

Coevolution of Language and Intentionality Sharing

Tao Gong, *Member, IEEE*, James W. Minett, *Member, IEEE*, and William S-Y. Wang

Abstract—We conduct an evolutionary simulation to explore the coevolution of language and a language-related ability, intentionality sharing. Our simulation shows that during the evolution of a simple informative language, communicative success helps optimize the level of intentionality sharing in the population. This study illustrates a selective role of language communications on language-related abilities, and assists the discussion of the uniqueness of language-related abilities based on comparative studies.

I. INTRODUCTION

LANGUAGE evolution takes place in *cultural transmissions* (communications among language users of the same or different generations [1]). Many empirical studies (e.g. [2]) have revealed a “*mosaic*” fashion of language evolution: a number of abilities (e.g., social cognition, imitation, etc.) all contribute significantly to language evolution at different times and with different levels [3]. The nature of these language-related competences has been widely discussed in linguistics, psychology and other disciplines (e.g., [2][4][5]). One such competence that is crucial for developing a human communication system is *intentionality sharing*; to recognize other individuals as intentional agents whose attention and behavior could be shared or manipulated [4]. Comparative studies between non-human primates and human infants have shown that the former rarely use their gestures or calls *referentially* (i.e., to attract the attention of others to some outside entities or events) or for *declarative* purpose (i.e., to direct the attention of others to something for the sake of sharing interests in it or commenting on it) [6]. In contrast, the latter acquire their languages through activities of *joint attention* that take place around 9 to 12 months of age. In these activities, human infants can share and direct the attentions of other persons. Based on this behavioral evidence, Tomasello [6] has developed a scenario of language evolution, which starts from intentional gestures of great apes, followed by requesting gestures of early hominids,

a simple language for informing events now and here, and a complex language with sophisticated grammar to narrate a series of events. Considering that intentionality sharing is prerequisite in all stages of development and there is a huge gap between the levels of this ability in chimpanzees and human infants, Tomasello and colleagues have claimed that intentionality sharing must be human-unique and language evolution rests crucially on this social cognitive skill [6][7].

An impressive level difference between modern humans and other primates in certain skills is insufficient to indicate the uniqueness of those skills in humans, because the level difference could result from a gradual evolution along with the development of a primitive communication system. Based on the “*socio-biological*” explanation [8], individuals having more successful communications could gain a selective advantage (e.g., a survival [9] or mating benefit [10]) so that they could produce more offspring than others and spread some of their strategies or abilities that lead to successful communications in future generations of individuals. The pace of language evolution is usually faster than that of the biological capacities for human language. But if certain abilities have already been present, at least as precursors, in both humans and our closest relatives in the animal kingdom, the adjustment on these abilities may proceed along with the emergence of a linguistic communication system. The mirror neuron system recently-found in non-human primates that enables these animals to grasp the intentions of some particular activities [11][12] could be the precursor for intentionality sharing, and the significant level of this ability in humans might be caused by the selective advantage of communicative acts during language evolution.

Lacking direct evidence on intermediate stages of language evolution restricts the anthropological studies to investigate the “*socio-biological*” explanation on the development of language and related abilities in humans. In addition, the comparative studies based on animal models can merely offer limited usefulness to illustrate this process [13], because non-human apes today only stay at a primitive level of intentionality sharing. Apart from these studies, evolutionary computation provides a promising way to tackle this problem. This approach can reasonably recapitulate the intermediate stages of language evolution and recover the missing links of language-related abilities between early hominids and modern humans [14]. The optimization mechanisms in evolutionary computation can conceptually simulate the selective role based on communicative success on language-related abilities. Although the real history may not follow exactly the same process shown in a simplified,

Manuscript received October 31, 2008. This work was supported in part by grants from 1224/02H and 1127/04H awarded by the Research Grants Council of Hong Kong. Gong is supported by the Alexander von Humboldt Foundation in Germany.

Tao Gong is with Department of Linguistics, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, Leipzig, 04103 Germany (phone: +49-341-3550-319; fax: +49-341-3550-311; e-mail: tao_gong@eva.mpg.de).

James W. Minett is with Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong (e-mail: jminett@ee.cuhk.edu.hk).

William S-Y. Wang is with Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong (e-mail: wsywang@ee.cuhk.edu.hk).

abstract model, such exploratory simulation can verify the internal coherence of the proposed theories and extend our insights in the discussion of the theoretical problems that lack direct evidence to evaluate.

We propose in this paper a multi-agent computational model to explore the coevolution [15] of language and intentionality sharing. Based on Tomasello's scenario on language evolution, we confine our model to simulate the emergence of a compositional language for informing events now and here. Intentionality sharing at this stage is boiled down to the availability of the topics from the immediate environment of communicative acts. We propose a framework that includes both cultural transmissions among individuals of the same generation and the genetic algorithm to adjust, during generation replacement, the levels of intentionality sharing among individuals.

