
Exploring Language Evolution using Behavioral Computational Models

Tao GONG

Department of Linguistics
Max Planck Institute for Evolutionary Linguistics

email: tao_gong@eva.mpg.de

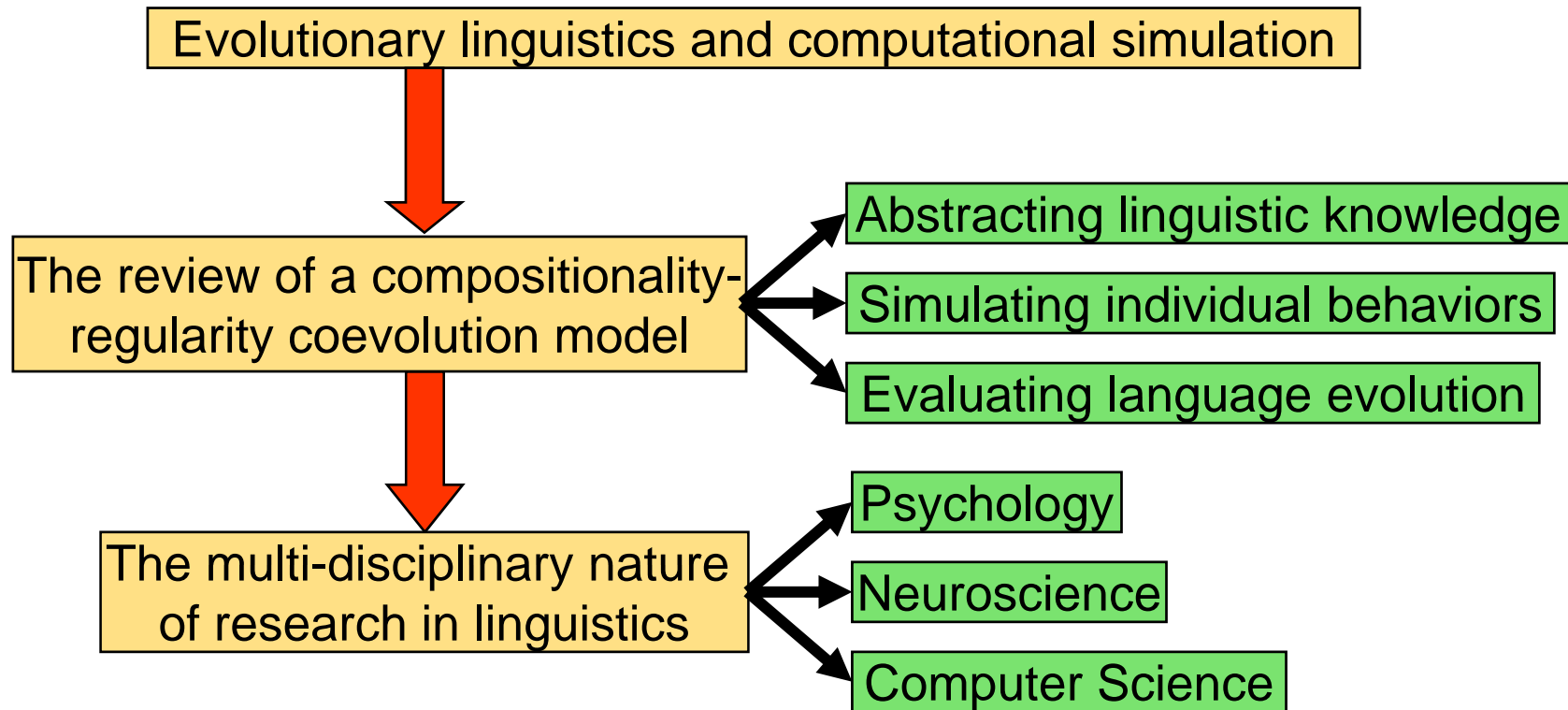
homepage: <http://www.eva.mpg.de/lingua/staff/gong/home.php>

Oct. 9, 2008





Outline



*My focus is **how computational simulation, as an efficient methodology, assists research on language evolution***

