



Title	A cross-model study on the effect of power-laws on language evolution
Author(s)	Gong, T; Shuai, L
Citation	The 2012 IEEE World Congress on Computational Intelligence(WCCI 2012), Brisbane, Australia, 10-15 June 2012. In Conference Proceedings, 2012, p. 3237-3244
Issued Date	2012
URL	http://hdl.handle.net/10722/179874
Rights	Congress on Evolutionary Computation Proceedings. Copyright © IEEE.

A Cross-Model Study on the Effect of Power-Laws on Language Evolution

Tao Gong

Department of Linguistics
University of Hong Kong
Hong Kong
gtojty@gmail.com

Lan Shuai

Department of Chinese, Translation and Linguistics
City University of Hong Kong
Hong Kong
susan.shuai@gmail.com

Abstract—Based on three evolutionary computational models that respectively simulate lexical, categorical and syntactic evolutions, we explore the effect of power-law distributed social popularity on language origin and change. Simulation results reveal a critical scaling degree ($\lambda \approx 1.0$) in power-law distributions that helps accelerate the diffusion of linguistic conventions and preserve high linguistic understandability in population. Other scaling degrees ($\lambda = 0.0$ or $\lambda > 1.0$), however, tend to delay such diffusion process and affect linguistic understandability. Apart from the conventionalization nature of language communications in these models, increase in population size could also contribute to select the critical scaling degree, since this scaling degree can accommodate the influence of population size on linguistic understandability and many power-laws in real-world systems have their scaling degrees around this critical value.

Keywords—computer simulation; language evolution; power-law; social popularity

I. INTRODUCTION

In many natural, social and technical systems, system components, such as their ranks, connectivity with others, probabilities for taking part in interactions with others, and so on, usually bear nonlinear relations with each other. Such relations can be described mathematically by primarily three types of probability distributions, including *normal distribution* ($f(x) \sim e^{-k(x)}$, $k(x)$ is a quadratic polynomial of x , in which the factor of the quadratic component is negative, and $f(x)$ is a density function), *exponential distribution* ($f(x) \sim ae^{-ax}$, $a > 0$), and *scaling distribution* ($f(x) \sim x^{-a}$, $a \neq 0$) [1]. Many latest findings in a number of disciplines (e.g., organism mass vs. metabolic rate across species [2], fractal structures of objects [3][4], numbers of recalled memories vs. recalling periods [5], and the rank of words vs. their occurrence frequencies [6]) further illustrate that scaling relations are ubiquitous in self-organizing systems involving multiple dependent components and intricate relations among them [1][7-10]. Among numerous scaling relations, *power-laws*, designated by *power-law distribution* ($f(x) \sim x^{-\lambda}$, $\lambda > 0$), are pervasive in many social and linguistic communities. For example, the relation between the rank of a word and its occurrence frequency in many language corpora follows a power-law (a.k.a. Zipf's law, $\lambda \approx 1.0$) [6]; the relation between the rank of a language family and the number of languages it has also follows a power-law ($\lambda \approx 2.0$, based on the

data from Ethnologue) [11-13]; and many social networks, such as email or collaboration network [14], are *scale-free*, having power-law distributed *node degree* (number of edge per node) [15].

Facing these many phenomena, more and more scholars, accepting the universality of power-laws in nature, have started to examine the causal mechanisms to such laws. Computer simulation and statistical analysis have recently served as new means of addressing this question. Recent work has shown that mechanisms such as preferential attachment [9] and constraints from geography [16] or kinship relations [17] can lead to power-law distributions in some social systems. Apart from the causal mechanisms, it is also worth studying how power-laws affect mutual information transmission during linguistic interactions or other social activities. Social structures and activities abound in human communities, and some of these structures and activities came into being prior to language [18]. Sociolinguists and historical linguists have observed that certain structures and activities could cast their influences on language origins and evolution. For example, different types of social constraints on communicative patterns among individuals can restrict the diffusion of linguistic variants within and across communities [19][20], and even identical social factors may exert different effects on various aspects of language evolution, due to different language processing mechanisms involved in those aspects.

In order to explore the effect of power-laws on language evolution, this paper conducts a cross-model study based on evolutionary computational models of language evolution. Instead of actual connections among individuals, we define an individual's *social popularity* as the probability for this individual to take part in linguistic communications, and use power-law distributions to manipulate all popularities. In addition, rather than one particular model on certain aspect(s) of language evolution, our cross-model analysis is concerned with three such models, including: 1) *the naming game* [21], which simulates the origin of consensus on a meaning-utterance association in population; 2) *the category game* [22], which studies the origins and diffusion of linguistic categorization patterns among individuals; and 3) *the lexicon-syntax coevolution model* [23][24], which examines the coevolution of lexical items and simple word orders during language origin and change. Our study reveals a correlation

This work is supported in part by the Society of Scholars in Humanities in the University of Hong Kong.

between the *scaling degree* (denoted by λ value) of a power-law distribution and linguistic understandability in population: a *critical scaling degree* ($\lambda \approx 1.0$) is able to accelerate the diffusion of linguistic conventions and keep a high level of linguistic understandability among individuals, whereas other scaling degrees ($\lambda = 0.0$ or $\lambda > 1.0$) tend to delay the diffusion process and affect mutual understandability in population. In addition, since the power-law with the critical scaling degree can accommodate the influence of population size on linguistic understandability, the critical scaling degree could result from increase in population size. In fact, many power-laws in real-world large-scaled systems have their scaling degrees around the critical value.

The rest of the paper is organized as follows: Sec. 2 describes the social popularity and simulation setting; Sec. 3 analyzes the simulation results of the three adopted models; and Sec. 4 discusses these results and concludes the paper.

