2006 IEEE Congress on Evolutionary Computation
Sheraton Vancouver Wall Centre Hotel, Vancouver, BC, Canada
July 16-21, 2006

Language Origin and the Effects of Individuals' Popularity

Tao Gong[a], James W. Minett[b], William S-Y Wang[c]
Language Engineering Laboratory, DSP and Speech Technology Laboratory
Department of Electronic Engineering, The Chinese University of Hong Kong
tgong@cuhk.edu.hk[a], jminett@ee.cuhk.edu.hk[b], wsywang@ee.cuhk.edu.hk[c]

***Abstract*—The emergence of a compositional language with a simple grammar and the effects of individuals' popularity on the phylogeny of language are studied based on a multi-agent computational model. In this model, a bottom-up syntactic development is traced, in which the global syntax in sentences is gradually formed from local sequential information. Assuming that the popularity of individuals follows a power-law distribution, we demonstrate that a common language can emerge efficiently only for certain power-law distributions and that these distributions could also be formed as a result of the language phylogeny.**

I. Introduction

THE issue of the origin and evolution of human language has been widely explored using computational modeling [1-2]. Many existing models are behavior-based, in which artificial language users (agents) are equipped with some computational mechanisms to develop individuals' idiolects [3] and to shape the communal language [3]. In each case, a population of agents is able to develop an artificial language with some linguistic features that resemble those in human language after numerous iterated communications. The linguistic features that have been modeled include the lexicon [4], semantic categories [5][6], and grammatical regularity such as linguistic morphology. Compared with the models focusing on lexical communication, there are few simulations studying the evolution of grammatical ability, a universal aspect of natural languages [3]. These models include: Batali's neural network model [7], an early version of the Iterated Learning Model (ILM) [8], and Fluid Construction Grammar (FCG) [9].

First, some of these models have adopted "top-down" learning mechanisms, or have studied the evolution of grammatical ability after the development of lexical items. For example, in ILM, before being fully decomposed, the global (sentence level) structure of the heard utterance is always preserved for future segmentation; agents acquire the complex meanings of utterances directly without reference to their own linguistic knowledge. In FCG, after all agents have come to share sufficiently many lexical items, morphology begins to develop by adding, identifying, and acquiring the morphological constituents of the constructions that consist of these lexical items. However, considering the "tinkerer" view of evolution [10], linguistic competence should develop gradually from the available materials or from some simple abilities, for example, certain biological and cognitive activities, discussed in [10]. Since it is necessary to express complex meanings, complex linguistic features such as syntax or morphology tend to develop along with the learning, production, and comprehension of the language [11]. As indicated by some empirical findings on language acquisition, syntactic development follows a "bottom-up" process (e.g., from a single-word stage to a multiple-word stage [12]) and the acquisition of lexical items and of grammar are inseparable [13]. Any theory of the acquisition of language should mirror this. Besides, the meanings contained in the heard utterances should result from a comprehension process that uses both lexical and grammatical knowledge, and, sometimes, available non-linguistic cues, which might be unreliable.

One possible "bottom-up" syntactic development scenario is implemented here in a model that we have modified from our previous work [14-15]. In that work, both the conventionalization of global syntax and the emergence of compositional linguistic materials were studied. In our new model, we extend that framework to study the formation of global syntax based on simple local orders. In this model, based on previous experience (sentences exchanged in previous communications), agents acquire compositional linguistic knowledge by detecting and learning recurrent patterns in the linguistic input, much like when someone learns a second language [16]. During the acquisition of lexical items, the simple relative orders between pairs of constituents in the input sentences are noticed and acquired as syntactic knowledge. The ability to manipulate simple orders is not specific to language or to humans; even chimpanzees have a similar ability [17]. Through some categorization mechanisms, agents gradually build up syntactic categories that associate sets of lexical items with simple orders. These categories are similar to the verb islands in [18]. Meanwhile, consistent word order in sentences consisting of more than two lexical items can emerge through the use of local orders that regulate these lexical items pair-wise. This linguistic knowledge of lexical items and local orders is used and updated during communications when agents exchange meaningful sentences. After many such communications, based on the local information, a set of common lexical items and a consistent local syntax are established and then diffuse among the agents.

Second, most of the available models of language evolution have adopted dyadic communications in which agents in the population are picked to take part in a particular communication with equal probability. This process disregards the effects of some social factors. Sociological research has already discovered that instead of such uniformity, in many social phenomena, the distribution of the elements or the interactions among them follows a "power-law" distribution (reviewed in [19]), which is also known to characterize a wide range of phenomena in the political and the natural world. These phenomena include sexual contact networks [20], the distribution of votes in legislator elections [21], the spread of epidemic diseases [22], rumors [23], etc. A power-law relationship between two scalar quantities $\boldsymbol{x}$ and $\boldsymbol{y}$ is defined as: $\boldsymbol{y = ax^{-\lambda}}$, where $\boldsymbol{x}$ represents the element or the interaction in a specific phenomenon, and $\boldsymbol{y}$ the frequency of this element or interaction. For example, in the rumor spreading network, $\boldsymbol{x}$ represents the number of people and $\boldsymbol{y}$ the probability for that number of people to spread the rumor. $\boldsymbol{a}$ is a scale parameter. The power-law distributions in different phenomena have different $\boldsymbol{\lambda}$ values. For example, as reported in [24], the $\boldsymbol{\lambda}$ value in the actor collaboration network is 2.3, 2.0 in the email message network, and 2.1 in the telephone call network. Drawn on a log-log graph, a power-law distribution appears as a straight line, whose slope increases with the $\boldsymbol{\lambda}$ value.