The emergence of linguistic universals

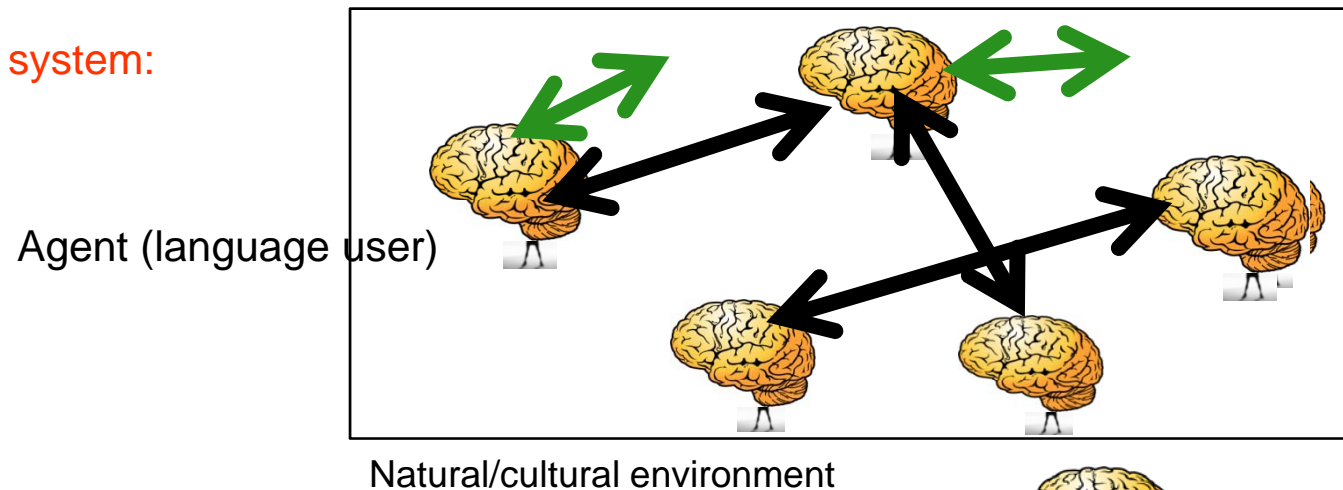
- **Linguistic universals:** language structure and use that hold across most but not necessarily all languages of the world (Christiansen & Kirby 2003).
 - Examples: arbitrariness; word order bias; recursion;
- **Compositionality:** the meanings of complex expressions are determined by the meanings of their components (lexical items).
 - English: /cats eat rats/;
 - Chinese: /猫吃老鼠/;
 - French: /un chat mange un rat/;
 - Japanese: /猫 が ネズミ を 食べる/;
Cat (nominative) rat (accusative) eat (canonical)
- **Regularity at the syntactic level:** many languages adopt conventionalized structures (e.g., **word order** or morphology) to build up complex expressions.
 - English: */cat rat eats/; /cat eats rat/ and /rat eats cat/;

Christiansen, M. H. & Kirby, S. 2003. Language evolution: Consensus and controversies. *Trends in Cognitive Sciences*, 7(7): 300–307.

Behavioral models

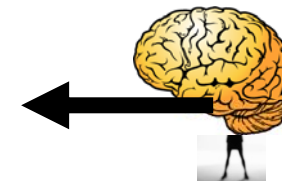
- Simulate **individuals' language-related behaviors** and/or the **social-cultural environment of language users**
- Explore **the minimum prerequisites** for the emergence of certain linguistic universals, or evaluate the effect of various **linguistic or nonlinguistic factors** on language evolution.

- **Multi-agent system:**



- **Rule-based system:**

“cat” \leftrightarrow /cat/; “run” \leftrightarrow /run/;
/cat/ \ll /run/



- Study the emergent communal language triggered by individual linguistic behaviors in linguistic communications



The general procedure to construct a behavioral model to study language evolution

- 1) Set up the working hypotheses;
 - 2) Set up the conceptual framework;
 - 3) Define the artificial language;
 - 4) Define the linguistic knowledge;
 - 5) Define the acquisition mechanisms;
 - 6) Define the communication scenarios;
 - 7) Define the indices to evaluate the simulation results;
 - 8) Analyze the simulation results;
 - 9) Verify the working hypotheses;
- *I will illustrate this procedure using a compositionality-regularity coevolution model*



The compositionality-regularity coevolution model

1) Set up the working hypotheses

- Language origin is a process of **coevolution of compositionality and regularity**: the acquisition and conventionalization of these universals **coevolve** during the transition from a holistic signalling system to a compositional language.
- The syntactic development follows a **“bottom-up” scenario**: the global word order at the sentence level can be formed based on reiterating local orders between lexical items.
- **Some domain-general abilities**, such as **pattern extraction** (Fillmore 1979; Monchi et al. 2001) and **sequential learning** (Christiansen & Ellefson 2002), could be precursors for the linguistic abilities to handle compositionality and regularity.