A comparison of the results with and without optimization has revealed some understanding that may shed light on the uniqueness of intentionality sharing in humans:

- 1) The minimum level of intentionality sharing required to trigger a communal language with a good understandability is not very high;
- 2) Communicative success can help optimize the level of intentionality sharing in the population to assist language evolution;
- 3) The optimization triggers the emergence of *displacement* in the communal language (the human language can efficiently describe the events not happening in the immediate environment of the conversation [16]), which paves the way for language evolution in the next stage where complex grammar emerges for narrating a series of events that are not ongoing.

The rest of the paper is structured as follows: Sec. 2 briefly reviews the language evolution model; Sec. 3 describes the framework to explore the coevolution and introduces the simulation setup; Sec. 4 summarizes the simulation results; and finally, Sec. 5 discusses these results and lists the conclusions.

II. THE LANGUAGE EVOLUTION MODEL

This computational model simulates how a population of interacting individuals gradually develops a communal language as a result of some general learning mechanisms, such as pattern detection and local order manipulation. The emergent language consists of a set of common lexical items and some consistent word order(s). It is well suited to describing simple integrated events and its development resembles the emergence of a simple informative language in Tomasello's scenario. A conceptual description of this model is given below, with the focus on the communicative acts that involve intentionality sharing. A detailed description of this model and the empirical bases of its mechanisms are in [17][18].

A. Language and Linguistic Knowledge

This model deals with the development of idiolects (an individual's linguistic knowledge) in the population. An idiolect is represented by 4 components: a *semantic space*, a *lexicon*, a set of *categories*, and a *syntax*.

All individuals share the same semantic space, which consists of a finite set of simple integrated meanings. During communications, individuals produce utterances to inform each other of these meanings. An integrated meaning consists of a predicate together with its one or two arguments: an agent and a patient. It is denoted, for example, by "run<wolf>" (meaning that "a wolf is running") or "chase<tiger, fox>" (meaning that "a tiger is chasing a fox"), in the latter example, the two arguments of the predicate "chase" are ordered, the first argument representing the *agent* (the entity that performs the action of the predicate) and the second representing the *patient* (the entity that undergoes the action).

The lexicon of an individual may contain both holistic and compositional rules. A *holistic rule* maps directly between an integrated meaning and a fully-formed utterance (a *sentence*). It is denoted, for example, by "run<tiger>" \leftrightarrow /abcd/, which indicates that the meaning "run<tiger>" may be produced as the utterance /abcd/ (a string of four syllables: 'a', 'b', 'c', and 'd') and also that /abcd/ may be comprehended as "run<tiger>". In other words, this mapping is bidirectional. A *compositional rule* maps between particular semantic constituent(s) and a sub-part of an utterance (a *word* or a *phrase*). It is denoted, for example, by "fox" \leftrightarrow /ef/. This mapping is also bidirectional. New lexical rules are formed whenever an individual perceives a novel recurrent pattern in both the meaning and utterance parts of at least two meaning-utterance mappings. Each individual has a buffer that stores a number of meaning-utterance mappings acquired in his/her previous communications with others. By comparing these mappings in pairs, for example, "run<fox>" \leftrightarrow /ab/ and "chase<fox, deer>" \leftrightarrow /bce/, this individual can encode the recurrent pattern "fox" and /b/ as a lexical rule "fox" \leftrightarrow /b/.

Holistic rules allow individuals to produce meaningful utterances (and comprehend perceived sentences) directly. However, the use of compositional rules in production requires that the rules be regulated in order so that they can form a meaningful sentence. In terms of meaning, a set of compositional rules may combine if they specify each constituent of an integrated meaning exactly once (i.e., 1 predicate, 1 agent, 1 patient). For example, two compositional rules with meanings "chase<tiger, #>" (the symbol # refers to either an unspecified or a variable meaning constituent) and "fox" can combine to form "chase<tiger, fox>". But the two rules "chase<tiger, #>" and "chase<#, fox>" cannot combine because the predicate is specified twice. In terms of utterance, the order of words or phrases in an utterance is regulated by the syntax. An individual's syntax consists of a set of syntactic rules, each specifying a relative, or local, order of

utterance items. For example, “tiger” << “fox” denotes that the meaning constituent “tiger” should be produced in the utterance before — but not necessarily immediately before — the constituent “fox”. This syntactic rule can also be denoted by “fox” >> “tiger”. This local order approach corresponds to the simple syntax in an informative language. Using a local order plus two lexical items, events like “run<wolf>” can be expressed. Similarly, by reiterating two or three local orders that regulate three lexical items, events like “chase<tiger, fox>” can be expressed.