To a certain degree, social factors such as friendship, geographical constraints, and political influence can affect the selection of participants in communications. Linguistic communications can also trigger some social, economical or political interactions by causing different individuals to attract and be attracted by others. In addition, some linguistic features have been shown to have power-law distributions, e.g., the distribution of the sizes of different language families [25] and the co-occurrence of syllables in some languages [26]. Therefore, power-law distributions are suitable for describing certain linguistic-related phenomena.

In this paper, we define an *individual's popularity* as the possibility for that individual to participate into linguistic communications. The distribution of an individual's popularity in the community is assumed to follow a power-law distribution. An individual's popularity represents the influence of a variety of social factors, e.g., friendship, political influence, economic incomes. The greater the popularity of an agent, the more frequently he takes part in communications. The computational simulation shows that for certain power-law distributions of individuals' popularity, a common language can be more efficiently triggered in the community for some λ values than for others. We believe that these distributions formed as a result of self-organization during language communication.

The remainder of the paper is organized as follows: Sec. 2 briefly describes the model; Sec. 3 and 4 discuss the emergence of language and the effects of individuals' popularity. Finally, conclusions are summarized in Sec. 5.

II. MODEL DESCRIPTION

The major components of our behavior-based model are introduced in this section, including the representation and acquisition of linguistic knowledge, and the communication scenario.

A. The Representation of Linguistic Knowledge

Language is represented by *Meaning-Utterance mappings* (M-U mappings) in this model. The semantic space contains two types of integrated meaning, each describing a complete event: Type-I: "Pr1<Ag>" (e.g., "hop<deer>") and Type-II: "Pr2<Ag, Pt>" (e.g., "fight<fox, wolf>"). "Pr1" is a predicate (action) having a single argument, "Pr2" is a predicate having two arguments: "Ag", the instigator of the action; and "Pt", the entity undergoing the action. Utterances consist of a string of combinable syllables which can be mapped to either a whole integrated meaning (such a string is called a *sentence*), or one or two semantic items (such strings are called a *word* or a *phrase*, respectively).

Lexical rules

Holistic rules:
(a) "chase<wolf, bear>"←→/a d/ (0.5)
(b) "hop<deer>"←→/a/ (0.4)

Compositional rules:
(c) "wolf"←→/d/ (0.6)
(d) "chase<#, bear>"←→/a b * d/ (0.7)

Detection of recurrent patterns
(1) "hop<fox>"←→/d h/
(2) "run<fox>"←→/d m/
(3) "run<wolf>"←→/a c m/
(4) "fight<wolf, deer>"←→/a c b d/
(5) "fight<wolf, gazelle>"←→/a c b m/
(6) "fight<fox, deer>"←→/d f k b/

New acquired lexical rules
(e) "fox"←→/d/ (0.5)
(f) "run<#>"←→/m/ (0.5)
(g) "wolf"←→/a c/ (0.5)
(h) "fight<wolf, #>"←→/a c b/ (0.5)
(i) "fight<#, #>"←→/b/ (0.5)

Syntactic categories and syntactic rules

Cat1 (S): *Lex-List:* rule (e) [0.5]
rule (g) [0.5]
Syn-List: (I) Cat1 << rule (f) (SV) (0.5)
(II) Cat1 << rule (i) (SV) (0.5)
⇨ (I) Cat1 << Cat2 (SV) (0.5)

Cat2 (V): *Lex-List:* rule (f) [0.5]
rule (i) [0.5]
Syn-List: (II) Cat1 << Cat2 (SV) (0.5)

Fig. 1. Linguistic rules and syntactic categories: "#" can be replaced by other semantic items, and "*" by other syllables. Lexical rules are itemized by letters, M-U mappings by Arabic numerals, and syntactic rules by Roman numerals. Numbers enclosed by ( ) denote rule strengths, and those by [ ] denote association weights. "<<" indicates the relative local order BEFORE.

An agent's linguistic knowledge is represented by rules and syntactic categories (see Fig. 1). Linguistic rules include *lexical* and *syntactic rules*. A lexical rule is a M-U mapping plus a strength, which indicates the probability of that mapping. A lexical rule can be *holistic* or *compositional*. The former is a mapping between an integrated meaning and a sentence (e.g., rules (a) and (b)). The latter is a mapping between a semantic item and a word (e.g., rules (c) and (d)), or between two semantic items that do not form an integrated meaning and a phrase (e.g., rule (d)). A syntactic rule is a relative order plus a strength, which indicates the probability of this relative order. These relative orders include BEFORE and AFTER; "relative" here means it is not necessarily immediately before or after.