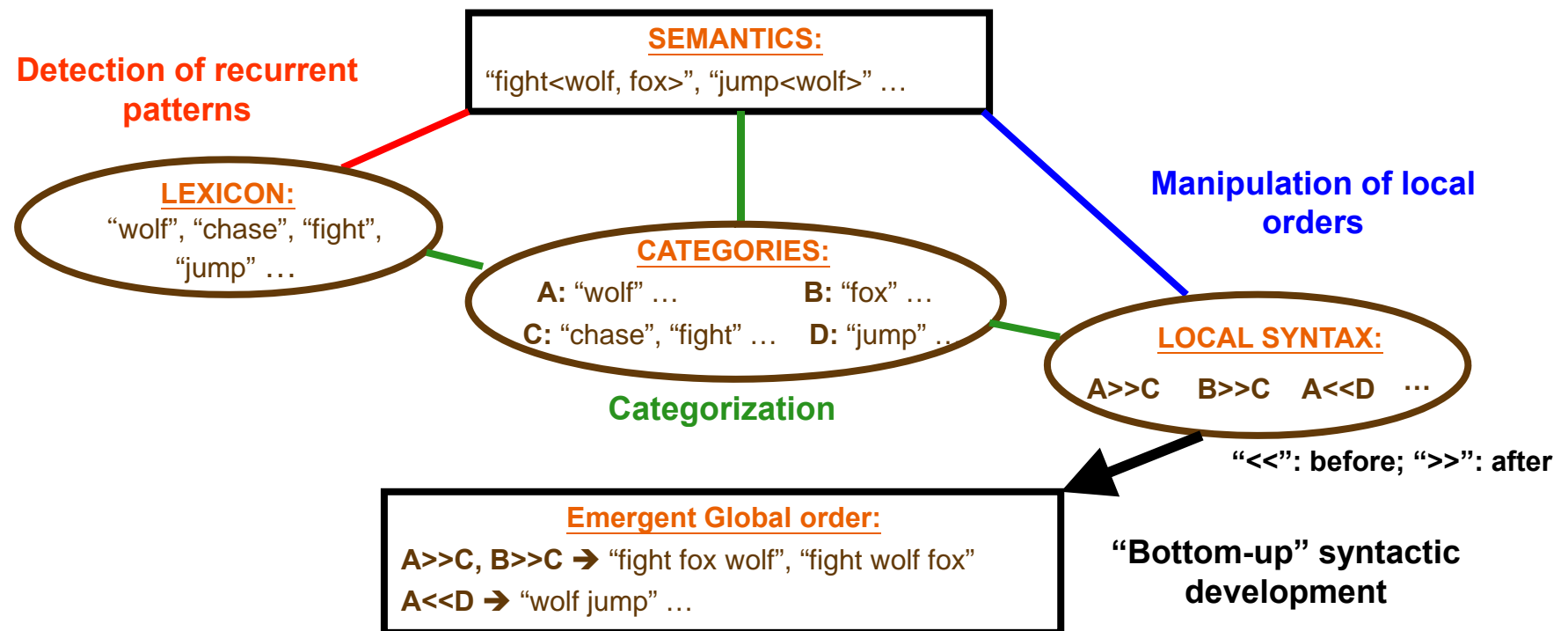
Fillmore, L. W. 1979. Individual differences in second language acquisition. In: C. J. Fillmore, D. Kempler, and W. S-Y. Wang, eds., *Individual differences in language ability and language behavior*, New York: Academic Press, 203–228.

