

Agent-based modelling using naming game for language evolution studies

Alexander Ioilyevich Ilyinsky^{1*}, *Galina Vladimirovna Klimova*^{2 †}, *Evgeniy Sergeevich Smakhtin*³, *Marina Aleksandrovna Amurskaya*³, and *Ekaterina Yurievna Rozhina*⁴

¹Financial University under the Government of the Russian Federation, Faculty of International Economic Relations, Moscow, Russia,

²Financial University under the Government of the Russian Federation, International Project Center, Moscow, Russia

³Financial University under the Government of the Russian Federation, Department of Foreign Languages and Intercultural Communication, Moscow, Russia

⁴Plekhanov Russian University of Economics, Chair of Foreign Languages N2, Moscow, Russia

Abstract. The article describes approaches to applying agent-based modelling and, particularly, the case of Naming Game, in linguistic studies and within teaching foreign languages. Computational modelling implementation has become a comprehensive and ambitious field of research, as its methods are applicable to solving tasks set within various aspects of contemporary society and science. The main purpose of this paper is to perform an analysis of Naming Game implementation in language emergence and evolution studies. To achieve this purpose we set several tasks: to present a vast literature review on agent-based modelling in linguistics and other adjacent sciences; to give an overview and description of the Naming Game; to perform simulations within the Naming Game and present their outcomes. As the main methodology the article uses simulations. The paper concludes that a clear hysteresis effect is present in the dependence of the size of the population vocabulary from the size of vocabulary of its average agent. At the point where the population vocabulary transitions into the uniform distribution the average agent's vocabulary reaches saturation and plateaus. Those dynamics also change as the population vocabulary grows and declines. Agent-based modelling is a relatively novel direction for linguistics with a modest number of research papers. Results, presented in the paper, give a fresh angle on the issues of language emergence and evolution.

Keywords: Naming Game, agent-based modelling, language emergence, language evolution

1 Introduction

* Corresponding author: ailyinsky@fa.ru

† Died July 25, 2021.

The Naming Game is studied comprehensively nowadays [1-4]. Being an interdisciplinary field of research, language evolution and emergence are investigated within various sciences. IT and computational modelling are gaining popularity rapidly in linguistics as well.

Research objective: to perform an analysis of Naming Game implementation in language evolution.

Methodology is based on simulations.

Processes operating within evolution of language as a complex adaptive system are known to strongly influence each other. By trying to analyze the emerging language, researchers hope that its similarity to human language evolution will lead to understanding main features typical for all languages, e.g. meaning-form mappings, origin of linguistic coherence, coevolutionary origin of grammar and meaning [5].

It is a common approach to distinguish two levels of studying and describing language. The first one represents individual language users. The population level, however, is established through effective communication of these individuals. Both levels are interconnected and interdependent [5].

Studying duration of the processes dealing with language emergence and evolution, scholars identify three basic time scales:

- ontogenetic time scale is characterized by the fastest dynamics at the individual level, including language acquisition and communication;
- glossogenetic time scale is represented by slower processes, e.g. migrations of language populations, dialect formation, language extinctions;
- phylogenetic time scale is defined by the slowest processes involved in the biological evolution of language users [5].

China is one of the countries where IT and various games have been developed and studied for the last decade. Practical implementation of the naming game while locating multiple information sources in social networks is addressed by Chinese scholars. The authors consider the naming game as the studies of shared lexicons development within a population of agents when they name an object under observation, and adopt this approach for selecting observations [6].

Guiyuan Fu and Weidong Zhang investigate the dynamics of the two-word naming game with biased assimilation, when the hearer may refuse to accept the word sent by the speaker. In the paper, a hearer is the only one who updates his memory with the speaker keeping it unchanged [7]. A Multi-Language Naming Game is applied to study the evolution of languages, assuming the agents are speakers of different languages who need a translator or interpreter to foster their communication [8].

Shi Xiaoming and Zhang Jiefang discuss common issues dealing with the emergence and development of Naming game models and present three stages of evolution:

- 1) the period of fast naming increase;
- 2) an active stage of communications;
- 3) a trend to balancing and naming decrease down to the minimum [9].