Syntactic categories represent the linguistic knowledge of how to regulate sequences of syllables of lexical rules belonging to one category and with those belonging to another. A syntactic category associates the semantic roles ("Ag", "Pr1/2" and "Pat") of the lexical items belonging to it with their respective syntactic roles ("Subject"("S"), "Object"("O"), "Verb"("V")), and contains the list of syntactic rules that apply to its lexical members. The association weight of a lexical rule to a particular syntactic category denotes the probability that the syntactic rules of the category are applied to that lexical rule. One lexical rule can be associated with many categories having identical syntactic roles but with different association weights. Moreover, lexical rules encoding semantic items like "fox" or "wolf" can be associated with both "S" and "O" categories, since in different integrated meanings, they can be either "Ag" or "Pt".

B. *The Acquisition of Linguistic Knowledge*

Lexical rules are acquired through the detection of recurrent patterns. Each agent has a buffer storing some previous experience (a limited list of M-U mappings obtained in previous communications). Newly acquired M-U mappings are compared with those already stored in the buffer before they too are inserted into the buffer. A recurrent pattern is defined as one or more semantic item(s) and one or more syllables that appear recurrently in at least two M-U mappings in the buffer. For instance, in Fig. 1, by comparing M-U mapping (2) with M-U mapping (1), the recurrent pattern "fox"←→/d/ is detected and so acquired as a lexical rule. The segmentation of holistic M-U mappings through the detection of recurrent patterns has been argued to be an effective way to acquire linguistic knowledge [27].

During the acquisition of lexical rules, syntactic rules and syntactic categories are also acquired. Evident in the previous experience (the M-U mappings (2) and (3) in Fig. 1), the words /d/ of rule (e) and /a c/ of rule (g) have the same relative order (BEFORE) with respect to the word /m/ of rule (f). Since "wolf" and "fox" share the same semantic role ("Ag"), rules (e) and (g) are associated into a new category, labeled "S". The association weights are all set initially to 0.5. Meanwhile, the local order (BEFORE) with respect to word (f) is acquired as a syntactic rule (I) in this category. This rule indicates that the words of lexical rules from the "S" category should precede the word (f). Similarly, checking M-U mappings (5) and (6), another syntactic rule (II) with respect to rule (i) is acquired. Furthermore, checking M-U mappings (2) and (6), the words /m/ of rule (f) and /b/ of rule (i) are found AFTER the word /d/ of rule (e). Consequently, a new "V" category associating words (f) and (i), which share the same semantic role "Pr1/2", is created together with a new syntactic rule. Now, since words (f) and (i) are already associated into a category, syntactic rules (I) and (II) are updated into one syntactic rule "Cat1<<Cat2 (0.5)" in both categories. This syntactic rule indicates that the words belonging to the "S" category should precede those of the "V" category. In addition to the creation of new categories, if rules (f) and (i) already belonged to different "V" categories, this previous experience would trigger a merging of the two "V" categories into one category comprising their lexical and syntactic members. Without directly acquiring the global orders in sentences that encode Type-II meanings, agents can use their local orders to regulate the syllables of compositional rules in pairs to build up these sentences. For example, to express "fight<fox, sheep>" based on the lexical rules (i), (e) and another lexical rule expressing "sheep", the SV local order in syntactic rule (I) can be used for regulating the words (i) and (e), another local order, say SO, from another syntactic rule is used for regulating rule (e) and the rule expressing "sheep". Then, the global order based on these local orders can be either SVO or SOV.

The formation of global orders based on local information introduces a certain degree of imprecision: the combination of some local orders can lead to multiple global orders (e.g., SV + SO lead to either SVO or SOV, as shown above), and a particular global order can be represented by the combination of different local orders (e.g., SVO can be represented by SV + VO or SO + VO). This imprecision increases not only the difficulty for a population of agents to acquire a common global word order but also the probability for word order change.

The rule strengths and the association weights make possible the rule competition (discussed later). After a communication, agents subtract a small amount (the forgetting rate) from the strengths of their linguistic rules and the association weights of their lexical rules. Rules that as a result of this subtraction have negative strengths or association weights are removed from the rule list or the syntactic categories containing these rules. The rule competition strengthens and maintains frequently used linguistic knowledge, and causes the language to self-organize.

Through the above acquisition, categorization and adjustment mechanisms, agents first learn some lexical rules and create some independent syntactic categories to associate them. Then, based on the information contained in their previous experience (e.g., recurrent patterns and local orders among them), they acquire more lexical rules, expand the lexical and syntactic members in their categories, and gradually merge categories that share identical syntactic roles. Finally, they may develop a communal language in which all lexical items having the same semantic role belong to the same syntactic category. Moreover, by using similar local syntactic rules to regulate the relative word order of lexical members from the syntactic categories, some common global word orders in sentences might emerge. The whole process simulates a "bottom-up" syntactic development based on the local, partial information available in previous experience.

C. The Communication Scenario

As a language emergence model, we assume that early communications are describing simple events represented by the two types of integrated meanings introduced above. One type of nonlinguistic information, environmental cues, is simulated. Environmental cues, which are represented as integrated meanings plus a fixed strength (0.5), assist the comprehension of the heard utterances. However, cues are not always reliable (otherwise the learning procedure would still involve mind-reading, as in [8]). The probability that one cue corresponds to the speaker's intended meaning is represented by a parameter, the Reliability of Cues (RC).