Categories gradually form in order for syntactic rules acquired for some words to be applied productively to other words having the same thematic relation (e.g., patient). A category consists of both a set of lexical items and a set of syntactic rules that may operate on all those lexical items and regulate their orders with respect to lexical items from other categories. These categories resemble the “verb islands” [19] formed as children gradually learn the constraints that apply to particular verbs and generalize them to apply to other verbs. A new category is formed when an individual observes that two meaning constituents having the same thematic relation in two different sentences follow the same order with respect to another constituent that appears in both sentences. For example, if in two meaning-utterance mappings in an individual’s buffer, the predicate “chase” comes before the agent “fox” in one sentence and before the agent “wolf” in another, then a new agent category is formed comprising both “fox” and “wolf”. This category is also referred to as a subject (S) category since the thematic relation of agent corresponds to the syntactic role of subject in sentences. Similarly, patient corresponds to object (O), and predicate to verb (V). In other words, the language that we simulate in this model is nominative-accusative, and all sentences are in active voice. In addition to the creation of this category, the individual also acquires a syntactic rule that regulates the members of this category to appear after the constituent “chase” in utterances. As other lexical items are absorbed into this category, the syntactic rules of the category may be applied to them as well. A syntactic rule that defines a local order between lexical members of two distinct categories can be denoted in terms of the syntactic roles of those categories. For example, a syntactic rule indicating members of an agent (S) category come before those of a predicate (V) category can be denoted as the local order S << V, or simply SV.

In this model, each lexical or syntactic rule has a *strength* and a lexical rule has an *association weight* to the category that contains this rule. The values of these strengths and association weights lie in [0.0 1.0]. The initial strengths of newly-acquired rules and the initial weights of new category associations are 0.5. These numerical parameters allow us to simulate a strength-based rule competition during linguistic communications, and a gradual loss of rarely-used linguistic knowledge. Both of these mechanisms serve to strengthen the frequently-used linguistic knowledge and to cause the self-organization of dialects toward a communal language.

Fig. 1 shows some examples of lexical rules, syntactic rules and categories. Based on this set of knowledge, if an individual wants to express “fight<wolf, fox>” using the lexical rules respectively from the three categories, the local orders SV and SO lead to two global orders SVO and SOV, then, the created utterance could be either /bcea/ or /bcae/.

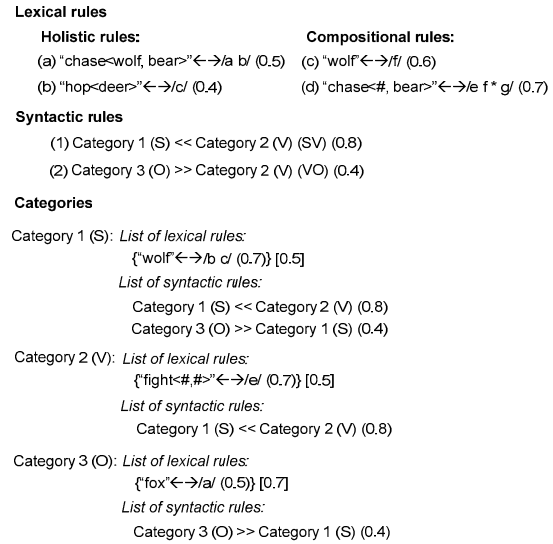


Fig. 1. Some examples of lexical rules (itemized by letters), syntactic rules (itemized by numbers) and categories. “S”, “V”, and “O” denote the syntactic roles of categories. Numbers enclosed by () denote rule strengths, those by [] association weights. “<<” indicates the relative local order “before”, “>>” is “after”.

B. Communication, Intentionality Sharing, and Rule Competition

Individuals acquire and apply their linguistic knowledge in iterated communications, each involving two randomly chosen individuals (a speaker and a listener) who perform a number of utterance exchanges. We assume that during the emergence of an informative language, intentionality sharing may assist comprehension by acquiring the meanings of linguistic utterances from some nonlinguistic information, via gaze following, pointing or other informative gestures. In an utterance exchange with a shared intention, the listener acquires the speaker’s intended meaning from a “correct” cue; in an utterance exchange without shared intentions, the listener acquires a “wrong” cue that simply contains a meaning distinct from the speaker’s intended one. In this model, a *cue* is represented by an integrated meaning plus a fixed strength. We use *reliability of cue* (RC) to manipulate the probability for speaker’s intended meaning to be available to the listener. For the speaker, RC indicates the probability of choosing the *ongoing events* that can be acquired by nonlinguistic information as the topics in communicative acts; for the listener, it refers to the probability of using the ongoing events to comprehend the heard utterances. For instance, if RC is 0.6, in an utterance exchange, there is a 60% percent chance that the speaker’s intended meaning is available to the listener via a cue. Although a cue participates

in comprehension, comprehension is not solely determined by cues; rather, it is a process determined by both linguistic and nonlinguistic information. By allowing changes in RC values, we can evaluate the relation between communication and intentionality sharing.

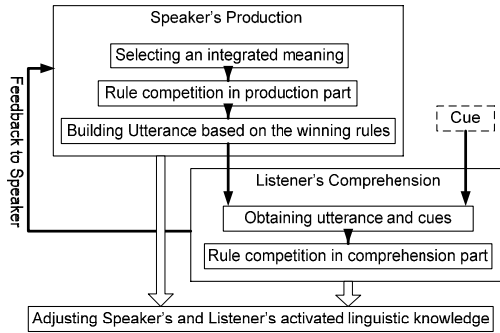


Fig. 2. The utterance exchange during communications.