Monchi, O., Petrides, P., Petre, V., Worsley, K., & Dagher, A. 2001. Wisconsin card sorting revisited: Distinct neural circuits participating in different stages of the task identified by event-related functional magnetic resonance imaging. *Journal of Neuroscience*, 21(19): 7733–7741.

Christiansen, M. H. & Ellefson, M. R. 2002. Linguistic adaptation without linguistic constraints: The role of sequential learning in language evolution. In: A. Wray, ed., *The transition to language*, Oxford, UK: Oxford University Press, 335–358.

2) Set up the conceptual framework

- Two types of integrated meanings:
 - Type1: “Pr₁<Ag>”: e.g., “hop<deer>”;
 - Type2: “Pr₂<Ag, Pat>”: e.g., “chase<fox, wolf>”;



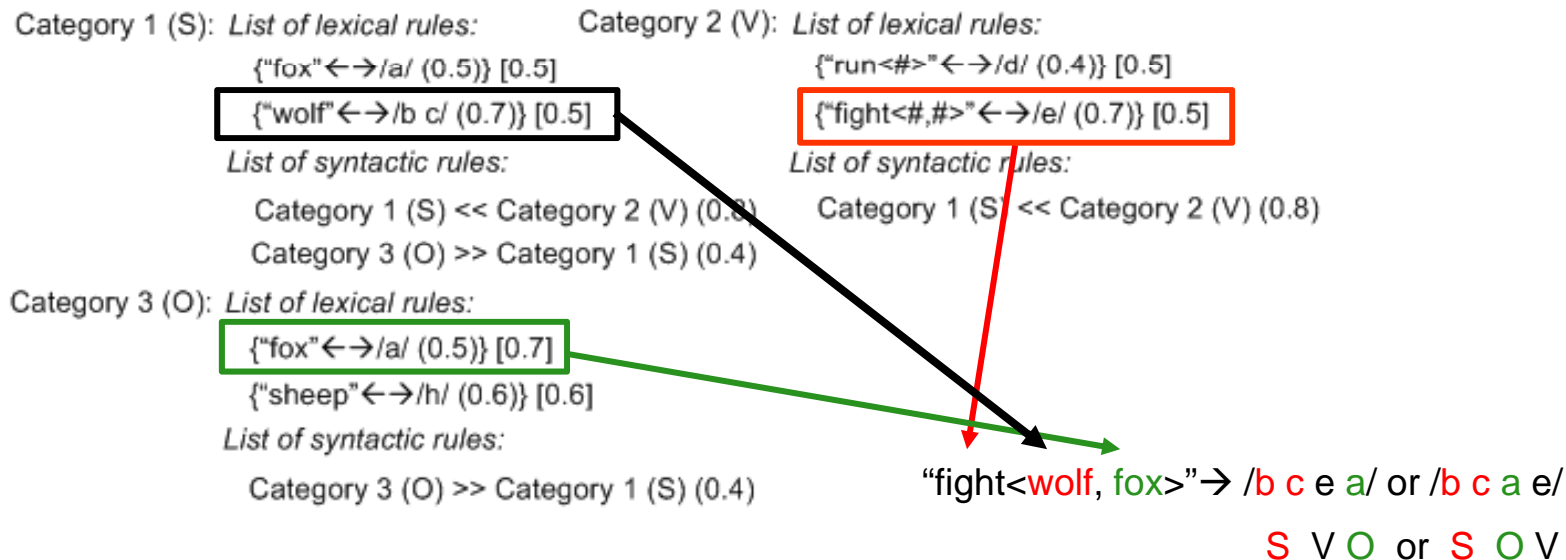
Local syntax: binary sequential relation (before or after) between 2 lexical items;

Global word order (e.g., SVO, SOV) results from local syntax; S: subject; V: verb; O, object;

4) Define the linguistic knowledge (cont.)