Pan Qihui et al. describe a model, which does not delete less frequently used names. The results show that main words used by a person to characterize an object or a phenomenon include at least one name, which is common for the majority of people representing a particular social group [10]. Models' implementation in marketing is discussed in the paper by Lu Yirun [11]. Huang Deheng et al. introduce a multi-layered affective model as a type of naming game, where agents are labelled with their own emotions towards objects and events emerging within a game [12]. Lin Boyu et al. study the application of naming game in political science, assuming that models can help solve

one of the main political problems, i.e. whether social groups can reach the same state without global outside control [13].

The Baldwin effect, well known within Darwinian approach, assumes the possibility of a selective pressure for the instinctive behavior evolution of a population to replace the so-called learned behavior, which, being beneficial, though stays rather costly [5]. Studying the Baldwin effect within a naming game, Lekvam et al. point out that, according to this theory, traits acquired and learned at the cultural level, can eventually become innate for a species [4].

An agent-based model application to study word order change in English is discussed in [14]. Implementation of agent-based modelling for the purposes of studying a historical word order change in Germanic languages is presented in the paper [15].

Focusing on the emergence of grammar through agent-based modelling, Luc Steels considers this issue as the final stage commonly recognized in language evolution research. Such aspects as word formation, syntax formation, and reduction, are addressed [16].

2 Methods

We investigate the dynamics of how the vocabulary of the population and the average vocabulary of an agent change using the minimal Naming Game model. Here, agents negotiate the names of objects in pairwise interactions without any centralized control. Agents are considered peers, in the sense that any agent can communicate with any other agent and obey the same rules of the model. To communicate, agents can use the words they already have in their vocabularies or create new ones. Vocabulary size is unlimited, however, the initial number of words prior to the first interaction is set at zero. Out of the population of N agents, two are selected at random. Each agent in the pair is assigned a role of either a hearer or a speaker. The speaker randomly chooses a word and communicates it to the hearer. If the hearer's vocabulary contains the word, both agents update their vocabularies to include only the word used in the interaction. In this scenario, their communication is designated as a success. If, on the other hand, the word is not in the hearer's vocabulary, the communication is a failure, and the hearer includes the word in the vocabulary [17].

To avoid confusion, homonymy is excluded by requiring that each newly invented word has never been used before.

3 Results

Results of simulating Naming Game within this paper can be seen below.

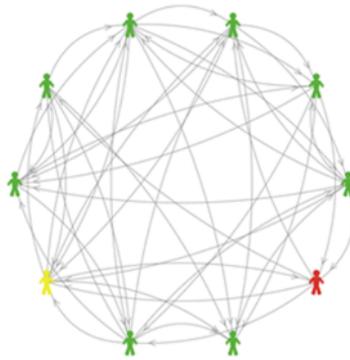


Fig. 1. Population of agents/vocabularies in Naming Game.

Fig. 1 demonstrates interaction procedure between agents within a population described above.

4 Discussion

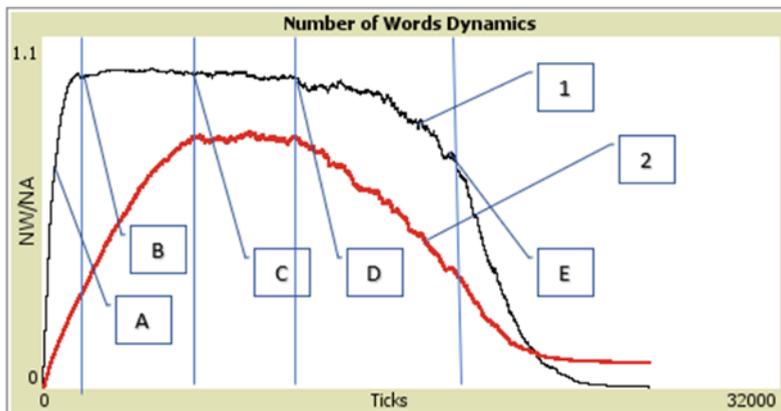


Fig. 2. Number of Words Dynamics.

Fig. 2 demonstrates dynamics of number of words, where:

Curve 1 is dynamics of a number of words within the population of agents, i.e. population's vocabulary.