A communication contains multiple rounds of integrated meaning exchange, each one of which proceeds as follows (for detail, see [14-15]): the speaker chooses an integrated meaning from the semantic space to express and activates certain lexical rules and related syntactic categories with which to encode this integrated meaning. Through a strength-based competition, the agent identifies the winning rules, builds up the utterance accordingly, and transmits the utterance to the listener. If he lacks a set of compositional rules that can represent all the semantic items contained in the chosen integrated meaning, the speaker may occasionally create a holistic rule to express the whole meaning (some models, e.g., [6], adopt a certain exploitation mechanism: if part—but not all—of the integrated meaning can be encoded by compositional rules, new compositional rules will be created to cover the remaining part. This mechanism provides a strong built-in bias toward compositionality).

The listener receives the utterance from the speaker and sometimes some cues from the environment. Then, he activates lexical rules whose syllables fully or partially match the heard utterance, and related categories. The listener selects that set of rules that allow him to comprehend an integrated meaning with the highest combined strength. His calculation of the combined strength considers the strengths of both linguistic rules and the available cues. Then, if the combined strength of the winning rules exceeds a Confidence Threshold (CT), the listener transmits a positive feedback to the speaker, and both agents reward their winning rules by increasing their strengths. Otherwise, a negative feedback is sent and these rules are penalized.

In this communication scenario, there is no direct check whether the speaker's intended meaning matches the listener's comprehended one. The listener's comprehension considers both linguistic and nonlinguistic information, which provides the opportunity for developing reliable linguistic knowledge that can withstand the effect of cues that do not match the speakers intended meaning. A reliable language that is capable of describing events not happing in the immediate space or time can be triggered.

Both linguistic rules and syntactic categories participate in the meaning exchange. For example, during production, the speaker first activates the compositional rules that can be combined to encode the chosen meaning. Then, based on the syntactic categories of these rules, he activates the syntactic rules (OS and VO in Fig. 2) by which these lexical rules can be regulated. Then, he judges which set of linguistic rules wins the strength-based competition, and produces an utterance accordingly. Similarly, during comprehension, after the listener identifies the lexical rules whose syllables partially match the heard utterance, the local orders (VS and OS) that are consistent with the locations of the syllables in the heard utterance (/a b/ before /e f/ and /d/ before /e f/) are detected. If these local orders match the syntactic rules of the categories to which these lexical rules belong, both the syntactic categories and their syntactic rules are activated. Then, based on the activated categories, the semantic roles of these lexical items in the comprehended meaning are specified ("lion" is "Ag", "fox" is "Pt" and "fight<#, #>" is "Pr2"). The calculation of the combined strength also considers the strength of cues that match this comprehended meaning ("fight<lion, fox>"). The listener then judges which set of rules wins the strength-based competition.

In both production and comprehension, under the guidance of syntactic categories, agents use their available building blocks (lexical rules and local orders) to produce utterances and comprehend integrated meanings. This shows how the conceptual-symbolic system (lexical items) and the regulatory system (syntactic categories and local orders) work together closely to process complex linguistic expressions.

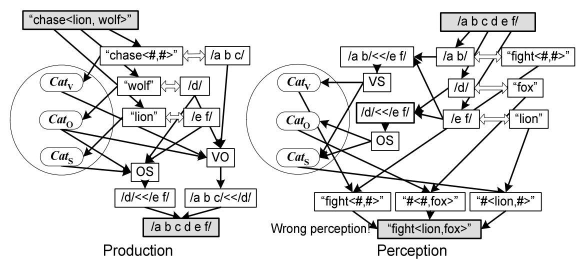


Fig. 2. Categories in production and comprehension: "<<" represents the relative order "before".

III. LANGUAGE EMERGENCE WITH BOTTOM-UP SYNTACTIC DEVELOPMENT

The implementation of the model that we discuss here adopts a semantic space having 16 Type-I and 48 Type-II integrated meanings (consisting of 4 "Ag"="Pt", 4 "Pr1" and 4 "Pr2" elements). Agents have choose to express Type-I and Type-II meanings with equal probability. Rule strengths and connection weights are bounded on the interval [0.0, 1.0], the initial value is 0.5, the update increment is 0.1, and the forgetting rate is 0.01. RC is set to 0.6 and CT to 0.75. Each communication contains 20 meaning exchanges. Each agent's buffer size is set to 40 and the rule list size is set to 60. The community has 10 agents, who, initially, share 8 holistic rules with which they can express 8 integrated meanings—initially, they have no syntactic rules or categories.

The agents then begin to communicate with each other in randomly selected pairs. The total number of communications is 5,000. Every 50 communications, the Rule Expressivity (RE, the average percentage of the total integrated meanings that all agents can produce), the understandability of each global and local order (the average percentage of Type-II integrated meanings that are comprehended using that global or local order), and the Understanding Rate (UR, the average percentage of the total integrated meanings accurately understood using linguistic knowledge without cues) are calculated.