An utterance exchange proceeds as follows (see Fig. 2): the speaker (referred to as “she”) first chooses an integrated meaning to express. She then activates the lexical rules that can encode some or all constituents in the intended meaning, as well as the syntactic rules and related categories by which the lexical rules can be combined and ordered to form a sentence. The activated linguistic rules form candidate sets for production, each allowing her to produce an utterance to encode the intended meaning. Following (1),

$$CS_{production} = Avg(str(LexRule(s))) + Avg(aso(Cats) \times str(SynRule(s))) \quad (1)$$

where “Avg” means taking average, “aso” taking the association weights of related lexical rules to related categories, and “str” taking the rule strengths of related lexical or syntactic rules, “LexRule”, “SynRule”, “Cat” representing related lexical rules, syntactic rules, and categories), the speaker calculates the *combined strength* ($CS_{production}$) of each set, which is the sum of two parts. The first part concerns lexical information, which is calculated as the average strength of the lexical rules in the set. The second part concerns syntactic information, which is the average product of two elements: the first element is the strengths of the syntactic rules that are used to regulate the syllables of the lexical rules in this set, and the second element is the association weights of those lexical rules to the categories in this set. The speaker identifies her *set of winning rules* that has the highest $CS_{production}$, builds up the sentence accordingly, and transmits it to the listener. In this model, we assume that an informative language starts from a holistic signaling system. Lexical items and simple syntax, which have a higher efficiency to express integrated meanings, can be gradually acquired by individuals using their learning mechanisms. If the speaker cannot construct a sentence to encode the intended meaning, she may occasionally create a holistic rule to express the entire meaning.

After receiving the speaker’s sentence and the cue, the listener (referred to as “he”) activates the lexical rules whose syllables fully or partially match the heard sentence, the categories that associate those lexical rules, and the syntactic rules of these categories that are consistent with the heard sentence. These activated linguistic rules form candidate sets, each providing an integrated meaning for comprehension.

Here, intentionality sharing affects comprehension via the cue. If the meaning in the cue exactly matches the one provided by a particular candidate set, the cue is combined with that set of linguistic rules. If some linguistic rules fail to provide an integrated meaning, but the meaning in the cue matches those constituent(s) specified by these rules, the cue is combined with those linguistic rules to form a candidate set and the cue’s meaning becomes the meaning of this set. Moreover, if the available rules cannot form an integrated meaning and the meaning of the cue does not match any constituent specified by those rules, or the listener simply has no linguistic rules, the cue itself forms a candidate set and its meaning becomes the meaning of this set. After that, following (2),

$$CS_{comprehension} = Avg(str(LexRule(s))) + Avg(aso(Cats) \times str(SynRule(s))) + str(Cue) \quad (2)$$

the listener calculates $CS_{comprehension}$ of each candidate set. $CS_{comprehension}$ of a set without a cue is calculated in the same way as $CS_{production}$. For a set that contains the cue, the nonlinguistic information, in the form of the cue strength, “str(Cue)”, is added to $CS_{comprehension}$. Then, the listener selects his set of winning rules that allow him to comprehend an integrated meaning with the highest $CS_{comprehension}$.

If the combined strength of the listener’s winning rules exceeds a *confidence threshold*, he adds the perceived meaning-utterance mapping to his buffer and transmits a positive feedback to the speaker. Then, both individuals reward their winning rules by increasing their strengths and association weights, and penalize other competing ones by decreasing their strengths and association weights. Otherwise, a negative feedback is sent and both individuals penalize only their winning rules. For the activated rules that have the initial strengths and association weights, i.e., 0.5, the contribution of linguistic (lexical and syntactic) information to the combined strength is 0.75 (0.5+0.5×0.5). Therefore, the cue strength and confidence threshold are both set to 0.75 to equally treat linguistic and non-linguistic information. This communication scenario simulates a multi-level selection on the lexical, syntactic and nonlinguistic information.

After a number of communications (scaled to the number of individuals), all individuals gradually “forget” their rules by deducting a fixed amount from their strengths and association weights. Rules having negative strengths or association weights to some categories are removed from the individual’s memory and the relevant categories. Categories

having no lexical members are also removed from the memory, together with their syntactic members.

III. THE CULTURAL TRANSMISSION FRAMEWORK AND SIMULATION SETUP

We assume that there is ongoing selective pressure toward shared intentions based on communicative success; during cultural transmissions, individuals who can better understand others in communications have opportunities to produce more offspring who can inherit the levels of their parents' intentionality sharing. Based on this assumption, we propose a cultural transmission framework to test how communicative success affects the level of RC, as in Fig. 3.

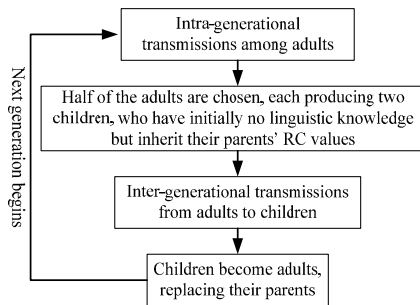


Fig. 3. The cultural transmission framework.