- **Syntactic categories:**
 - A group of lexical rules that express semantic items with the same semantic role, together with some syntactic rules to regulate the orders between these lexical rules and those from other categories;
 - **Syntactic role (S, O or V)**, a list of lexical rules and a list of syntactic rules;
 - **[Association weight]**: the membership value for a lexical rule to be associated into a category;

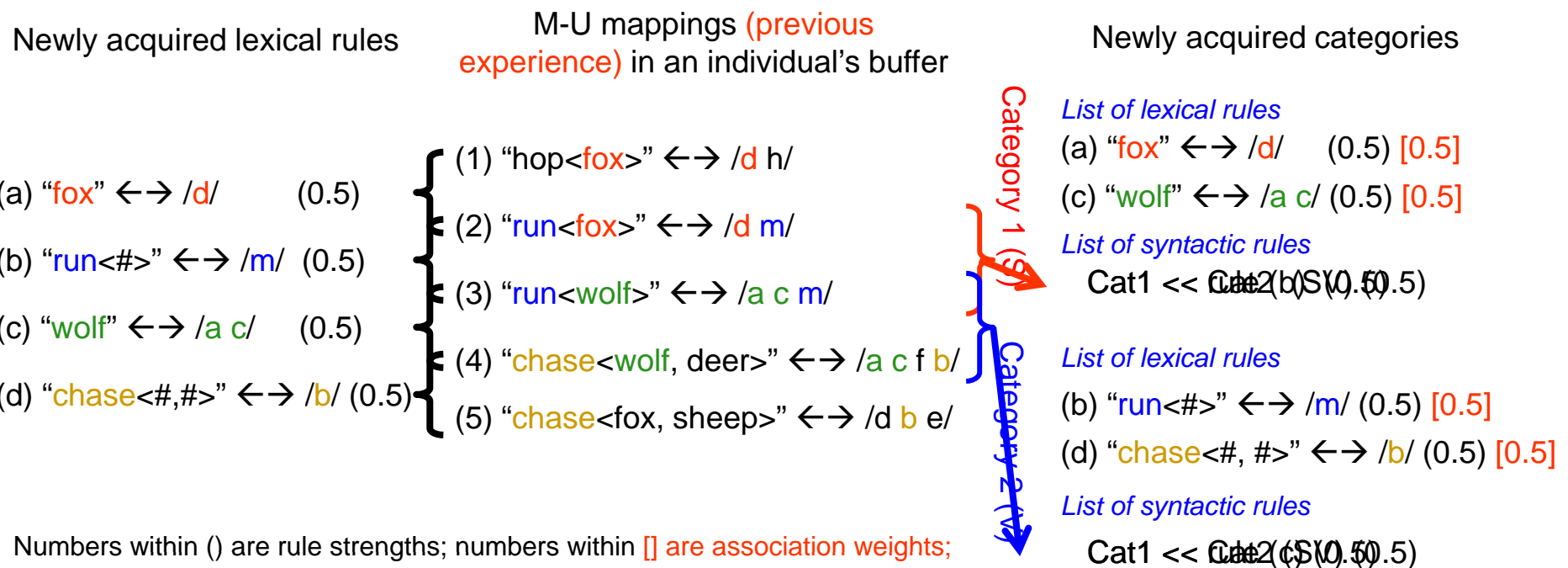
Syntactic categories



Numbers inside () are rule strengths; numbers inside [] are association weights;

5) Define the acquisition mechanisms

- **Random creation** of linguistic instances (holistic expressions): without it, language could never get off the ground.
- **“Learning from experience”**:
 - Acquisition of lexical rules: **detection of recurrent patterns**;
 - Acquisition of local orders and syntactic categories:
 - Lexical rules **with the same semantic role (Ag, Pat or Pr_{1/2})** and **similarly used (local order)** in M-U mappings can be associated into categories;



5) Define the acquisition mechanisms (cont.)

■ Category merge:

- Two categories having the same syntactic roles share some lexical member(s);
- The linguistic instances in the buffer indicate that some of lexical members in these two categories are similarly used (with the same local order) with respect to another lexical rule;
- The merged category comprises all lexical and syntactic members of the original categories;

M-U mappings in the buffer

- (1) "fight<fox, wolf>" \leftrightarrow /d a c b/
- (2) "fight<deer, gazelle>" \leftrightarrow /k b n/

Acquired lexical rule

"fight<# , #>" \leftrightarrow /b/ (0.5)

Available categories

Category 1 (S): List of lexical rules:

{(a) "fox" \leftrightarrow /d/ (0.6)} [0.5]

{(b) "wolf" \leftrightarrow /a c/ (0.6)} [0.7]

List of syntactic rules:

(I) Category 1 (S) << Category 3 (O) (0.6)