The results for the above parameter settings are shown in Fig. 3. During iterated communications, agents can acquire new linguistic materials, which increase the RE of both holistic and compositional rules. After a number of communications, the original and newly acquired holistic expressions are replaced by the newly acquired compositional ones; the RE of the holistic rules drops and the RE of the compositional rules gradually rises to almost 100%. Meanwhile, the diffusion of rules among the agents increases the understandability of the emergent language. During the period in which the compositional rules and holistic rules are competing, the UR follows a U-shaped curve. However, when the holistic rules come to be replaced by the compositional rules, the UR curve exhibits a sharp, S-shaped growth as the population negotiates a set of common compositional rules, indicating the transition from an initial holistic signaling system to a compositional language.

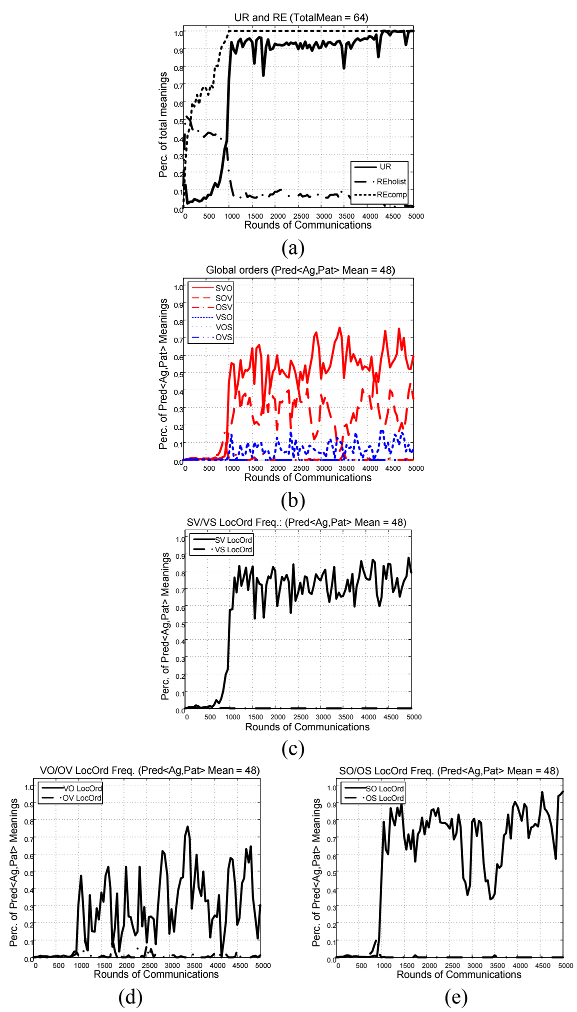


Fig. 3. Language emergence results: (a) UR and RE; (b) Global orders understandability (Type-II meanings, for Type-I meanings, see (c)); (c)(d)(e): Local orders understandability (SV/VS, VO/OV, SO/OS).

During the acquisition of common compositional rules, a certain degree of regularity emerges. Fig. 3(b-e) shows the understandability of different global and local orders among the agents. Along with the increase of UR, some global orders (SVO and SOV) become prevalent, i.e., most agents use these orders to comprehend most integrated meanings. From Fig. 3(c-e), it can be seen that these global orders result from the combination of the prevalent local orders (SV and SO). The absolute strengths of the local orders VO and OV fluctuate, but have no significant impact on the global word order. This indicates that the VO and OV local orders are not unified among agents. If the average strength of the VO local order were to greatly exceed that of OV, the number of meanings comprehended using SVO would then greatly exceeds that of meanings comprehended using SOV. The emergent global syntax is merely an emergent property of the simple sequential information specified by the local order. Changes in the surface word order result from changes in the local sequential information; a detailed study of the influence of local orders on the emergent global word order is given in [28]. Furthermore, changes in the strengths of the local orders do not greatly influence the understanding rate (UR), which suggests that the whole system can efficiently and robustly adapt to new situations.

Language emergence in this model can be viewed as a self-organizing process: each individual's organization of his own linguistic knowledge leads to convergence in the language of the entire population. Communications provide opportunities for linguistic knowledge to diffuse among agents. The acquisition of compositional rules and the development of local orders boost each other and are achieved simultaneously. The model also shows the viability of a bottom-up syntactic development process; complex linguistic features can develop based on some general competences which might not be unique to humans (such as detecting recurrent patterns and manipulating simple, local sequential information).

The model only simulates horizontal transmission (language communication among agents in the same generation), but can be easily extended to simulate vertical transmission across generations.

IV. EFFECTS OF INDIVIDUAL'S POPULARITY

In this section, we study language emergence in the community for different power-law distributions of agent popularity. In these distributions, the value of the ***a*** parameter is set to 1.0; the value of λ lies in the interval [0.0, 3.0], which corresponds to the range of λ values observed in many real world power-law distributions. $\lambda = 0$ corresponds to the case in which every agent has equal chances to communicate with each other. The particular power-law distributions of agent popularity that we investigate are shown in Tab. 1 and Fig. 4. On a logarithmic scale, as λ is increased, so the distribution line becomes steeper, i.e., most agents are significantly less popular than the few agents having high popularity. These few agents frequently communicate with each other and with other agents.