In all simulations, the semantic space contains 4 items as agent constituents, 4 as single-argument predicate, 4 as double-argument predicate, and 4 as patient, in which the constituents being agent are identical to those being patient. These items form 64 integrated meanings, 16 (4×4) of which involve the single-argument predicates and 48 (4×4×(4-1)) involve the double-argument predicates. An individual's communicative success is calculated as the average percentage of integrated meanings in the semantic space that this individual can accurately understand when others speak to him. During the calculation, there is no cue available to the listener and individuals do not adjust their linguistic knowledge. The average value of all individuals' communicative successes is called *understanding rate* (UR), which indicates the understandability of the communal language in the population. The main parameters in the simulations are listed in Table 1.

TABLE I
PARAMETER SETTINGS

| Parameter | Value |
|---|-------|
| Population size | 10 |
| Individual's buffer size | 40 |
| No. utterance exchange per transmission | 20 |
| Rate of random creation of holistic rules | 0.25 |
| Adjusting amount on rule strengths and association weights in competition | 0.1 |
| Deducting amount on rule strengths and association weights in forgetting | 0.01 |
| No. generations | 500 |
| No. intra-generational transmissions | 200 |
| No. inter-generational transmissions | 100 |
| Mutation rate | 0.02 |
| Adjusting amount on RC during mutation | 0.1 |

We set up three sets of simulations. The first two sets evaluate whether intentionality sharing can coevolve with the emergence of an informative language. In these simulations, individuals in the first generation only share 8 holistic rules to encode 8 integrated meanings. In the third set, however, the individuals in the first generation share a compositional language containing 12 word rules to encode those semantic constituents, 3 categories to associate these rules, and 3 syntactic rules (SV, VO and SO) to regulate the order of the lexical members of these categories. All the rule strengths and association weights are 1.0. The second and third sets of simulations compare the optimization based on communicative success under an initial holistic signaling system and an initial compositional language.

Parents are randomly chosen in the first set of simulations, and children directly copy their parents' RC values. However, in the other two sets, parents are chosen based on their communicative success. In each generation of those simulations, after intra-generational communications among adults, each adult's linguistic understandability is calculated and the 5 adults having higher understandabilities are chosen as parents. During the reproduction, the mutation on RC occurs. The mutation values in these simulations are set so that qualitative results can be seen within 500 generations of transmissions.

There are 10 conditions in each set of simulations, determined by the RC values of the individuals in the first generation. These values are chosen randomly from Gaussian distributions. The standard deviations of these distributions are 0.01, but their means range from 0.0 to 1.0 in different conditions, with a step of 0.1. In every condition of the simulations, we conduct 20 runs for statistical analysis.

IV. THE SIMULATION RESULTS

A. With and without Optimization on Intentionality Sharing

In the first set of simulations, there is no adjustment on the initial RC values among individuals. These simulations provide us a baseline of the minimum level of intentionality sharing required to trigger a communal language with a good understandability. In Fig. 4, the solid line traces the average and standard deviation values of the highest UR in simulations with different average values of the initial RC. When the initial RC is low (below 0.4), UR is rather low, around the value triggered by the initially shared holistic rules. Once RC exceeds 0.5, after 500 generations of cultural transmissions, a communal language with a high UR (over 0.8) can emerge in many simulations. These results indicate that given several generations of cultural transmission, a high level of intentionality sharing surely helps trigger a communal language with a good understandability, but the minimum level of this ability to achieve the same goal is not very high. In addition, a small change in the level of this ability greater than 0.3 may cause a qualitative change in the understandability of the communal language. Furthermore, a

check of shared linguistic knowledge among individuals reveals that the emergent communal language with a good understandability consists mainly of word rules, and the local orders in it can form a consistent global order to regulate these word rules. As shown in Fig. 5, this communal language consists of 8 common word rules and a consistent word order SVO, formed by SV and VO. It has a UR of 0.903, which indicates that individuals can use this language to accurately inform each other of many integrated meanings.

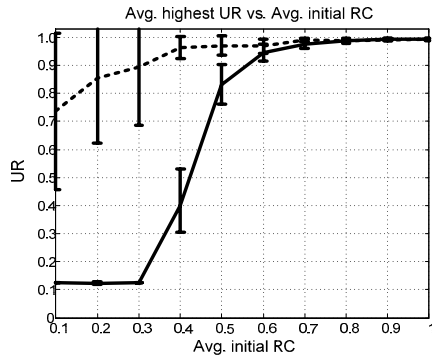


Fig. 4. The average highest UR vs. the average initial RC in the first (the solid line) and second (the dash line) sets of simulations. The distance between the top and bottom bars of a data point is twice of the standard deviation.

```

Generation 499:
UR= 0.903;

Common lexical rules = 8
Rule 1 (SIO) (0.91): "fox" <-> /6 /
Rule 2 (SIO) (0.93): "wolf" <-> /8 /
Rule 3 (SIO) (0.94): "deer" <-> /24 /
Rule 4 (SIO) (0.92): "goat" <-> /5 /
Rule 5 (V) (0.77): "run" <#> <-> /22 16 /
Rule 6 (V) (0.86): "chase" <#> <-> /1 /
Rule 7 (V) (0.85): "stalk" <#> <-> /13 /
Rule 8 (V) (0.86): "attack" <#> <-> /12 /

Local orders
SV (0.619) VS (0.000)
VO (0.578) OV (0.000)
SO (0.056) OS (0.000)

