

Social Computational Model of Language Endangerment and Recovery

YUN Jian¹, LI Xiao-Tong², LIU Shuang^{1,*} and WANG Chun-Xia¹

¹Department of Computer Science and Technology, Dalian Nationalities University, Dalian Liaoning 116600, China;

²School of Computer Science and Technology, Jilin University, Changchun Jilin 130012, China

Abstract: How to protect the endangered languages is a typical and important social problem around the world. Now, the existing social computational models of language evolution are generally short of language endangerment warning and language recovery. These two parts are very important for language evolution. In this paper, an artificial society is cultivated based on multi-agents with purpose of observing language endangerment and its recovery phenomenon. In view of endangerment factor of population language, resilience inner protective factor and resilience composite protective factor, corresponding social computational model is given. From experimental results we can see that the computational model draws the endangerment trend of the weaker one in the language competition and the recovery process of weaker endangered language under the effect of resilience composite protective factor. To break through research bottleneck for high cost of language endangerment experiment in real society, the proposed computational model is working in an artificial society. The introduction of artificial society plays an important and practical role in the design and evaluation of endangered language protection scheme.

Keywords: Language competition, endangered language protection, agents, social computing, artificial society.

According to data provided by Ethnologue, authoritative publication journal gathering language information all over the world, there are 7202 known languages in the world [1]. Based on the prediction admitted by most scholars, more than half of these languages will disappear after a century. Language is very important for the heritage and spread of human culture. Each language contains the quintessence of one country or one nation, so language extinction means culture extinction. Known this, it's not surprising that language endangerment and extinction has become a worldwide social problem.

Based on MAT (Multi-Agents Technology), a social computational model for language endangerment and recovery is proposed in this paper. The computational model draws the endangerment trend and the recovery process under the effect of "resilience protection factor" of the weaker one in language competition in the artificial society.

This paper is structured as follows. Related research work is introduced in Section 1 and boundary of language endangerment status is presented. Social computational model on language endangerment and recovery and related computational experiments to prove the validity of the model are given in Section 2. The paper is concluded in Section 3.

1. ANALYSIS OF RELATED WORK AND DEFINITION OF LANGUAGE ENDANGERMENT STATUS

Language endangerment is a typical social problem. Now, it has attracted much attention from many linguists in many countries. But there is a difficulty we can't solve in

traditional social science research. Language endangerment is the fact we can't avoid in history and culture. Like other traditional social science problem, it's impossible to conduct computational experiments that can be repeated and assumed on language endangerment in reality. The computational experiments are at the risk of culture destruction. None of any nations want its own language to be endangered, so the "authenticity experiment" is absolutely rejected. So many researches on language endangerment are conducted after language endangerment happens. These researches are ex-post, which belongs to negative analysis after language endangerment. To break through this bottleneck, what we should put priority to is to try computational experiments on language endangerment in a risk-free space. In other words, parameters adjustment are under control and the evolution and corresponding results are repeatable, variable and controlled during the whole evolution process of language endangerment and recovery. In recent years, social computing has got much attention from researchers of information discipline and related interdisciplinary at home and abroad. Here, social computing means computing theory, methods and applications which orient social activity, social structure, social process, social organization and social function [2]. Lazer *et al.* published one paper named "Computational Social Science" in "Science" in 2009 [3], which marks a hot topic for social computing in the mainstream academic. As an emerging interdisciplinary research field, social computing has got much attention from the information discipline and related inter-disciplines in the world in recent years. At present, social computing is applied to many social science fields including linguistics. Chinese scholar Wang Feiyue proposed ACP (Artificial Societies + Computational Experiments + Parallel Execution) methodology [4]. ACP can solve the difficulty we encounter in "authenticity experiment".