II. POWER-LAW DISTRIBUTED SOCIAL POPULARITY

Modeling social structures as complex networks has been widely adopted in simulation studies exploring social structure effect on language evolution. In such networks, individuals are treated as *nodes*, and their communications *edges* among nodes [25]. Various types of networks, such as row [26], lattice [27], ring [28], and small-world or scale-free network [29-31], have been used to denote human communities. However, using connections among individuals to denote social relations is only reasonable in large-scale communities. It is unrealistic to define such connections in small-scale societies, where any individual may directly interact with anyone else. A weighted fully-connected network helps overcome this limitation, but it is hard to estimate pairwise connection weights from empirical data usually obtained at the population level. Noting this, instead of local connectivity among individuals, we define social popularity as the probability for an individual to participate into communications with others, and adopt power-laws to manipulate all individuals' popularities in population. In this way, the common features of different types of networks having different connectivity patterns can be easily generalized, and the influence of network size can be diminished. Meanwhile, such way of abstraction requires a small set of parameters whose values can be estimated from group data, and statistical analyses can help get some quantitative understanding of the correlation between language evolution and social factors.

The power-law relation between individual rank and social popularity in this paper is designated by:

$$y = ax^{-\lambda} \quad (1)$$

Here x is an individual's rank (from 1 to N , N is population size), y is the social popularity of this individual, a is a parameter ensuring the sum of all popularities as 1.0, and λ characterizes power-law distributions. According to [14], many real-world power-law distributions usually have their λ within [0.0 3.0]. If $\lambda = 0.0$, all individuals have the same probability to interact with each other, which resembles the case of random communications. If λ is within (0.0 1.0), popularities are sensitive to network size, but once $\lambda > 1.0$, the influence of

network size is not obvious, especially in big networks. Considering these, we only consider some λ values, including 0.0, 1.0, 1.5, 2.0, 2.5, and 3.0.

Mathematical analysis reveals a correlation between λ of a power-law distribution of social popularity and λ' of a power-law degree distribution. This correlation is shown in Eq. (2) and proved in Eq. (3):

$$\lambda' = 1 + 1/\lambda \quad (2)$$

$$\left. \begin{aligned} r(k) &\propto \int_k^\infty p'(k)dk = \int_k^\infty k^{-\lambda'}dk \propto k^{1-\lambda'} \\ p(r(k)) &= r(k)^{-\lambda} \propto k \end{aligned} \right\} \rightarrow -\lambda = \frac{1}{1-\lambda'} \quad (3)$$

Here, $p(k)=k^{-\lambda}$ is a power-law distribution of social popularity, $p'(k)=k^{-\lambda'}$ is a power-law degree distribution, and $r(k)$ is the rank of individuals having more than k edges. Scaling factors are omitted here. Such correlation holds when both λ and λ' are bigger than 1.0.

In order to illustrate the general effect of power-law distributed social popularity on language evolution, we set the population size N as 50, 100, 150, 200, 300, 400, and 500. Power-law distributions in these populations can be calculated following the definition in Eq. (1). Under each of the 6 distributions and in each of the 7 populations for statistical analysis, we conduct 20 simulations.

III. EFFECT OF POWER-LAW DISTRIBUTED SOCIAL POPULARITY ON LANGUAGE EVOLUTION

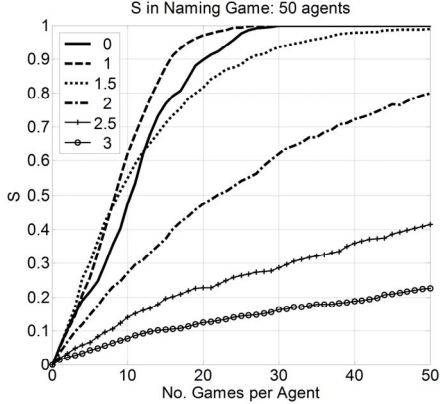
The three adopted models, namely the naming game [21], category game [22], and lexicon-syntax coevolution model [23][24], have been designed to respectively study the evolution of lexical items, categorization patterns, and simple syntax in terms of basic word orders among at most three lexical items. Instead of implementation details, which can be found in relevant papers, our work focuses on the performance of language evolution in these models under different types of power-laws. Due to different processing mechanisms involved in these models, the dynamics of language evolution in these models are manifest in different timescales, and traced by different indices. Based on the indices that primarily trace mutual linguistic understandability, we examine the effect of power-law distributed social popularity on language evolution from two aspects. We first evaluate the effects of power-laws and population sizes on linguistic understandability in each of these models, and then, summarize the general effect across these models.

A. The Naming Game

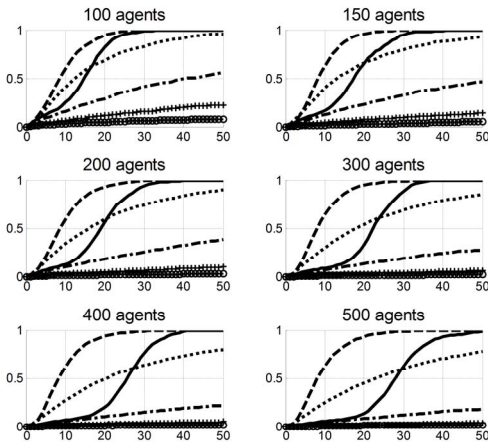
This model simulates the origin of consensus on naming a particular semantic item within a population of individuals. Linguistic convention refers to the common lexical name for the semantic item. The dynamics of this game is traced primarily by the number of different names in population (N_d) and the rate of successful games among individuals (S). In our study, we focus on S , which reflects linguistic understandability in population. Due to the simple processing mechanisms towards lexical names, such as randomly creating a name or

deleting some competing ones, the dynamics of this game is manifest in a short timescale. We set the number of games at 50 (per agent, actual numbers of games are dependent on N , and due to scaled social popularity, not all agents participate into exactly the same number of games during simulations).