```

Fig. 5. The communal language at the end of 499 generations of cultural transmissions in a simulation of the first set with the initial RC as 0.7. In the common lexical rules, “S”, “O” and “V” denote the syntactic roles of the categories to which these lexical rules belong (the lexical rules encoding constituents with the same semantic roles do not necessarily belong to the same category with the corresponding syntactic roles). Numbers within () are the average strengths of these rules in all individuals, those enclosed by // are the utterance syllables. In the local orders, since the syntactic rules that contain these local orders may belong to different categories with the corresponding syntactic roles, we only record, inside (), the percentages of integrated meanings that individuals can accurately comprehend using the local orders in those syntactic rules.

In the second set of simulations, the levels of intentionality sharing can be adjusted, when children inherit them from adults who have higher values of communicative success in cultural transmissions. In Fig. 4, the dashed line traces the corresponding UR values in the second set of simulations. In these simulations, a much lower initial RC (around 0.2) can already trigger a communal language with a high UR (over 0.8) in some simulations. Once the initial RC exceeds 0.3, a communal language with a high UR emerges in most simulations. In the simulations with rather low initial RC values, the RC values have to be adjusted so that a communal language with a good understandability can emerge after

several generations of cultural transmissions.

B. The Evolution of RC during Language Emergence

The evolution of RC in the second set of simulations is shown in Fig. 6, which traces the differences between the average RC values at the end of 500 generations and those at the first generation. This figure illustrates two roles of communicative success in optimizing the level of intentionality sharing.

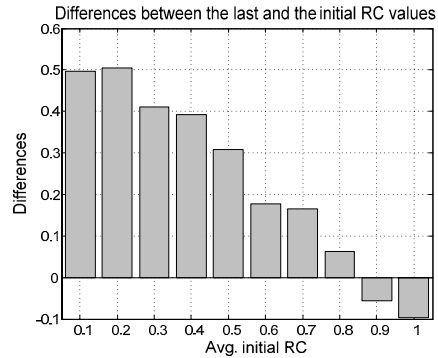


Fig. 6. The differences between the last and the initial RC values in different conditions in the second set of simulations. The average final RC values can be calculated by adding the average initial RC values with the height of the corresponding bars.

On the one hand, if the initial RC values are small (below 0.5), after generations of cultural transmissions, these values can be greatly increased. Intentionality sharing contributes to understanding of idiolects. Individuals having slightly higher levels of this ability than others may have higher values of communicative success. Then, they can be chosen as parents and spread their levels of this ability in the future population. The levels of this ability can be further increased (with 50% chances) slightly as a result of reproduction. Then, an initial, slightly higher level of intentionality sharing can be selected, and increased after several generations. And the average level of this ability in the population increases in respond.

On the other hand, if the initial RC values are already high (within [0.6 0.8]), after several generations, these values are not greatly changed. A high initial RC value (0.9 or 1.0) drops after several generations, which can be seen from the negative values of differences in Fig. 6. In these situations, the levels of intentionality sharing among individuals are similarly high. All individuals can have similar values of communicative success and similar chances to be chosen as parents. Communicative success cannot efficiently select the level of intentionality sharing. Since mutation during the reproduction is neutral, having equal chances to slightly increase or decrease the level of intentionality sharing, after several generations, the average level of intentionality sharing in the population is not greatly different (around 0.1) from that in the first generation. In addition, if the level of this ability is much high, a drop of it may not greatly affect the understanding of other idiolects, especially when a communal language with a good understandability is formed in the population. Then, individuals with a lower level of

intentionality sharing may still be chosen as parents, which will cause the average RC in the population to drop after several generations.

Slightly reducing a rather high level of RC can be viewed as a “side effect” of the optimization. As for the speaker, a drop in RC means that she tends to select events not occurring now and here as the topics in communications; as for the listener, this means that he tends to obtain “wrong” cues having meanings different from the speaker’s intended ones. Once RC drops, during cultural transmissions, there are some communications in which shared intentions are not formed. These communications provide opportunities for individuals to develop some robust linguistic knowledge that needs no assistance of cues or even resists distractions of “wrong” cues. This language can gradually liberate itself from the restriction of nonlinguistic information, and can be used efficiently in communications without cues or other nonlinguistic assistance. Therefore, a moderate level of intentionality sharing can allow language to become an efficient means of communication, independent of other available ones [16]. That communicative success can slightly reduce the level of intentionality sharing is implicit in the short run, but it is crucial for language evolution in the long run.