Research on language evolution is one of the main fields in social computing. Language evolution includes the evolution of language structure and language function. Social computational model of language evolution once focused on the research of language structure [5-9]. The computational model was used to describe the individual learning of pronunciation, grammar, vocabulary and other language elements or simulate the appearance of language microscopically. Currently, the main social computing used to describe the evolution of language elements like pronunciation, grammar, vocabulary are biological evolution-based methods, complex system methods based on multi-agents, complex networks-based methods or methods based on artificial neural network [10-14]. However, the evolution of language function is the essence of language evolution. One of the most typical achievements on language function evolution is the paper written by Abrams and Strogatz published in Nature [15]. The social computational model in this paper says that a weak language in society will be threatened by foreign languages and it's likely to be extinct at last. The model is called AS model. Baker conducted computer simulation on AS model [16]. Considering the bilinguals, Mira et.al conducted similarity and stability analysis of the two languages in competition [17]. Castello et.al described the language competition through social network [18]. Kosmidis et.al simulated the language competition through Bit-string model which can save more storage [19]. Based on the above, we can see that existing language evolution computing models are generally short of language endangerment warning and endangered language recovery protection.

The following is our assumptions and definitions on language endangerment and the related problems. 1) Once more than half of the people using their own languages are over 40 years old in the population, the language is taken as endangered language. 2) Once more than one fifth people abandon their own languages in one population, the language is taken as endangered language. 3) Before 30 years old, saying ability for an individual is positively increasing with age. Between 30 and 50 years old, saying ability for an individual is relatively stable. After 50 years old, saying ability for an individual is declined with age. Listening ability for an individual is related to surrounding individuals. The above assumptions and definitions are the foundation of the social computational model in this paper.

2. SOCIAL COMPUTATIONAL MODEL ON LANGUAGE ENDANGERMENT AND RECOVERY

2.1. Construction of Social Computational Model on Language Endangerment and Recovery

Based on the language extinction computing model proposed by Abrams and Strogatz and the language competition computing model proposed by Baker, one social computational model is presented and improved in this paper. As the earliest mathematical model that describes language competition and language extinction, AS model constructed system dynamics equation taking the number of people and language status in a population as parameters. Based on Multi-Agents Simulation, Baker realized a computer simulation model on language competition. There is something in common between the two models. 1) The purpose of the two

models is to describe language competition without language endangerment warning or protective recovery. 2) In essence, two models are both static without considering the change of age, saying and listening ability, and social influence. Obviously, language competition is not just a dynamics process of the number of agents changing with time, but a comprehensive description considering the change of age, saying and listening ability and social influence of an agent in a population.

A social computational model with language endangerment warning and protective recovery is proposed in this paper. In one hand, the age, saying and listening ability and social influence of each agent in a population changes with time goes on. With age changes, the ability for saying and listening changes. Social influence changes with saying and listening ability changes. Two populations are given a certain birth and death rates. On the other hand, the static population in original model is replaced by artificial society dynamic population in this paper. Given this, the computational model will give warning once the weaker one in language competition is endangered. After warning, it conducts "resilience protection" to protect the endangered language efficiently.

Improvement of our model for the first aspect includes following works. 1) At the beginning, each agent is given an initial age. After a certain period, the age of each agent increases one. Once the agent is over 80 years old, it's dead. 2) During each certain period, each population is given a birth rate and non-normal mortality rate. The birth rate is 2% and its standard deviation is 0.01. Non-normal mortality rate is 5% and its standard deviation is 0.05. 3) Each agent in the population is able to speak and hear. With age changes, their ability of saying and listening changes. The influence radius changes with the ability of saying and listening changes. Details of our improvement can be found in our related paper [20].

For the second aspect, we introduced "resilience inner protective factor" into the simulation on the protection of endangered language in an artificial society. In psychology, resilience is the adaptation process when the individual encounter adversities or threatening. In other words, it's the bounce ability when encountering difficulty. Resilience inner protective factor means internal protective factor in language competition, which describes a fact that the members of a population voluntarily protect their own language because of cherishing their language when the language is endangered. The external protective factor in language competition means that the government of a population will take measures to push forward the language agency to protect the language. In our works, agents losing its own mother language are set as bilingual when weak language needs protection because of too many individuals losing their mother languages. Based on considering the above two aspects, composite protective factor is formed for each agent of the weak language populations.

There are two populations in our cultivated artificial society, that is, population1 and population2 supposing agents in population1 use language 0 at initial status and agents in population2 use language 1.