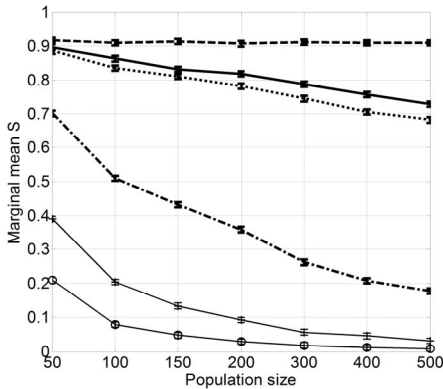
Fig. 1(a) traces the dynamics of this game in a 50-agent population. In the case of random communications ($\lambda = 0.0$),



(a)



(b)



(c)

Figure 1. Successful rate of the naming game (S) under different power-law distributed social popularities in a 50-agent population (a) and other populations (b). (c): marginal mean S under different power-law distributed social popularities in different populations. Legends in (b) and (c) are the same as that in (a). Each line is averaged over 20 simulations.

the dynamics of this game includes two phases: first, S remains low, indicating that individuals keep inventing new names, but many games fail due to using different names, and then, after some games, S starts to increase and gradually reaches 1.0, indicating that individuals begin to share a common name to reach success in most games. Compared with the case of random games, when $\lambda = 1.0$ or 1.5, the second phase, increase in S , occurs earlier; when $\lambda > 1.5$, it occurs later. Meanwhile, when $\lambda = 1.0$, increase in S is slightly faster; when $\lambda > 1.0$, it becomes slower, and after 50 games per agent, S cannot reach 1.0, compared with cases where $\lambda = 0.0$, 1.0 or 1.5.

Fig. 1(b) shows the results in other populations. In the case of random games, with increase in N , increase in S occurs later and later. This tendency is consistent with the statistical analysis in [21], and also shows in cases where $\lambda > 1.0$. In those cases, increase in S takes place much later and final values of S after 50 agents per agent are smaller. However, the trajectory of S does not change much in the case where $\lambda = 1.0$ across different N ; after a small number of games per agent, S can quickly reach 1.0.

These observations are confirmed by a two-way analysis of covariance (ANCOVA) (dependent variable: S ; fixed factor: 6 λ values; random factor: 7 N values; covariate: 20 sampling points in all games per agent). Rather than ANOVA, we use ANCOVA and treat number of games per agent as a covariate. This aims to partial out the influence of number of games per agent. In addition, since population size is not limited to these 7 values, we treat N as a random factor, instead of a fixed factor as λ . The ANCOVA reveals that both λ ($F(5, 30) = 222.026$, $p < .001$, $\eta_p^2 = .974$) and N ($F(6, 30) = 9.899$, $p < .001$, $\eta_p^2 = .664$) have significant main effects on S and these two factors significantly interact with each other ($F(30, 16757) = 69.092$, $p < .001$, $\eta_p^2 = .110$). The covariate, number of games per agent, also significantly correlates with S ($F(1, 16757) = 12935.405$, $p < .001$, $\eta_p^2 = .436$). Fig. 1(c) traces the marginal mean S under different power-law distributions in different populations. It is shown that the marginal mean S in the case where $\lambda = 1.0$ remains similarly high across different N , whereas those in other cases drop along with increase in N .

To summary, in this game, compared with the case of random games, power-law distributed social popularity with a particular scaling degree ($\lambda = 1.0$) can efficiently accelerate the diffusion of consensus in population. This effect is less affected by population size. However, other scaling degrees ($\lambda = 0.0$ or $\lambda > 1.0$) will delay the diffusion process and affect S . This effect is more explicit in bigger populations.

B. The Category Game

Extended from the naming game, the category game simulates the origins and diffusion of linguistic categorization patterns among individuals. During the category games, individuals develop categorical knowledge to discriminate semantic items from a continuous perceptual space. Linguistic conventions consist of both common lexical items and linguistic categories with similar perceptual boundaries across individuals, and processing mechanisms towards both lexical names and categories, such as inventing or updating new lexical names and new categories with particular boundaries for

the purpose of discrimination, are simulated. The evolution of linguistic categorization patterns proceeds in a much bigger timescale. Noting this, we set the number of games at 10^6 (per agent). The dynamics of this game is traced primarily by the degree of boundary alignment of linguistic categories in individuals, the number of shared lexical names of linguistic categories, and the successful rate of category games among individuals (S), similar to the naming game. In our study, we focus on S .

Fig. 2(a) traces the dynamics of the naming game in a 50-agent population. In the case of random communication ($\lambda = 0.0$), the dynamics of this game consists of two phases. First, new perceptual categories with different boundaries and lexical names are created for discrimination, but misunderstanding is frequent and S remains low. After many games, new perceptual categories keep emerging, but due to boundary mismatch, adjacent perceptual categories in some individuals start to share common lexical names and merge into *linguistic categories* (see [22] for an example). Then, although the boundaries of most perceptual categories remain mismatched, those of linguistic categories gradually become roughly aligned. Then, S starts to increase and reach above 0.9. After this stage, the system remains stable for a very long time. If one waits for a longer time (say, $10^5 \sim 10^6$ games per agent), a slight drop of S can be observed [22]. Compared with the case of random games, when $\lambda = 1.0, 1.5$ or 2.0 , the second phase, increase in S , occurs earlier; when $\lambda > 2.0$, it occurs much later. Meanwhile, when $\lambda = 1.0$, increase in S becomes faster; when $\lambda > 1.0$, it is slower, and after 10^6 games per agent, S finally reaches high values around 1.0, much more slowly than cases where $\lambda = 0.0$ and 1.0. Note that the X-axis in Fig. 2 is in a logarithm scale; a small distance in X corresponds to a big number of games per agent.