C. Optimization on Intentionality Sharing based on Compositional Language

In the second set of simulations, the optimization on intentionality sharing takes place during the emergence of a compositional language out of a limited number of holistic signals. But what happens if the initial language is already compositional? Fig. 7 compares the RC differences in the second and third sets of simulations. As shown in this figure, the optimization based on communicative success on intentionality sharing is similar in both sets of simulations.

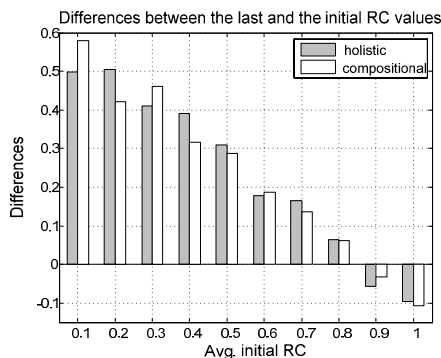


Fig. 7. The differences between the last and the initial RC values in different conditions in the second (the grey bars) and third (the white bars) sets of simulations.

In the third set of simulations, the initial language can accurately exchange all meanings in the semantic space. Individuals in the first generation should have similar values of communicative success and similar chances to be parents. However, the children initially have no linguistic knowledge and have to develop their idiolects from parents during

inter-generational communications. Although they can easily notice the recurrent patterns in sentences created by parents using a compositional language, they have to rely on intentionality sharing to first grasp the meanings contained in those sentences, and then, acquire that compositional language. The level of intentionality sharing in a child can affect his acquisition of linguistic knowledge, and later on, after he becomes an adult, his value of communicative success when others speak to him. Therefore, communicative success in these simulations can play a role similar to that in the second set of simulations on the optimization of intentionality sharing. In addition, the acquisition process described here also occurs in later generations in the second set of simulations, where the communal language may already contain some compositional materials to describe integrated meanings. The simulation results in these two sets indicate that the optimization based on communicative success on intentionality sharing is not dependent on the communal language that evolves during cultural transmissions.

V. DISCUSSIONS AND CONCLUSIONS

This paper develops a computational model that adopts evolutionary mechanisms to explore whether the ability of intentionality sharing in communications can coevolve with language. Following the “socio-biological” explanation, we build in the level of this ability in individual’s genome and use individual’s communicative success to select the level of this ability and some a genetic algorithm to adjust this level in future generations. During cultural transmissions, once the optimization based on communicative success is allowed, along with the emergence of an informative language, an initially low level of intentionality sharing can be increased, whereas a high level of this ability can be reduced.

In the proposed framework, it is worth noticing that it is communicative success that selects the level of intentionality sharing and guides the optimization. But it cannot adjust the level of this ability. It is the mutation mechanism that changes the level of this ability in an individual. Without the guidance of communicative success, the levels of intentionality sharing may undergo a random-walk, since the adjustment during the mutation is neutral. With the guidance of communicative success, the amount in the adjustment can affect the efficiency of the optimization. In general, other kinds of mechanisms can be adopted to adjust the level of intentionality sharing, and the whole framework can be adapted to explore the evolution of other language-related or general cognitive abilities.

Different from the uniqueness view based on comparative studies, our simulation study provides an alternative explanation on the level difference of some language-related abilities between humans and other animals. Once intentionality sharing starts assisting communications, this ability becomes “piggyback” on language, and the evolution of language may cast its influence on this ability in a

phylogenetic timescale. This idea partially reflects the “mimetic” view of language evolution [20], which claims that language, in order to be easily acquired or efficiently used, could adjust some abilities in language users. In addition, intentionality sharing has a reverse role on language evolution, helping trigger a displaced language. This role will lead to the emergence of complex syntax and hierarchical structures for describing complicated or non-existent items and narrating a series of events not necessarily occurring now and here. Furthermore, communicative success is calculated based on cultural transmissions. Its optimization role on intentionality sharing is part of the effects of cultural transmissions. Our study illustrates that cultural transmissions provide an important medium for language evolution, during which not only linguistic features, such as lexical items or simple syntax, can emerge, but also language-related abilities to assist language evolution, such as intentionality sharing, can be adjusted (optimized), even though these abilities are not directly related to linguistic structures. This may explain why the optimization on sharing intentionality continues after a compositional language emerges. As shown in the empirical studies on human infants, an optimized level of this ability still takes effect during the early stage of language acquisition.

As an exploratory study, this model adopts some uncertain or arbitrary assumptions. From the empirical aspect, there is no decisive evidence showing that sharing intentionality is genetically encoded and transmitted. From the theoretical aspect, apart from the “socio-biological” explanation, there is the “socio-cultural” explanation [21], following which, sharing intentionality, as a strategy leading to communicative success, is not necessarily innate; instead, it can be nurtured based on linguistic or nonlinguistic experiences. In addition, our model arbitrarily defines the communicative scenario and linguistic knowledge. Whether the primitive communicative acts among early hominids happen like this and whether there are rule-like behaviors in human brains to process language are still in doubt [22]–[24]. Furthermore, sharing intentionality in this model directly assists linguistic comprehension, but the selective pressure for this is unknown. Nevertheless, this simulation study can assist the discussion of the uniqueness of some language-related behaviors, inspire some empirical studies to further evaluate the details of the coevolutionary hypothesis, and contribute to a complete picture on the evolution of language and its related abilities.