Definition 1. Endangerment factor of population language. k_{pi} is the ratio between sum of agents number with age over 40 years old and agents number abandoning mother languages and total number of agents in this population, which can be computed by Equation (1).

$$k_{pi} = \frac{oldagents_{pi} + changedagents_{pi}}{totalagents_{pi}} \quad (1)$$

Where i is code number for population, $oldagents_{pi}$ is agents number with age over 40 years old, and $changedagents_{pi}$ is agents number abandoning mother languages.

Definition 2. Resilience inner protective factor u_{ai} . Resilience composite protective factor works on a specific agent and is in direct proportion to endangerment factor of the population language the agent belongs to. And it is in inverse proportion to sum of radiuses for

agents saying another language within radius of this agent. Resilience inner protective factor can be computed by Equation (2).

$$u_{ai} = \frac{k_{pi}}{\sum_{anotherla \in S} I(|P_{anotherla} - P_{ai}| \leq R_{ai}) \times R_{anotherla}} \quad (2)$$

Where P_{ai} , R_{ai} are coordinates and radius of the

agent, $P_{anotherla}$, $R_{anotherla}$ are coordinates and radius

of the related agents mentioned above. S is the set of all speakers in the model and I is the indicator function.

Our cultivated artificial society will give a warning if a language is in an endangerment status. And at the same time, resilience inner protective factor will decrease the probability of agents changing to speak another language in the endangered population. In our computational simulations, probability of changing to speak another language is given in Equation (3) based on resilience inner protective factor, which is lower than Baker's model

$$P_{aichangela} = u_{ai} \times \left(\frac{\sum_{anotherla \in S} I(|P_{ai} - P_{anotherla}| \leq R_{anotherla})}{\sum_{anotherla \in S} I(|P_{ai} - P_{anotherla}| \leq R_{anotherla})} \right)^{1.31} \left(\frac{\sum_{anotherla \in S} I(|P_{ai} - P_{anotherla}| \leq R_{anotherla}) \times R_{anotherla}}{\sum_{anotherla \in S} I(|P_{ai} - P_{anotherla}| \leq R_{anotherla}) \times R_{anotherla}} \right) \quad (3)$$

When the endangered point happens to the language

used by a population in language competition, composite protective factor including the resilience inner protective factor and external protective factor will start to protect the endangered language. If the endangerment warning occurs at the same time in two populations, the one with fewer agents is protected.

We created a process to execute its own events list for each agent. In the whole development process of artificial society, we recorded the change of agent and the interaction among agents through time-step method. Within a specific period, we adopted event-driven method. Event-driven method is that once the status of an agent changes, other agents are all influenced. So we should determine and compare the choices agents make continuously. When the endangered point appears and recovery protection is started, the

artificial society we cultivated transmits the endangerment factor k to all agents in the population. Based on this, these agents will get the computation value of resilience inner protective factor by combining external protective factor and finally achieve the language recovery and protection.

2.2. Attributes of Agents and Model Parameters Setting

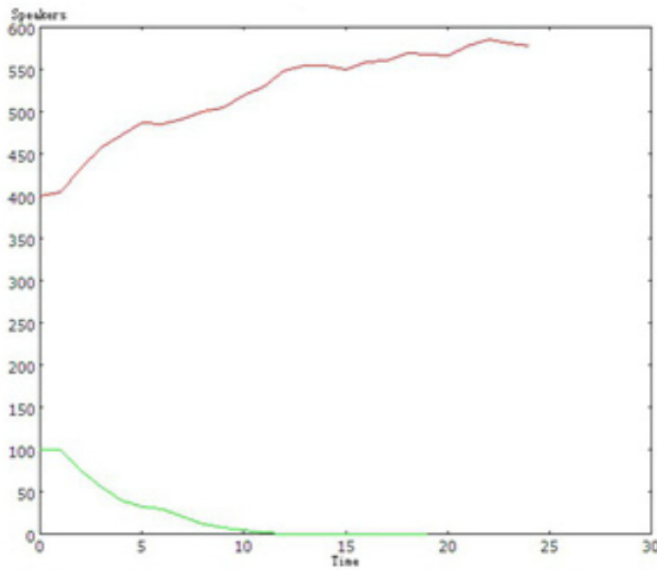
Attributes like age, id, position, radius, language, livestatus, sayability and listenability are given to agents. Some attributes can be affected by agent's activities, for example, sayability and listenability. In addition, some attributes may change after one time step. There are three reasons why these attributes change. The first reason is that the agent itself changes, for example, age increases. The second one is the rule in artificial society. The deadlist contains all the ids of non-normal dead agents. If the agent's id exists in deadlist, its livestatus is false. The third one is the other agents around this agent. Surrounding agents give the values of attributes including position, language and radius to the agent. These change the values of this agent's attributes, among which language chosen is the most important one.