Fig. 2(b) shows the results in other populations. In the case of random games, with increase in N , increase in S occurs later and later. This tendency is also shown in cases where $\lambda > 1.0$. However, the trajectory of S does not change much in the case where $\lambda = 1.0$ across different N ; after a small number of games per agent, S reaches a high value around 1.0.

These observations are confirmed by an ANCOVA similar to that in the naming game. It reveals that both λ ($F(5, 30) = 95.266, p < .001, \eta_p^2 = .941$) and N ($F(6, 30) = 6.874, p < .001, \eta_p^2 = .579$) have significant main effects on S and these two factors significantly interact with each other ($F(30, 16757) = 62.743, p < .001, \eta_p^2 = .101$). The covariate, number of games per agent, also significantly correlates with S ($F(1, 16757) = 10802.522, p < .001, \eta_p^2 = .392$). As shown in Fig. 2(c), similar to the naming game, the marginal mean S in the case where $\lambda = 1.0$ is similarly high across all populations, whereas those in other cases drop with increase in N .

These results lead to a conclusion similar to that in the naming game. In the category game, a particular scaling degree ($\lambda = 1.0$) in power-law distributed social popularity can accelerate the diffusion of common linguistic categorization patterns among individuals. This effect is less affected by population size. However, other scaling degrees ($\lambda = 0.0$ or $\lambda > 1.0$) will delay the diffusion process and affect S . This effect is more explicit in bigger populations.

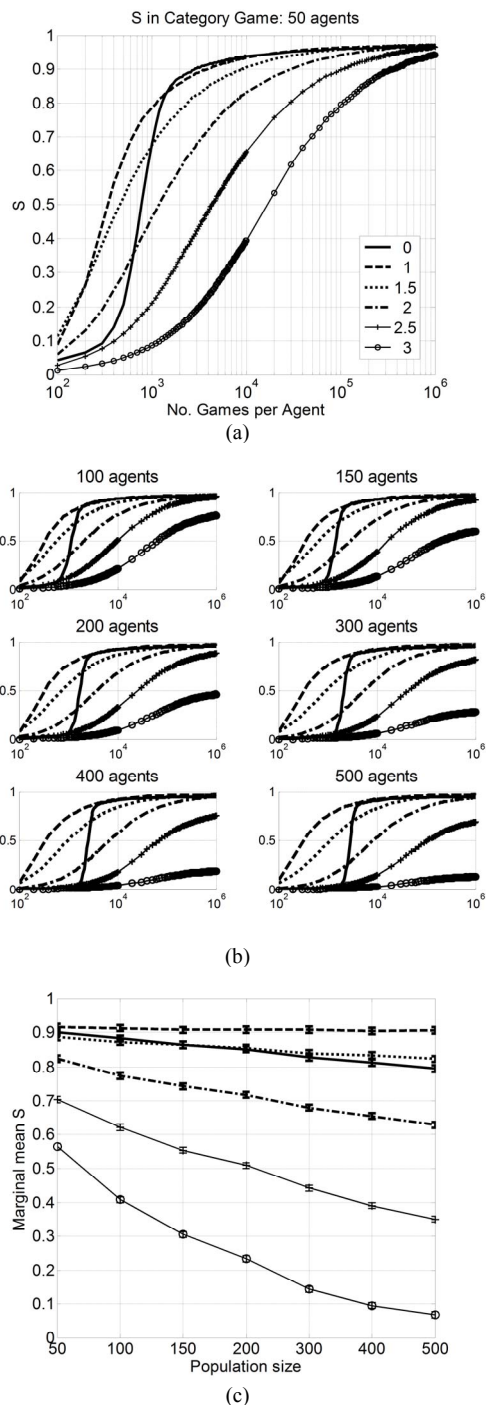


Figure 2. Successful rate of the category game (S) under different power-law distributed social popularities in a 50-agent population (a) and other populations (b). (c): marginal mean S under different power-law distributions in different populations. Legends in (b) and (c) are the same as that in (a). Each line is averaged over 20 simulations.

C. The Lexicon-Syntax Coevolution Model

This model simulates the coevolution of lexical items and simple syntax during language origin and change. Instead of lexical evolution as in the previous two models, this model examines both lexical and syntactic evolutions. Instead of lexical items, the artificial language evolved in this model

encodes semantic expressions with simple predicate-argument structures into sentences with simple syntactic structures. This artificial language resembles many aspects of real languages. Moreover, apart from language origin, this model can simulate language change, by letting all individuals initially share a set of common linguistic knowledge. Linguistic conventions consist of not only common lexical items and syntactic categories, but also similar syntactic knowledge to regulate lexical items in sentences. Processing mechanisms towards both lexicon and syntax, such as acquiring lexical items according to recurrent patterns among exchanged meaning-utterance mappings in communications, learning knowledge about word orders on the basis of sequential information of lexical items in utterances, and assigning lexical items and word orders into syntactic categories based on semantic and ordering relations of lexical items in exchanged utterances, are equipped by individuals to develop linguistic knowledge during communications. Due to these factors, language evolution in this model proceeds in a timescale distinct from those in the above models.

The dynamics of this model is traced primarily by the expressivity of linguistic knowledge in individuals and the understandability of linguistic knowledge among individuals (UR). In our study, we focus on UR , set the number of communications at 600 (per agent), and conduct both language origin and change simulations. In the origin simulations, individuals initially share 8 rules to encode only 8 out of 64 semantic expressions; in the change simulations, individuals initially share a set of linguistic rules capable of encoding all 64 semantic expressions.