Curve 2 is dynamics of an average number of words in a vocabulary of an agent, i.e. agent's vocabulary.

The Naming Game assumes several modes:

1) from the initial state to point A, we can see a simultaneous increase of numbers of different words (population's vocabulary) and this of an average word number in the vocabulary (agent's average vocabulary), as well as the linear connection between these values;

2) from point A to point B, there is a simultaneous increase population's vocabulary and that of an agent's average vocabulary, as well as non-linear connection between these values;

3) from point B to point C, a full set of words in the system is fixed while agent's average vocabulary is increasing;

4) from point C to point D, the size of both population's and agent's average vocabularies is fixed;

5) from point D to point E, there is a simultaneous reduction of population's vocabulary and that of an agent's average vocabulary, as well as non-linear connection between these values;

6) from point E to final state, we can see a simultaneous decrease of population's vocabulary and agent's average vocabulary, as well as linear connection between these values.

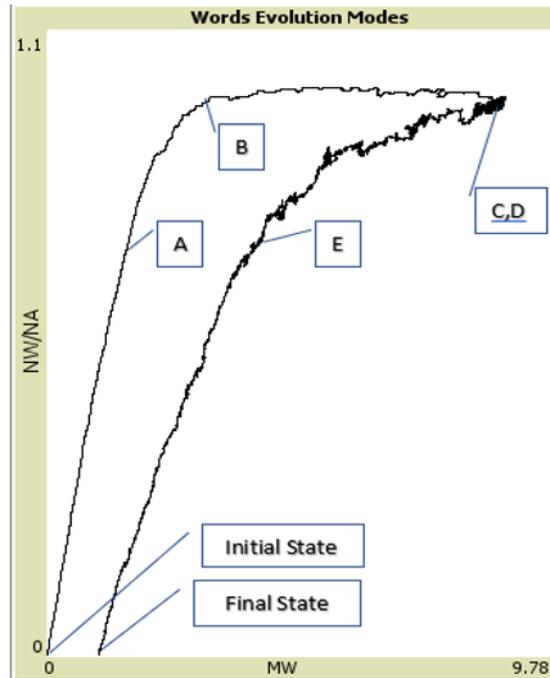


Fig. 3. Words Evolution Modes.

The phase plane of the Naming game model in Figure 3 shows the dependence of the vocabulary dynamics in the population (Y-axis) and the average agent (X-axis) on the direction of growth and decline in the richness of vocabulary. It allows to trace the hysteresis effect of the vocabulary dynamics:

1) Starting from the initial state and up to point A, an increase in average agent's vocabulary size by one word leads to an increase in the population's vocabulary by $0.5 \cdot NA$.

2) From point A to point B new word additions to the average agent's vocabulary lead to the increase in the population vocabulary albeit at a lower rate of NW/NA increase. Simultaneously, the vocabulary variety grows along the average agent's vocabulary and exhibit a nonlinear relationship with one another.

3) From point B to point C the population's vocabulary reaches the level of saturation and keeps a constant value with monotonous growth of the agent's average vocabulary.

4) From point C to point D the size of both the population's and the agent's mean vocabulary are fixed. Thus, the associated line segment degenerates into a point on the phase plane.

5) From point D to point E there is a simultaneous decrease in the population's vocabulary variety and the average agent's vocabulary with a non-linear relationship between these quantities.

6) From point E to the final state, a decrease in the average agent's vocabulary by one word leads to a decrease in the population vocabulary by $0.26 \cdot NA$.