Parameters of the model are initial number of agents, position, position offset (standard deviation), radius and radius offset (standard deviation). In order to make the populations to fully interact with each other, we usually set the center of two populations at same coordinates.

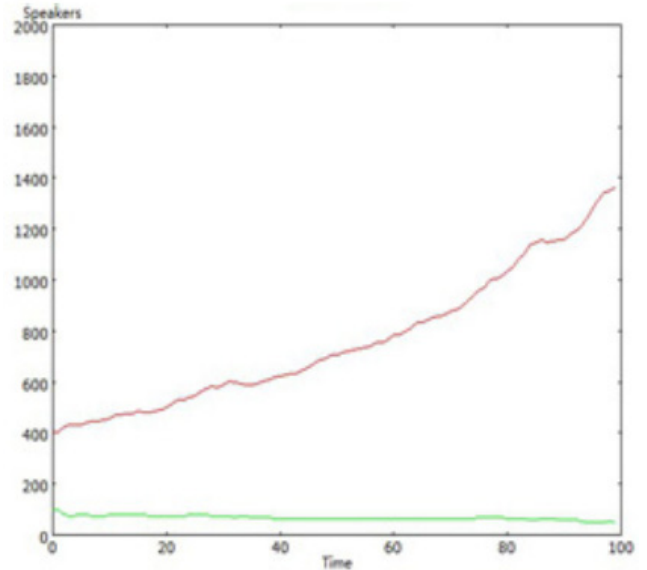
2.3. Computational Experiments

Experimental parameters are set with two populations. Parameters of the first population are as following. Numbers of agents for the two populations are 400 and 100. Position and position offset (standard deviation), and radius and radius offset (standard deviation) of population 1 using language 0 are 50, 5, 5, and 0.5. Position and position offset (standard deviation), and radius and radius offset (standard deviation) of population2 using language1 are 50, 3, 1, and 0.1. Parameters of the second population are as following. Initial numbers of agents for two populations are 300 and 100. Position and position offset (standard deviation), and radius and radius offset (standard deviation) of population1 using language 0 are 50, 20, 7, and 0.8. Position and position offset (standard deviation), radius and radius offset (standard deviation) of population2 using language 1 are 50, 1, 5, 0.5. Whether to introduce the resilience composite protective will influence the future of endangered language (extinction or recovery). Fig. 1(a), Fig. 1(b), Fig. 2(a), and Fig. 2(b) show the difference.

In fact, according to the definition in this paper on endangered language with parameter values of the two populations, language 1 (indicated by the green curve) has been endangered at 1% iteration. By the final statistical results, language 1 has been endangered 18 times and 20 times for the two times computations while language 0 has never been endangered. The future of language 1 in Fig. 1(a) and Fig. 2(a) makes us realize that we should have found that language 1 is endangered earlier. And language 1 goes extinct eventually in Fig. 1(a) and Fig. 2(a). From Fig. 1(b) and Fig. 2(b) we can see that language 1 doesn't go extinct during the whole simulation (100 years in total, the period is set as 20 because the time simulation is long) because of introduction of composite protective factor.