Category 2 (S): List of lexical rules:

{(b) "wolf" \leftrightarrow /a c/ (0.6)} [0.5]

{(c) "deer" \leftrightarrow /k/ (0.7)} [0.8]

List of syntactic rules:

(II) Category 2 (S) >> Category 4 (V) (0.7)



Merged category

Category 1 (S): List of lexical rules:

{(a) "fox" \leftrightarrow /d/ (0.6)} [0.5]

{(b) "wolf" \leftrightarrow /a c/ (0.6)} [0.7]

{(c) "deer" \leftrightarrow /k/ (0.7)} [0.8]

List of syntactic rules:

(I) Category 1 (S) << Category 3 (O) (0.6)

(II) Category 1 (S) >> Category 4 (V) (0.7)

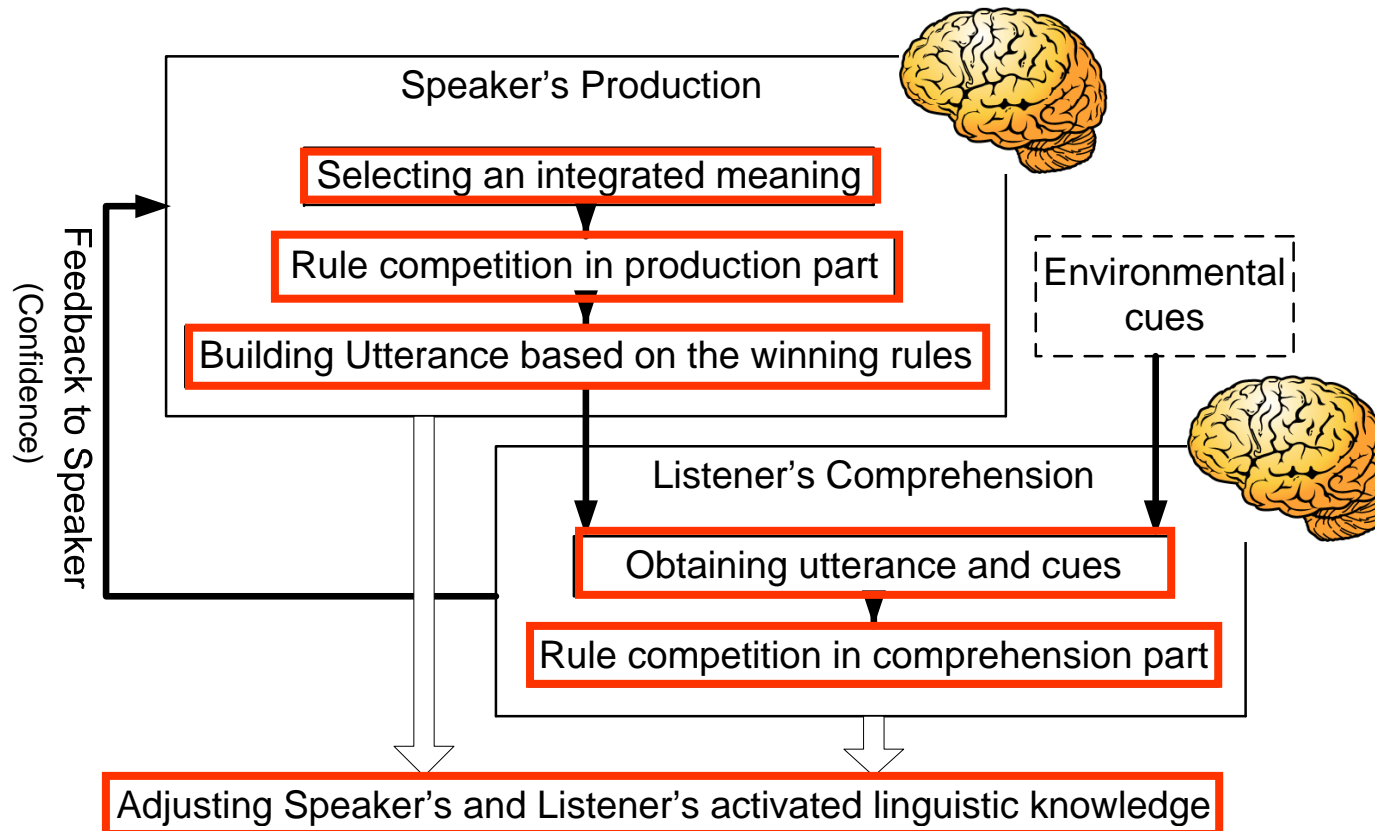
(III) Category 1 (S) << Lexical rule (d) (0.5)