TABLE I.
POPULARITY OF EACH AGENT IN DIFFERENT POWER-LAW DISTRIBUTION

Index / λ	0	1	2	3
1	0.1	0.342	0.645	0.835
2	0.1	0.171	0.161	0.105
3	0.1	0.114	0.072	0.031
4	0.1	0.085	0.040	0.013
5	0.1	0.068	0.026	0.007
6	0.1	0.057	0.018	0.004
7	0.1	0.049	0.013	0.002
8	0.1	0.043	0.010	0.002
9	0.1	0.038	0.008	0.001
10	0.1	0.034	0.006	0.0008

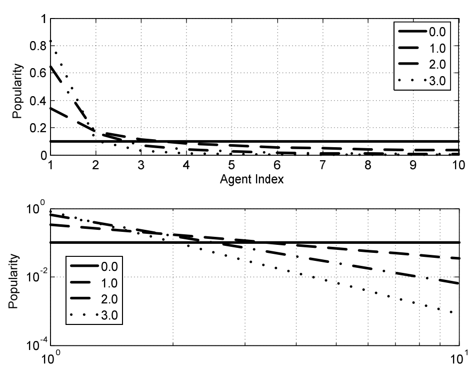
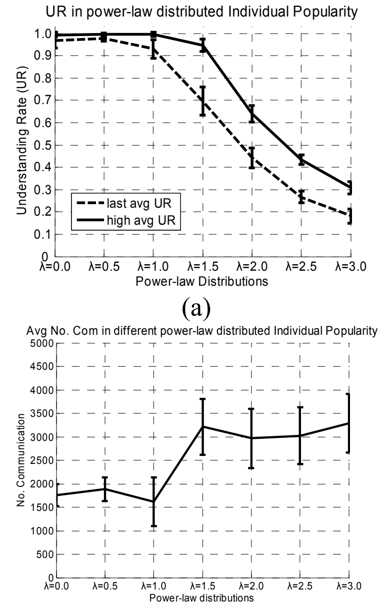


Fig. 4. Power-law distribution of popularity: the upper panel shows the distribution on linearly scaled axes, the lower on logarithmically scaled axes.

The UR value and the number of communications required for the UR to reach its peak value are recorded for different power-law distributions. Fig. 5 shows these results for a 10-agent population. In Fig. 5(a), the solid line traces the average peak UR of the emergent language for each distribution; the dashed line traces the average UR of the emergent language at the end of 5,000 communications. Fig. 5(b) traces the average number of communications for the UR to reach its peak value for each distribution.



(a)

(b)

Fig. 5. Language emergence in different power-law distributions of popularity: (a) average UR in different power-law distributions; (b) average number of communications for UR to reach its peak value. Under each power-law distribution, the results of 20 simulations are collected.

For small λ values (within the interval [0.0 1.5]), a language with high UR emerge. The emergence of high UR is achieved in fewer than 3,000 communications. After 5,000 communications, the high understandability is maintained. With the increase of the value of λ from 1.5 to 3.0, not only does the understandability of the emergent language begin to drop, but also the number of communications required to reach the peak UR value increases. When λ is set to 3.0, the understandability of the language drops to 40%; after 5,000 communications, the understandability falls to below 20%. In a 10-agent community, steeper power-law distributions of agent popularity provide a reduced probability for a common language to emerge, and understandability of the emergent language is rarely maintained.

Linguistic communications provide opportunities for salient linguistic materials created by different individuals to diffuse across the population. Given sufficiently many communications, the development of common linguistic knowledge and mutual understanding based on this knowledge can be achieved. Therefore, in order to develop a common language, each agent is required to have sufficient opportunity to communicate with other agents in order to conventionalize his idiolect to the communal language.

For power-law distributions with small λ values, the absolute value of each agent's popularity is still high and the difference of popularity among agents is not that significant, i.e., each agent has the opportunity to communicate with each other agent. Therefore, a communal language with high UR can emerge under these conditions.

When the value of λ is high, the opportunity for many agents to take part in communication drops significantly.

Besides, the majority of agents tend only to interact with the few extremely popular agents. In this situation, those popular agents act as hubs connecting closely among themselves and loosely with the unpopular ones. Some research on complex networks [29] has demonstrated that by introducing hubs, information transmission among the nodes and synchronization of the whole network can be accelerated. A similar acceleration effect exists in our model, although, there are other factors which may reduce the impact of the acceleration effect.

First, considering the imprecision introduced by the local orders, although it is easier to share lexical items through frequent communications, the syntactic knowledge may not easily converge. Based on different previous experiences agents may develop different local orders from the same global syntax. Furthermore, the global syntax may diverge when one agent changes his local orders as a result of communicating with another. Therefore, the information transmission via the popular agents is not always efficient transmitted to the entire population.

Second, affective transmission of information through the hubs requires that these hubs be stable. However, these popular agents are also language learners themselves who continue to update their idiolects during communication. If some popular agents update their linguistic knowledge as a result of listening to some unpopular agents whose idiolects differ greatly from their own, then in future, when the popular agents talk to others, the communal language shared by these agents might be damaged. Not only does the understandability of the communal language drop, but also more communications are needed either to recover the originally shared linguistic knowledge or to diffuse the linguistic knowledge newly acquired from the unpopular agents.