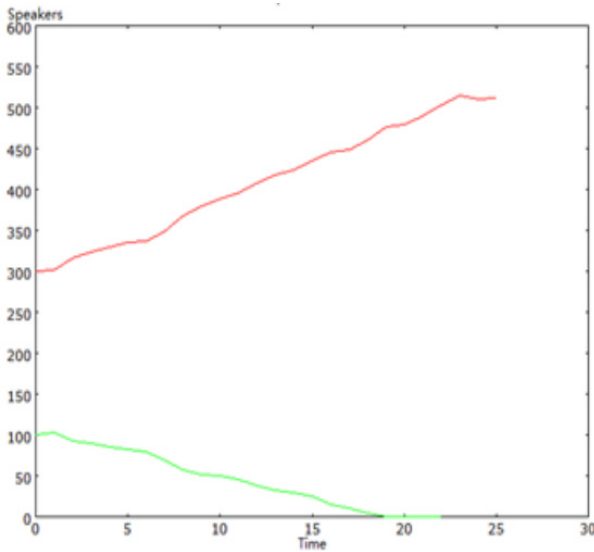


(a) without resilience composite protective factor

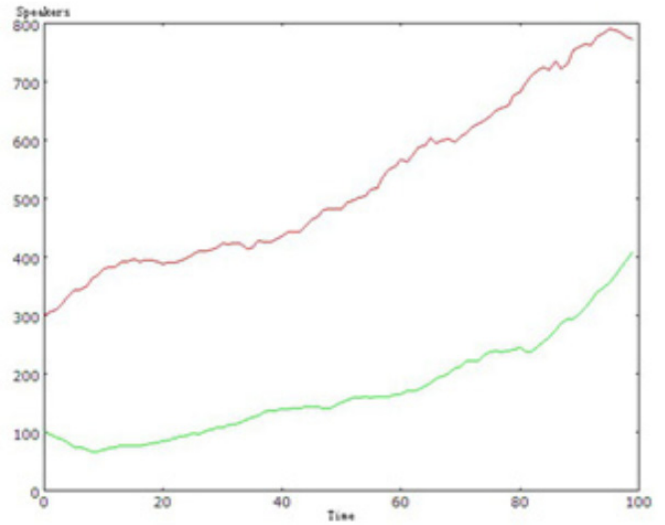


(b) introduce of composite protective factor

Fig. (1). Different future of endangered language with parameters of population1.



(a) without resilience composite protective factor



(b) introduce of composite protective factor

Fig. (2). Different future of endangered language with parameters of population2.

Table 1(a). Change of numbers of agents in the two populations without composite protective factor under parameters of population1

Time(year)	0-1	2-3	4-5	6-7	8-9	10-11	12-13	14-15	16-17	18-19
population1	400-403	407-413	415-420	423-429	431-439	442-453	457-469	473-487	490-503	508-512
population2	100-98	95-82	80-66	60-45	43-35	34-24	18-12	11-7	5-4	3-0

From Table 1 and Table 2 we can see that the weaker language population (population2) goes extinct soon without introduction of composite protective factor. While the composite protective factor is introduced into the model, the number of agents in population2 (the weaker language population) doesn't go extinct and even increases in some extent, which verifies the effect of recovery and protection.

To make the computational experiments more convincing, we conducted statistical analysis for a number of experimental results after introduction of composite protective factor. In Table 3, population2 using language1 is the weaker language population with smaller population radius than population1 using language 0.

From Table 3 we can see that population2 as a weaker language population is not bound to go extinct in five experiments after introduction of composite protective factor into the model. Instead, population2 survives together with the stronger language population.

SUMMARY

The language endangerment computational experiments we have finished are different from the real language system. The difference is that our computational model runs in an artificial society. The computational experiments are designed to be similar to the real language system as possible as we can. But it can also be a substitute to the real language system. Based on the social computational model on language endangerment and recovery, we can conduct active and repeatable experiments. Hence, language endangerment from original non-repeatable at moral ethical risks changes into a risk-free and a complete and deep insight provision in the research process.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by Humanity and Social Science Foundation of Ministry of Education of China: Computational Modeling of Alarm System of Cross-culture Language Transmission and Endangered Languages Preservation (No. 12YJCZH263). And this work also was supported by the Fundamental Research Funds for the Central Universities of China(No.DC201502030401).The authors would like to thank Dr. Baker for the some code of literature[16],and Yujing Ye for useful assistances.

REFERENCES

- [1] W.W.Gibbs, "Saving dying languages," *Scientific American*, pp. 80-85, 2002.