Figs. 3(a) and 4(a) trace the dynamics of this model in a 50-agent population. As for the origin simulations (Fig. 3(a)), in the case of random communications ($\lambda = 0.0$), the dynamics of the model comprises two phrases. First, using their learning mechanisms, individuals begin to acquire new linguistic knowledge to encode more semantic expressions. However, since much of this newly-acquired knowledge is not yet largely shared, UR remains low and may even drop, when some new knowledge competes with the original knowledge. Then, after many communications, competition causes certain knowledge to be widely shared among individuals, and based on this knowledge, mutual understanding becomes frequent, and UR starts to increase and reaches around 1.0. Compared with the case of random communications, when $\lambda = 1.0$ or 1.5, increase in UR occurs earlier; when $\lambda > 1.5$, increase in UR is not obvious, and after 600 communications per agent, UR remains low. Meanwhile, although increase in UR occurs early in cases where $\lambda = 1.0$ and 1.5, UR reaches a high value only in the case where $\lambda = 1.0$, similar to that in the case of random communications.

As for the change simulations (Fig. 4(a)), in the case of random communications, the dynamics is straightforward: UR remains stable and relatively high (over 0.8) throughout the simulation, but some lexical and/or syntactic rules may change. Compared with the case of random communications, a relatively high UR is kept only in the case where $\lambda = 1.0$. In other cases, UR drops and never rises up to a high value within 600 communications per agent. By tracing shared linguistic knowledge, we find that even in cases where UR remains high,

some linguistic knowledge gradually becomes different during communications; the utterances of some shared lexical items are different. This reflects an inevitable change of language during cultural transmission [32]. What power-law distributed social popularity helps preserve is linguistic understandability, based on a set of consistently changing linguistic knowledge.

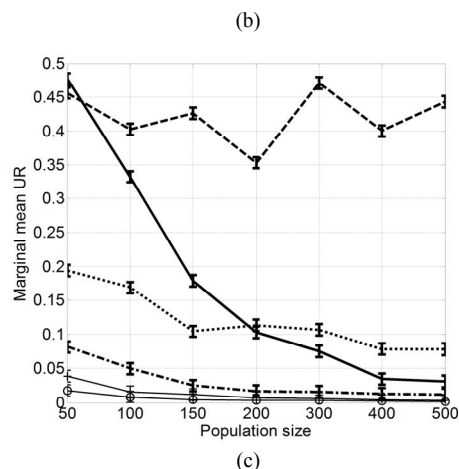
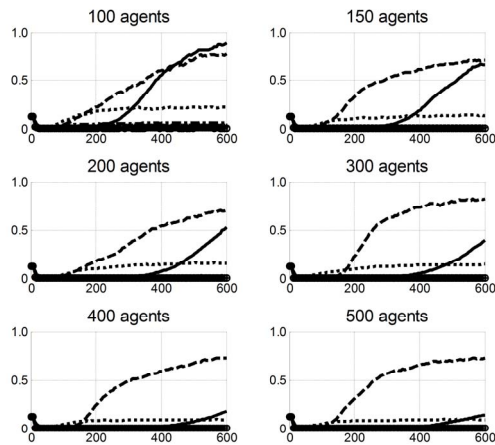
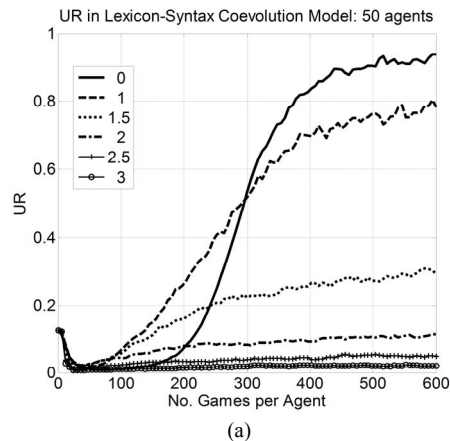


Figure 3. Understanding rate (UR) in the language origin simulations based on the lexicon-syntax coevolution model under different power-law distributed social popularities in a 50-agent population (a) and other populations (b). (c): marginal mean UR under different power-law distributions in different populations. Legends in (b) and (c) are the same as that in (a). Each line is averaged over 20 simulations.

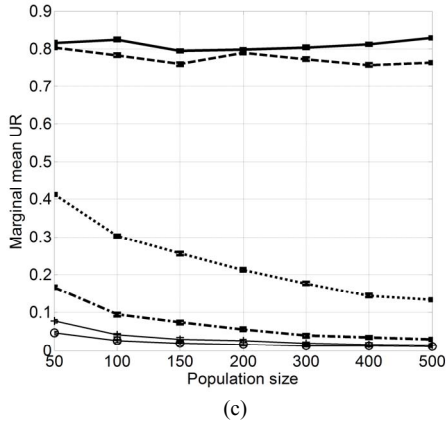
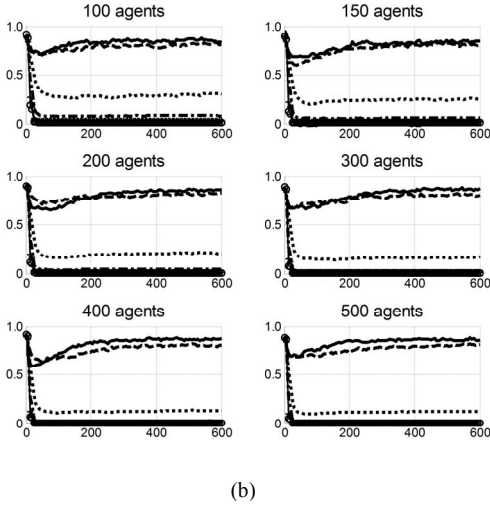
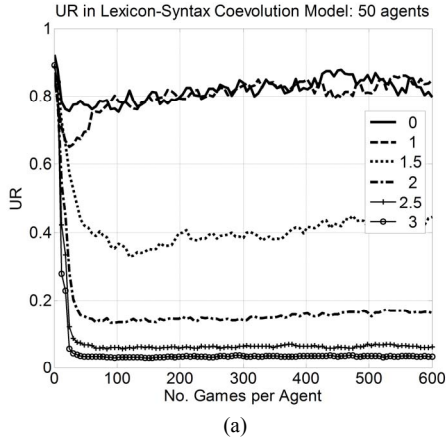