Numbers within () are rule strengths; numbers within [] are association weights;

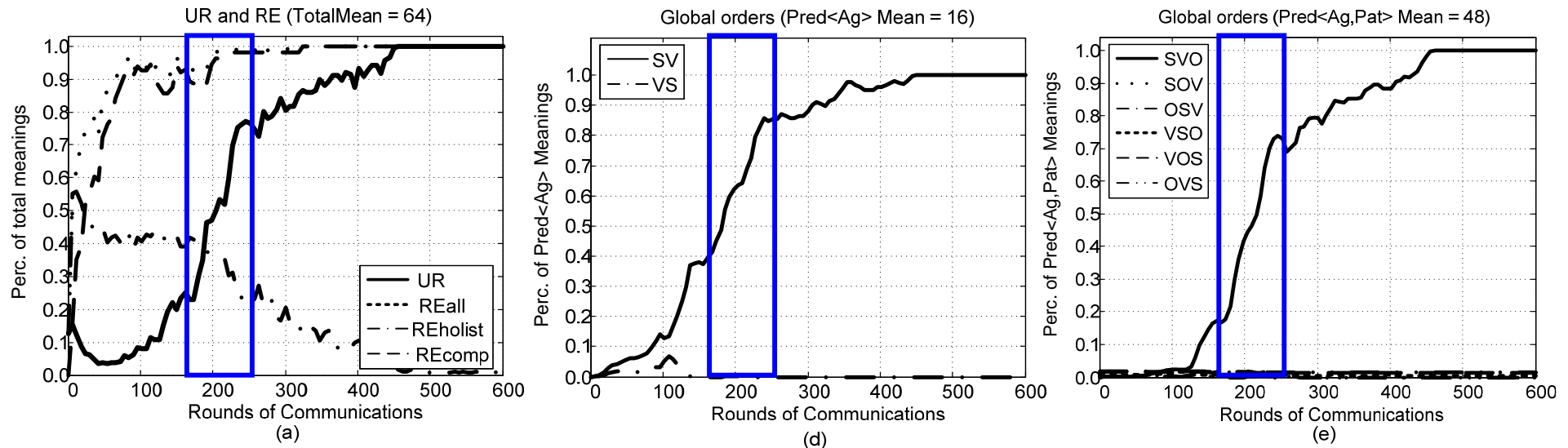
What we simulated here are mechanisms reported in empirical studies (e.g., the "Verb-Island" hypothesis, Tomasello, 2003)!

6) Define the communication scenario

- Nonlinguistic information (**environmental cues**) may assist primitive language communications.
- **Cues are unreliable**, which may or may not contain the speaker's intended meaning (the probability with which the speaker's intended meaning is contained in one of the cues available to the listener is controlled by **Reliability of Cue**);



8) Analyze the simulation results



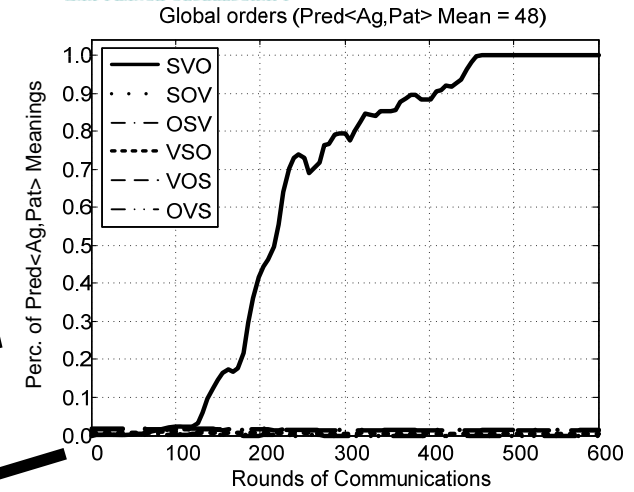
- **Emergence of compositionality:**
 - **RE** → Holistic signals → compositional expressions;
 - **UR** → The emergence of a common compositional language (**the S-shape sharp transition**).
- **Emergence of regularity:** some global word orders become **prevalent**;
 - **Prevalent order:** the global / local orders that are frequently used in comprehensions;
 - e.g., SV for “Pr₁<Ag>” meanings; SVO for “Pr₂<Ag, Pat>” meanings;
- **Coevolution of compositionality and regularity:** the emergence of common lexical items and that of prevalent global orders are almost synchronized; they coevolve during the transition from a holistic signaling system to a compositional language.