The above two factors can both delay the emergence of a communal language and reduce the understandability of the emergent language. These acceleration and deceleration effects coexist and compete with each other in our model. In a 10-agent community, it is shown that the deceleration effects becomes more obvious as the value of λ is increased. The optimal value of λ for triggering a language with high UR in this population lies within the interval [0.0 1.5]. The optimal value for λ differs according to the population size. Fig. 6 shows the UR for a 50-agent population after 150,000 communications. Here, the optimal λ value is around 2.0, which matches the λ values of many social phenomena's power-law distributions, e.g., the email network, the telephone call network [24]. In those networks, the community sizes are much greater than 50. In addition, in Fig. 6, where λ = 0, the UR value is just around 50%. This indicates that in communities having larger population sizes, it is unrealistic to assume that each agent has equal opportunities to participate into communications. Under this assumption, high understanding rate will not be obtained, and, as shown in many social networks, such uniform structure does not exist.

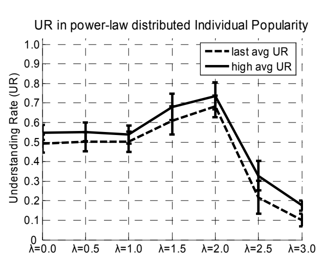


Fig. 6. Language emergence in different power-law distributions of popularity in the community of 50 agents. Under each power-law distribution, the results of 10 simulations are collected.

In the above simulations, the power-law distribution of individual popularity is predefined without discussing how these distributions are formed. There are some explanations for the formation of the power-law distributions in complex networks, social and physical phenomena. For example, preferential attachment [30] is claimed to be a key mechanism to form a scale-free network with power-law degree distribution. Many social phenomena are simulated based on similar preferential attachment mechanisms (e.g., [21][23]). Physicists believe that the power-law distribution is the characteristic of many self-organizing systems [31]. In addition, factors like geographical constraints [32] and kinship relations [33] can also trigger certain power-law distributions.

Besides these factors, we suggest that communications during the phylogenetic emergence of language can be another factor to trigger such power-law distribution of individual popularity. Self-organization during linguistic communications and mutual understanding based on the evolving language can adjust the possibility for an individual to participate in future communications. For example, an agent whose language is understandable to others would gradually gain more opportunities to communicate with others. And, similar to preferential attachment, others will tend to prefer to communicate with these popular agents. Then, a scaling emerges among the originally equal popularities, and a power-law distribution of individual's popularity could be formed.

In other words, the phylogeny of language and the individual's popularity may coevolve. Some preliminary work (e.g., [34]) has already touched upon the coevolution of language emergence and the formation of certain social structure.

V. Conclusions

By implementing a "bottom-up" syntactic developmental process, we have shown that some complex linguistic features can develop based on general competences not unique to humans. This supports Emergentism [36] rather than Innatism [35].

We have studied the influence of individuals' popularity on language emergence. We suggest that different power-law distributed popularities can affect the emergence of the communal language in different communities, and that the process of phylogenetic emergence of language may have selected for certain power-law distributions.

Besides the understandability, it is necessary to observe whether the global word order is also influenced by the individuals' popularity. Furthermore, in communities having greater population, it is interesting to know whether the distribution of individuals' popularity can trigger a segmentation of the whole group into subgroups and whether a common language can emerge in those subgroups. The emergence of different languages in different subgroups could maintain the global social structure and the popularity distribution. All these questions provide promising future directions for this research.

Acknowledgment

The authors would like to thank Professors John H. Holland and Thomas Lee, and lab mates Francis Wong and Susan Shuai for useful comments and suggestions. The research is supported by research grants from the RGC Hong Kong: CUHK-1224/02H & CUHK-1127/04H.

References

[1] A. Cangelosi and D. Parasi, Eds., *Simulating the evolution of language*. London: Springer, 2002.

[2] M. H. Christiansen and S. Kirby, Eds., *Language evolution*. Oxford: Oxford University Press, 2003.

[3] L. Steels, "The emergence and evolution of linguistic structure: from lexical to grammatical communications," *Connection Science*, vol. 17, pp. 213-230, 2005.

[4] J. Ke, J. W. Minett, C. Au, and W. S-Y Wang, "Self-organization and selection in the emergence of vocabulary," *Complexity*, vol. 7, pp. 41-54, 2002.

[5] T. Belpaeme and J. Bleys, "Explaining universal color categories through a constrained acquisition process," *Adaptive Behavior*, vol. 13, pp. 293-310, 2005.

[6] P. Vogt, "The emergence of compositional structures in perceptually grounded language games," *Artificial Intelligence*, vol. 167, pp. 206-242, 2005.

[7] J. Batali, "The negotiation and acquisition of recursive grammars as a result of competition among exemplars," in *Linguistic Evolution through Language Acquisition: Formal and computational Models*, T. Briscoe, Ed., Cambridge, MA: Cambridge University Press, 2002, pp. 111-172.