Figure 4. Understanding rate (UR) in the language change simulations based on the lexicon-syntax coevolution model under different power-law distributed social popularities in a 50-agent population (a) and other populations (b). (c): marginal mean UR under different power-law distributions in different populations. Legends in (b) and (c) are the same as that in (a). Each line is averaged over 20 simulations.

Fig. 3(b) and 4(b) show the results in other populations. As for the origin simulations (Fig. 3(b)), in the case of random communications, with increase in N , increase in UR occurs later and later, and the peak UR becomes smaller and smaller, within 600 communications per agent. This tendency is also manifest in cases where $\lambda > 1.0$. However, in the case where λ

$= 1.0$, the trajectory of UR does not change much across different N ; within 600 communications per agent, UR reaches a high value around 0.8, even in bigger populations. As for the change simulations (Fig. 4(b)), with increase in N , relatively high values of UR are kept only in cases where $\lambda = 0.0$ and 1.0. In other cases, UR drops with increase in N .

These observations are confirmed by ANCOVA. As for the origin simulations, the ANCOVA reveals that both λ ($F(5, 30) = 41.593, p < .001, \eta_p^2 = .874$) and N ($F(6, 30) = 2.763, p < .001, \eta_p^2 = .356$) have significant main effects on UR and these two factors significantly interact with each other ($F(30, 16757) = 60.596, p < .001, \eta_p^2 = .098$). The covariate, number of communications per agent, also significantly correlates with UR ($F(1, 16757) = 3738.736, p < .001, \eta_p^2 = .182$). Fig. 3(c) shows that the marginal mean UR in the case where $\lambda = 1.0$ is similarly high across different N , whereas those in other cases drop with increase in N .

As for the change simulations, the ANCOVA reveals that both λ ($F(5, 30) = 702.010, p < .001, \eta_p^2 = .992$) and N ($F(6, 30) = 4.615, p < .001, \eta_p^2 = .480$) have significant main effects on UR and these two factors significantly interact with each other ($F(30, 16757) = 86.398, p < .001, \eta_p^2 = .134$). However, the covariate, number of communications per agent, only marginally correlates with UR ($F(1, 16757) = 3.267, p = .071$), indicating that in some conditions, number of communications is not correlated with UR . Fig. 4(c) shows that the marginal mean UR in cases where $\lambda = 0.0$ and 1.0 remains high across N , whereas those in other cases drop with increase in N .

Despite of different natures of the artificial language and processing mechanisms involved in this model, the simulation results show a conclusion similar to those in the previous two models. Compared with the case of random communications, a particular scaling degree ($\lambda = 1.0$) in power-law distributed social popularity can accelerate the diffusion of lexical and syntactic knowledge in population. This effect is less affected by population size. However, other scaling degrees ($\lambda = 0.0$ or $\lambda > 1.0$) will delay the diffusion process and affect linguistic understandability in population. This effect is more explicit in bigger populations. In addition, certain scaling degrees ($\lambda = 0.0$ or 1.0) can also help preserve a high level of linguistic understandability among individuals, but other scaling degrees ($\lambda > 1.0$) will destroy mutual understandability in population. These effects are more explicit in bigger populations.

D. Cross-Model Analysis

These three models examine different aspects of language evolution (e.g., lexical and syntactic evolutions, and origin, diffusion and change of linguistic conventions) and simulate many language processing mechanisms (e.g., lexical and syntactic processing mechanisms). Due to these various linguistic components and processing mechanisms involved, conventionalization of different types of linguistic knowledge (e.g., lexical items, categorization patterns, or syntactic structures) in these models proceeds in different timescales. Nonetheless, there are similar tendencies observed in these models under power-law distributed social popularities. When $\lambda \approx 1.0$, diffusion of linguistic conventions is accelerated; increase in S or UR takes place earlier than that in the case of

random games or communications. Meanwhile, a high level of linguistic understandability is triggered and preserved; S or UR reaches a high value, and a relatively high UR is preserved throughout the simulation. In addition, when $\lambda > 1.0$, linguistic diffusion is delayed or nearly possible, and a high level of linguistic understandability is neither achieved nor maintained. Since λ reflects the scaling degree of power-law distributed social popularity, these simulation results show a *critical scaling degree* ($\lambda \approx 1.0$), under which linguistic conventions can efficiently diffuse in population and a high level of linguistic understandability among individuals can be well preserved. However, below or above this critical degree, language evolution becomes less efficient.

IV. DISCUSSIONS AND CONCLUSIONS

In the above section, we evaluate the effect of power-laws on language evolution, by manipulating power-law distributed social popularities among individuals. The simulation results of three independent evolutionary computational models show that: a critical scaling degree in power-laws ($\lambda \approx 1.0$) can efficiently spread linguistic conventions and keep a high level of linguistic understandability in population, whereas other scaling degrees ($\lambda = 0.0$ or $\lambda > 1.0$) tend to delay the diffusion process and destroy mutual understandability in population, especially in big populations.