- [2] F. Wang, X. Li, W. Mao, and W. Tao, "The Basic Method and Application of Social Computing," Hang Zhou:Zhejiang University Press, 2013.
- [3] D. Lazer, A. Pentland, L. Adamic, A.L. Barabási, D. Brewer, N. Christakis, N. Contractor7 J. Fowler, M.Gutmann, T. Jebara, G. King, M. Macy, D.Roy, M. Van Alstyne, "Computational social science", *Science*, vol. 323, no.5915, pp.721-723, 2009.
- [4] F. Wang,"Toward a paradigm shift in social computing: the ACP approach", *IEEE Intelligent Systems*, vol. 22, no.5, pp. 65-67, 2007.
- [5] V. Loreto, and L. Steels, "Social, dynamics: Emergence of language," *Nature Physics*, vol. 3, pp.758-760, 2007.
- [6] H. Cornish, M. Tamariz, and S. Kirby, "Complex Adaptive Systems and the Origins of Adaptive Structure: What Experiments Can Tell Us", *Language Learning*, vol. 59, no. 1, pp. 187-205, 2009.
- [7] S. Kirby, H. Cornish, and K. Smith, "Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language", *PNAS*, vol. 105, no.31, pp.10681-10686, 2008.
- [8] D. Parisi, and M. Mirolli, "The emergence of language: How to simulate it", *Emergence of Communication and Language*, Springer-Verlag, London, pp. 269-285, 2007.
- [9] J.Y. Ke, J. W. Minett, C.P. Au, and S.Y. Wang, "Self-organization and selection in the emergence of vocabulary",*Complexity*, vol. 7, no.3, pp. 41-54, 2002.
- [10] J. Ke, M. Ogura, and W.S. Y. Wang, "Optimization Models of Sound Systems Using Genetic Algorithms", *Computational-Linguistics*, vol. 29, no.1, pp.1-18, 2003.
- [11] H. Jaeger., S. Luc, B. Andrea, B. Ted, H. C. Morten, G. Thomas, G. Jäger, K. Simon, L. K. Natalia, J. R. Peter, and T. Jochen, "What can mathematical, computational, and robotic models tell us about the origins of syntax?", *Biological Foundations and Origin of Syntax*, Cambridge, CA: MIT Press, pp. 385-410, 2010.
- [12] P. de Bie., and B. de Boer, "An agent-based model of linguistic diversity", In: A. Benz, C. Ebert & R. van Rooij (Eds.) *Proceedings of the ESSL1 2007 Workshop on Language, Games, and Evolution*, 2007.
- [13] A. Mukherjee, M. Choudhury, A. Basu, and N. Ganguly, "Emergence of community structures in vowel inventories:an analysis based on complex networks", In *Proceedings of Ninth Meeting of the ACL Special Interest Group in Computational Morphology and Phonology*, Prague, 2007.
- [14] P.Y. Oudeyer, "Coupled Neural Maps for the Origins of Vowel Systems", In G. Dorffner and H. Bischof and K. Hornik, editors, In: *Proceedings of the International Conference on Artificial Neural Networks*, LNCS2130, Springer Verlag, pp. 1171-1176, 2001.
- [15] Abrams and Strogatz, "Modelling the dynamics of language death", *Nature Publishing Group*, vol. 424, no.21, pp. 900, 2003.
- [16] A. C. Baker, "Social Proximity, Influence and the Dynamics of Language Death", Chicago, <http://home.uchicago.edu/~adamc/papers/language-death.pdf>, 2007.
- [17] J. Mira, L. F. Seoane, and J. J. Nieto, "Importance of interlinguistic similarity and stable bilingualism when two languages compete," <http://arxiv.org/pdf/1006.2737v2.pdf>, 2011.
- [18] X. Castello, R. Toivonen, V. M. Eguiluz, J. Saram aki, K.Kaski and M.San Miguel, "Anomalous lifetime distributions and topological traps in ordering dynamics", *EPL-EUROPHYS LETT*, vol. 79, no.6, pp. 1-6, 2007.
- [19] K. Kosmidis, J.M. Halley, and P.Argyarakis, "Language Evolution and Population Dynamics in a System of Two Interacting Species", *Physica A*, vol. 353, pp.595-612, 2005.
- [20] J. Yun, X. Li, and C. Wang, "Simulation of Conducting Early-warning to the Endangered State of Language", *Journal of Multimedia*, vol. 8, no. 5, pp. 475-480, 2013.