[8] S. Kirby, "Learning, bottlenecks and the evolution of recursive syntax," in *Linguistic Evolution through Language Acquisition: Formal and computational Models*, T. Briscoe, Ed., Cambridge, MA: Cambridge University Press, 2002, pp. 173-204.

[9] L. Steels, "Constructivist development of grounded construction grammars," in Proceedings of the Annual Meeting of Association for Computational Linguistics, W. Daelemans, Ed., Barcelona, Spain: MIT Press, 2004.

[10] F. Jacob, "Evolution and tinkering," *Science*, vol. 4295, pp. 1161-1166, 1977.

[11] T. P. Schoenemann, "Syntax as an emergent characteristic of the evolution of semantic complexity," *Minds and Machines*, vol. 9, pp. 309-346, 1999.

[12] E. V. Clark, *First language acquisition*. Cambridge, MA: Cambridge University Press, 2003.

[13] E. Bates and J. C. Goodman, "On the inseparability of grammar and the lexicon: evidence from acquisition" in *Language development: the essential readings,* M. Tomasello and E. Bates, Eds., Oxford: Blackwell, 2001, pp. 134-162.

[14] T. Gong and W. S-Y Wang, "Computational modeling on language emergence: a coevolution model of lexicon, syntax and social structure," *Language and Linguistics*, vol. 6, pp. 1-41, 2005.

[15] T. Gong, J. W. Minett, J. Ke, J. H. Holland, and W. S-Y Wang, "A computational model of the coevolution of lexicon and syntax," *Complexity*, vol. 10, pp. 50-62, 2005.

[16] L. W. Fillmore, "Individual differences in second language acquisition," in *Individual differences in language ability and language behavior*, C. J. Fillmore, D. Kempler, and W. S-Y Wang, Eds., New York: Academic Press, 1979, pp. 203-228.

[17] M. D. Hauser, *The evolution of communication*. Cambridge, MA: MIT Press, 1996.

[18] M. Tomasello, *Constructing a language: a usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press, 2003.

[19] M. E. J. Newman, "Power laws, pareto distributions and Zipf's law," *Contemporary Physics*, vol. 46, pp. 323-351, 2005.

[20] F. Liljeros, C. R. Edling, and L. A. N. Amaral, "Sexual networks: implications for the transmission of sexually transmitted infections," Microbes and Infection, vol. 5, pp. 189-196, 2003.

[21] H. Situngkir, "Power law signature in Indonesian legislative election 1999-2004," Research Paper WPL2004, Bandung Fe Institute, 2004.

[22] M. D. F. Shirley and S. P. Rushton, "The impact of network topology on disease spread," *Ecological Complexity*, vol. 2, pp. 287-299, 2005.

[23] Y. Moreno, M. Nekovee, and A. F. Pacheco, "Dynamics of rumor spreading in complex networks," Physical Review E, vol. 69, pp. 066130: 1-7, 2004.

[24] M. E. J. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, pp. 167-256, 2003.

[25] S. Wichmann, "On the power-law distribution of language family sizes," *Journal of Linguistics*, vol. 41, pp. 117-131, 2005.

[26] M. M. Soares, G. Corso, and L. S. Luneca, "The network of syllables in Portuguese," *Physica A*, vol. 335, pp. 678-684, 2005.

[27] A. Wray, *Formulaic language and the lexicon*. Cambridge, MA: Cambridge University Press, 2002.

[28] J. W. Minett, T. Gong, and W. S-Y Wang, "A language emergence model predicts word order bias," in *Proceedings of sixth international conference on language evolution*, Rome, Italy. In press.

[29] P. Crucitti, V. Latora, M. Marchiori, and A. Rapisarda, "Efficiency of scale-free networks: error and attack tolerance," *Physica A*, vol. 320, pp. 622-642, 2003.

[30] A-L Barabási and R. Albert, "Emergence of scaling in random networks," Science, vol. 286, pp. 509-512, 1999.

[31] P. Bak, How nature works: the science of self-organized criticality. New York: Copernicus, 1996.

[32] C. P. Warren, L. M. Sander, and I. M. Sokolov, "Geography in a scale-free network model," Physical Review E, vol. 66, pp. 056102: 1-5, 2002.

[33] R. Coelho, Z. Néda, J. J. Ramasco, and M. A. Santos, "A family-network model for wealth distribution in societies," Physica A, vol. 353, pp. 515-528, 2005.

[34] T. Gong, J. Ke, J. W. Minett, and W. S-Y Wang, "A computational framework to simulate the coevolution of language and social structure" in *Proceedings of the IX International Conference on the Simulation and Synthesis of Living Systems*, J. Pollack, M. Bedau, P. Husbands, T. Ikegami, and R. Watson, Eds., Boston, USA: MIT Press, 2004.

[35] S. R. Anderson and D. W. Lightfoot, *The language organ*. Cambridge, MA: Cambridge University Press, 2002.

[36] C. Knight, M. Studdert-Kennedy, and J. R. Hurford, Eds., *The evolutionary emergence of language: social function and the origins of linguistic form*. Oxford: Oxford University Press, 2000.