These similar results across different models are primarily due to two factors. First, linguistic behaviors in these models help develop conventions and achieve mutual understanding via local communications among individuals. Frequent communications within a small group of individuals can efficiently trigger and share common linguistic knowledge among these individuals. Second, the scaling degree of a power-law distribution adjusts the ratios between two types of communications, including: 1) communications among *popular individuals* (individuals having small rank values) and 2) communications between popular and *unpopular individuals* (individuals having big rank values). Such ratio affects conventionalization of linguistic knowledge among different types of individuals.

Under these two factors, the existence of a critical scaling degree ($\lambda \approx 1.0$) is elaborated as follows. First, in the case of random communications, both of the abovementioned two types of communications are frequent and conventionalization proceeds in the whole population. Then, with increase in population size, conventionalization becomes delayed (see Fig. 1(b), 2(b), and 3(b)).

Second, when λ slightly increases, the first type of communications becomes more frequent, whereas the second type less so. Then, conventionalization in popular individuals becomes easier and faster. Meanwhile, since the second type of communications is also sufficient, unpopular individuals can efficiently acquire the shared knowledge from popular individuals. Compared with the uniform conventionalization in the case of random communications, such “popular first, unpopular later” conventionalization is more efficient, as shown in Figs. 1(a), 2(a), and 3(a).

Third, when λ keeps increasing, the second type of communications becomes insufficient. Then, unpopular

individuals will lack enough chances to interact with popular ones or with each other. Then, although popular individuals can quickly develop common linguistic knowledge, much of this knowledge cannot efficiently diffuse into unpopular ones. Therefore, conventionalization in the whole group is affected. As for the naming game and category game, without knowledge forgetting, more games can give unpopular individuals more chances to communicate with popular ones, thus leading to mutual understanding. As for the lexicon-syntax coevolution model, however, due to competition and regular forgetting of linguistic knowledge, unpopular individuals cannot grasp enough shared knowledge in time, even given more communications. Then, UR remains low in the origin simulations. Meanwhile, in the change simulations, if unpopular individuals do not have enough chances to use initially shared knowledge, that knowledge will gradually be forgotten by those individuals. Then, when λ keeps increasing, UR will drop.

Apart from these two factors primarily concerning the conventionalization nature of language communications in those models, increase in population size also helps select the critical scaling degree. On the one hand, under the critical scaling degree, with increase in N , the trajectories of linguistic understandability do not change much; a high S or UR in big populations can be achieved within the same timescales as in small populations, and in the lexicon-syntax coevolution model, a high UR is kept in both small and big populations. However, under other bigger scaling degrees, with increase in N , a high level of linguistic understandability fails to be efficiently triggered or maintained. These results suggest that the critical scaling degree can accommodate the influence of population size on linguistic understandability, since it can efficiently trigger and maintain a high level of linguistic understandability in big populations.

On the other hand, due to uneven social, economic, or political status among individuals, a uniform distribution of social popularity (equivalent to a power-law distribution with $\lambda = 0.0$) is nearly possible in a big population, and there is a trend of increasing the scaling degree along with population expansion. Meanwhile, population expansion also requires a relatively high and stable level of mutual understandability; otherwise, frequent misunderstanding will affect social bonding among individuals, which may ultimately lead to a split of the whole population. Therefore, in order to maintain a high level of mutual understandability during population expansion, the scaling degree of power-laws could not increase too much, but remain around the critical value.

This general claim summarized from the simulation results can gain some support from empirical data. Recalling the correlation between λ in power-law distributed social popularities and λ' in power-law degree distributions of social networks (see Eq. (1)), we know that the critical λ value around 1.0 corresponds to a critical λ' value around 2.0. As shown in [14], many real-world, scale-free social networks that involve language-related activities do have their λ' around 2.0, e.g., the film actors cooperation network (449,913 nodes, $\lambda'=2.3$), the email network (59,912 nodes, $\lambda'=2.0$) and the telephone call network (47,000,000 nodes, $\lambda'=2.1$). In addition, many natural and technical systems that involve general information

exchange or conventionalization also have their λ' around 2.0, e.g., the metabolic network (765 nodes, $\lambda'=2.2$) and the peer-to-peer network (880 nodes, $\lambda'=2.1$) [14].

Finally, apart from the theoretical findings, let us evaluate the *cross-model comparison* adopted in this study to evaluate the general or universal effect of power-laws on language evolution. Such approach is not frequent in previous work. Previous studies usually adopt a *within-model comparison* approach, focusing on a particular model and evaluating its performances in various conditions (e.g., [29-31][33]). Without further analysis, the conclusions obtained from a single model focusing on a particular aspect of language evolution should not be directly applied to other aspects of language evolution. Besides the within-model comparison, the cross-model comparison as in our study is necessary, especially when one wants to summarize the “universal” effect of certain factor(s) on language evolution in general. Although such cross-model comparison is currently restricted at the conceptual or qualitative level, unification of both within- and cross-model approaches is very promising to reach a comprehensive understanding of the correlation between language evolution and social factors. With many computer models on language evolution being available, more and more studies adopting the cross-model comparison approach are foreseen.

ACKNOWLEDGMENT

We thank Prof. Vittorio Loreto from the Sapienza University of Rome and Prof. Umberto Ansaldi from the University of Hong Kong for their comments on this work.

REFERENCES

[1] C. T. Kello, G. D. A. Brown, R. Ferrer-i-Cancho, J. G. Holden, K. Linkenkaer-Hansen, T. Rhodes, and G. C. van Orden, “Scaling laws in cognitive sciences,” *Trends in Cog. Sci.*, vol. 14, pp. 223–232, May 2010.

[2] A. J. Spence, “Scaling in biology,” *Cur. Bio.*, vol. 19, pp. R57–R61, January 2009.

[3] P. Bak, *How Nature Works: The Science of Self-organized Criticality*. New York: Copernicus, 1996.

[4] B. Mandelbrot, “How long is the coast of Britain? Statistical self-similarity and fractional dimension,” *Science*, vol. 156, pp. 636–638, May 1967.