ACKNOWLEDGMENT

The authors are grateful to Prof. Michael Tomasello from Max Planck Institute for Evolutionary Anthropology, and colleagues Susan L. Shuai, Hongying Zheng and Francis Wong from The Chinese University of Hong Kong for the useful comments and discussions on this work.

REFERENCES

- [1] M. H. Christiansen and S. Kirby, “Language evolution: Consensus and controversies,” *Trends in Cognitive Sciences*, vol. 7, pp. 300–307, July 2003.
- [2] K. Oller and U. Griebel, Eds., *Evolution of Communication Systems: A Comparative Approach*. Cambridge, MA: MIT Press, 2000.
- [3] W. S-Y. Wang, “The three scales of diachrony,” in *Explorations in Language*. W. S-Y. Wang, Ed. Taipei; Seattle, WA: Pyramid Press, 1991, pp. 60–71.
- [4] M. Tomasello, “Some facts about primate (including humans) communication and social learning,” in *Simulating the Evolution of Language*. A. Cangelosi and D. Parisi, Eds. London: Springer-Verlag, 2002, pp. 327–340.
- [5] M. H. Christiansen and S. Kirby, Eds. *Language Evolution*. Oxford; New York: Oxford University Press, 2003.
- [6] M. Tomasello, *Origins of Human Communication*. Cambridge, MA: MIT Press, 2008.
- [7] M. Tomasello, M. Carpenter, J. Call, T. Behne, and H. Moll, “Understanding and sharing intentions: The origins of cultural cognition,” *Behavioral and Brain Sciences*, vol. 28, pp. 675–691, October 2005.
- [8] J. R. Hurford, “Biological evolution of the Saussurean sign as a component of the Language Acquisition Device,” *Lingua*, vol. 77, pp. 187–222, February 1989.
- [9] A. Cangelosi and D. Parisi, “The emergence of a ‘language’ in an evolving population of neural networks,” *Connection Science*, vol. 10(2): 83–97, June 1998.
- [10] K. Okanoya, “Sexual display as a syntactic vehicle: The evolution of syntax in birdsongs and human language through sexual selection,” in *The Transition to Language*, A. Wray, Ed. Oxford, UK: Oxford University Press, 2002, pp. 47–63.
- [11] G. Rizzolatti, L. Fadiga, V. Gallese, and L. Fogassi, “Premotor cortex and the recognition of motor actions,” *Cognitive Brain Research*, vol. 3, pp. 131–141, March 1996.
- [12] M. A. Arbib, Ed. *Action to Language via the Mirror Neuron System*. Cambridge, MA: Cambridge University Press, 2006.
- [13] K. L. Chow, “Speech and language: A human trait defined by molecular genetics,” in *Language Acquisition, Change and Emergence: Essays in Evolutionary Linguistics*, J. W. Minett and W. S-Y. Wang, Eds. Hong Kong: City University of Hong Kong Press, 2005, pp. 21–45.
- [14] A. Cangelosi and D. Parisi, Eds. *Simulating the Evolution of Language*. London: Springer-Verlag, 2002.
- [15] J. N. Thompson, *The Coevolutionary Process*. Chicago, IL: University of Chicago Press, 1994.
- [16] C. F. Hockett, “The origin of speech,” *Scientific American*, vol. 203, pp. 88–96, 1960.
- [17] T. Gong, *Language Evolution from a Simulation Perspective: On the Coevolution of Compositionality and Regularity*. Doctoral Dissertation. Hong Kong: The Chinese University of Hong Kong, 2007.
- [18] T. Gong, *Computational Simulation in Evolutionary Linguistics: A Study on Language Emergence* (Frontiers in Linguistics, Monograph IV). Taipei: Institute of Linguistics, Academia Sinica, 2008 in press.
- [19] M. Tomasello, *Constructing a Language: A Usage-based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press, 2003.
- [20] S. Blackmore, *The Meme Machine*. Oxford: Oxford University Press, 1999.
- [21] L. Steels, “The emergence and evolution of linguistic structure: From lexical to grammatical communication systems,” *Connection Science*, vol. 17, pp. 213–230, September–December 2005.
- [22] M. Tallerman, “Did our ancestors speak a holistic protolanguage?” *Lingua*, vol. 117, pp. 579–604, March 2007.
- [23] K. Smith, “Is a holistic protolanguage a plausible precursor to language? A test case for a modern evolutionary linguistics,” *Interaction Studies*, vol. 9, pp. 1–17, March 2008.
- [24] J. Elman, “An alternative view of the mental lexicon,” *Trends in Cognitive Sciences*, vol. 8, pp. 301–306, July 2004.