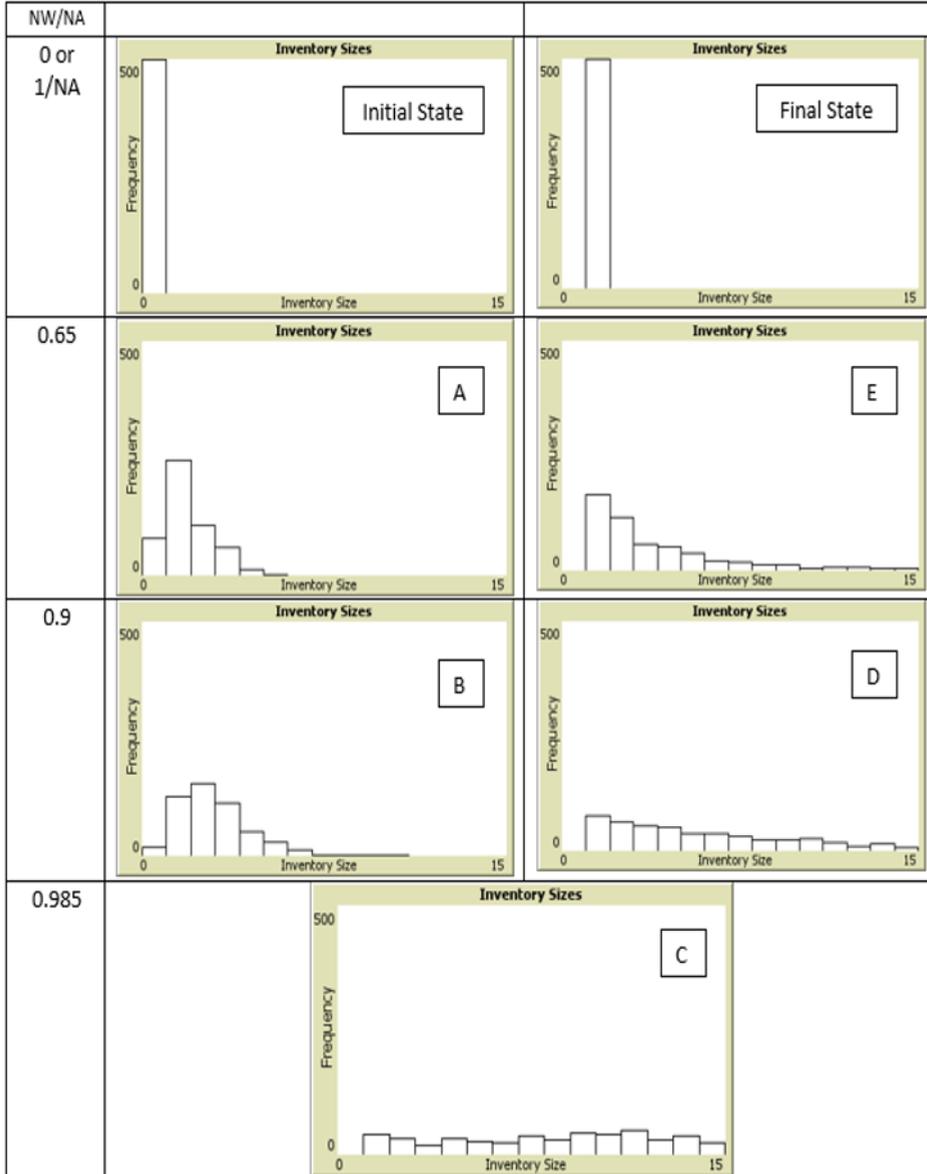


Fig. 4. Histograms of word distribution in the agents' population.

The X-axis shows the cells corresponding to the sets of 0, 1, 2, ..., 14 words in the agent's vocabulary, and the height of the bars shows the number of agents in the population with the specified number of words. Comments are as follows:

1) In the initial state, all agents have zero vocabulary. Therefore, the leftmost column shows the total number of agents in the system.

2) At point A, words have been created and distributed in the system due to the interaction of agents. The number of agents with zero vocabulary has decreased, and the maximum number of agents has a vocabulary of one word.

3) At point B, the probability density continues to transform. The maximum value shifts to the right and the packet becomes blurred. The number of agents with empty vocabulary decreases further.

4) At point C, the number of agents with zero vocabulary drops to zero, and the distribution takes a homogeneous form. Hence, the probabilities of picking an agent with a vocabulary ranging from 1 to 14 words are about the same.

5) At point D, the extinction of agents with a rich vocabulary and the growth in the number of agents with a smaller vocabulary begins. There are no agents with empty vocabulary in the system.