[5] E. A. Maylor, N. Chater, and G. D. A. Brown, “Scale invariance in the retrieval of retrospective and prospective memories,” *Psycho. Bul. Rev.*, vol. 8, pp. 162–167, March 2001.

[6] G. K. Zipf, *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Cambridge, MA: Addison-Wesley, 1949.

[7] A. Clauset, C. R. Shalizi, and M. E. J. Newman, “Power-law distributions in empirical data,” *SIAM Rev.*, vol. 51, pp. 661–703, February 2009.

[8] T. Gisiger, “Scale invariance in biology: coincidence or footprint of a universal mechanism?” *Bio. Rev.*, vol. 76, pp. 161–209, May 2001.

[9] A. L. Barabási and R. Albert, “Emergence of scaling in random networks,” *Science*, vol. 286, pp. 509–512, October 1999.

[10] D. W. Sims, E. J. Southall, N. E. Humphries, G. C. Hays, C. J. A. Bradshaw, J. W. Pitchford, A. James, M. Z. Ahmed, A. S. Brierley, M. A. Hindell, D. Morritt, M. K. Musy, D. Righton, E. L. C. Shepard, V. J. Wearmouth, R. P. Wilson, M. J. Witt, and J. D. Metcalfe, “Scaling laws

of marine predator search behavior,” *Nature*, vol. 451, pp. 1098–1102, February 2008.

[11] B. F. Grimes, *Ethnologue: Languages of the World*, 14th ed. Dallas: Summer Institute of Linguistics, 2000.

[12] D. Stauffer, C. Schulze, F. W. S. Lima, S. Wichmann, and S. Solomon, “Non-equilibrium and irreversible simulation of competition among languages,” *Physica A*, vol. 371, pp. 719–724, January 2006.

[13] S. Wichmann, “On the power-law distribution of language family sizes,” *J. Ling.*, vol. 41, pp. 117–131, March 2005.

[14] M. E. J. Newman, “The structure and function of complex networks,” *SIAM Rev.*, vol. 45, pp. 167–256, March 2003.

[15] M. E. J. Newman, “Power laws, pareto distributions and Zipf’s law,” *Cont. Phys.*, vol. 46, pp. 323–351, May 2006.

[16] C. P. Warren, L. M. Sander, and I. M. Sokolov, “Geography in a scale-free network model,” *Phy. Rev. E*, vol. 66, pp. 056105, November 2002.

[17] R. Coelho, Z. Nédá, J. J. Ramasco, and M. A. Santos, “A family network model for wealth distribution in societies,” *Physica A*, vol. 353, pp. 515–528, December 2004.

[18] R. Dunbar, *Grooming, Gossip, and the Evolution of Language*. Cambridge, MA: Harvard University Press, 1996.

[19] W. Labov, *Principles of Linguistic Change: Social Factors*. Oxford: Basil Blackwell, 2001.

[20] S. G. Thomason and T. Kaufman, *Language Contact, Creolization, and Genetic Linguistics*. Berkeley: University of California Press, 1988.

[21] A. Baronchelli, M. Felici, V. Loreto, E. Caglioti, and L. Steels, “Sharp transition towards shared vocabularies in multi-agent systems,” *J. Stat. Mech. Theo. Exp.*, vol. 9, pp. P06014, May 2005.

[22] A. Puglisi, A. Baronchelli, and V. Loreto, “Cultural route to the emergence of linguistic categories,” *Proc. Nat. Acad. Sci.*, vol. 105, pp. 7936–7940, June 2008.

[23] T. Gong, *Computational Simulation in Evolutionary Linguistics: A Study on Language Emergence (Frontiers in Linguistics, Monograph IV)*. Taipei: Institute of Linguistics, Academia Sinica, 2009.

[24] T. Gong, “Simulating the coevolution of compositionality and word order regularity,” *Interact. Stud.*, vol. 12, pp. 63–106, April 2011.

[25] T. Malsch and I. Schulz-Schaeffer, “Socionics: sociological concepts for social systems of artificial (and human) agents,” *J. Arti. Socie. Soc. Simu.*, vol. 10, January 2007. Available at <http://jasss/soc/surrey.ac.uk/10/1/11.html>.

[26] D. Livingstone, “The evolution of dialect diversity,” in *Simulating the Evolution of Language*, A. Cangelosi and D. Parisi, Eds.. London: Springer-Verlag, 2002, pp. 99–117.

[27] D. Nettle, *Linguistic Diversity*. Oxford & New York: Oxford University Press, 1999.

[28] S. Kirby, “Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners,” in *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, C. Knight, Ed. Cambridge: Cambridge University Press, 2000, pp. 303–323.

[29] L. Dall’Asta, A. Baronchelli, A. Barrat, and V. Loreto, “Nonequilibrium dynamics of language games on complex networks,” *Phy. Rev. E*, vol. 74, pp. 036105, September 2006.

[30] A. Kalampokis, K. Kosmidis, and P. Argyrakis, “Evolution of vocabulary on scale-free and random networks,” *Physica A*, vol. 379, pp. 665–671, January 2007.

[31] J. Y. Ke, T. Gong, and W. S. Y. Wang, “Language change and social networks,” *Comm. Comp. Phys.*, vol. 3, pp. 935–949, December 2008.

[32] T. Gong, “Exploring the roles of horizontal, vertical, and oblique transmissions in language evolution,” *Adap. Beha.*, vol. 18, pp. 356–376, June/August 2010.

[33] T. Gong, A. Baronchelli, A. Puglisi, and V. Loreto, “Exploring the roles of complex networks in linguistic categorization,” *Artificial Life*, vol. 18, pp. 1–15, Winter 2012.