of states in our treatments was 18.

[6] See also Weber and Camerer (2003) on the emergence of a common code based on natural language in the context of organizational culture and mergers.