6.) At point E, the probability of picking an agent with a vocabulary of one word becomes dominant. The rate of extinction of agents with a rich vocabulary sharply increases. There are no agents with empty vocabulary in the system.

7) In the final state, the system reaches consensus. The vocabulary of the population reaches 1. So, all the agents use only one word to describe a given object. That word emerged in the system as the result of the agreement among all members of the population. Consensus in the population is achieved through only paired interactions between randomly chosen agents without any centralized influence on the system as a whole.

5 Conclusion

Thus, while implementing the Naming Game model, we were able to define the modes where the population's vocabulary and agent's average vocabulary are increasing and decreasing as a result of pairwise interactions in the system and the diffusion and compression of the probability distributions of agents' vocabularies.

In the transition to uniform distribution of agents' vocabularies, the curves representing the population and average agent vocabularies show a saturation plateau, which leads to degeneration of saturation plots C – D into a point on the phase plane.

The curves that show the dependence of the population vocabulary on the average vocabulary of an agent clearly demonstrate the effect of hysteresis and significant differences in dynamics with the growth and decrease of the population vocabulary.

References

1. A. Baronchelli, V. Loreto, L. Steels, *Int J Mod Phys C*, **19(05)**, 785-812 (2008).
<https://doi.org/10.1142/S0129183108012522>
2. G. Marchetti, E. Heinsalu, M. Patriarca, *Chaos*, **30**, 063119 (2020).
<https://doi.org/10.1063/5.0009569>
3. J. Nevens, P. Eecke, K. Beuls, *A Practical Guide to Studying Emergent Communication through Grounded Language Games*, in "Language Learning for Artificial Agents (L2A2) Symposium" of the 2019 Artificial Intelligence and Simulation of Behaviour (AISB) Convention Falmouth University, Cornwall, UK, 16-18 April, 2019 (2020)
4. T. Lekvam, B. Gambäck, L. Bungum, *Agent-based modeling of language evolution*, in Proceedings of the 5th Workshop on Cognitive Aspects of Computational Language Learning (CogACLL), 49-54, April 2014, Gothenburg, Sweden (2014).
<https://doi.org/10.3115/v1/W14-0510>

5. D. Lipowska, *Comp Methods Sci Techn*, **17(1-2)**, 41-51 (2011).
<https://doi.org/10.12921/cmst.2011.17.01.41-51>
6. X. Yang, et al., *Physics Let A*, **384(35)**, 126908 (2020).
<https://doi.org/10.1016/j.physleta.2020.126908>
7. G. Fu, W. Zhang, *Physica A: Stat Mech Applic*, **490**, 260-268 (2018).
<https://doi.org/10.1016/j.physa.2017.08.016>
8. J. Zhou, et al., *Physica A: Stat Mech Applic*, **496**, 620-634 (2018).
<https://doi.org/10.1016/j.physa.2017.12.124>
9. X. Shi, J. Zhang, *J Zhejiang Normal Univ*, **33(2)** (2010).
<https://doi.org/10.3969/j.issn.1001-5051.2010.02.009>
10. Q. Pan, et al., *J Dalian Univ Techn*, **54(3)** (2014)
11. Y. Lu, *Shanghai Packaging*, **04** (1996)
12. H. Deheng, Zh. Changle, J. Min, *Mind Calc*, **6(1)** (2012)
13. B. Lin, et al., *Naming Game on Small World Networks: The Influence of Cluster Structure*, in *Proceedings of the 2006 National Conference on Complex Networks* (2006)
14. M. Beeksmā, et al., *Comp Linguis Netherlands J*, **7**, 17-38 (2017)
15. J. Bloem, A. Versloot, F. Weerman. *An agent-based model of a historical word order change*, in *Proceedings of the Sixth Workshop on Cognitive Aspects of Computational Language Learning*. Lisbon, Portugal September 2015, 22-27 (2015).
<https://doi.org/10.18653/v1/W15-2404>
16. L. Steels, *Phil Trans R Soc*, **371(1701)** (2016). <http://doi.org/10.1098/rstb.2015.0447>
17. A. Baronchelli, *Belg J Linguist*, **30(1)**, 171-192 (2016).
<https://doi.org/10.1075/bjl.30.08bar>