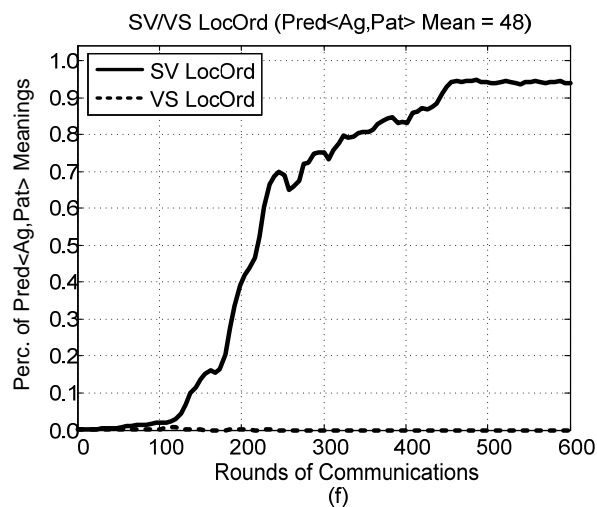
8) Analyze the simulation results (cont.)

Global order's understandability

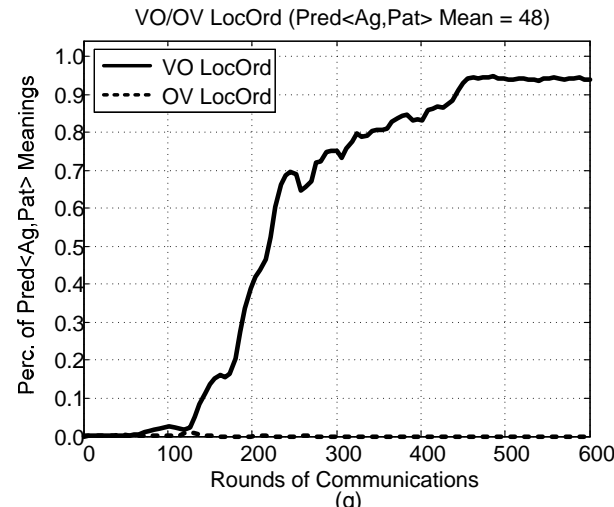
Local order's understandability



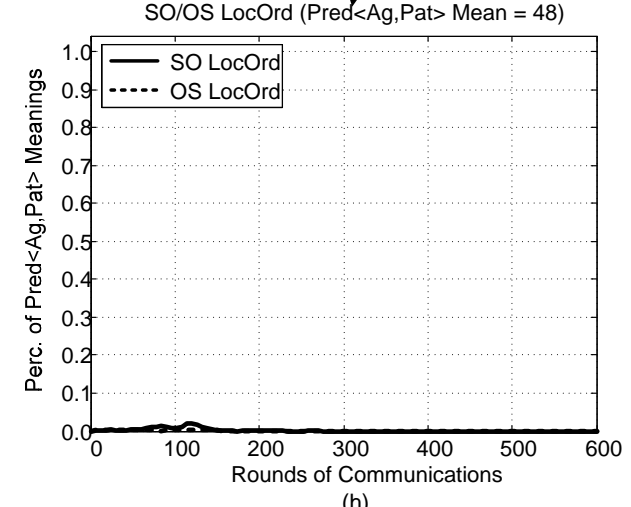
(e)



(f)



(g)



(h)

- The “bottom-up” syntactic development: the prevalent global word order reflects the local sequential information specified by the prevalent local orders;

9) Verify the working hypotheses

- **The working hypotheses:**
 - ✓ □ Language origin is a process of **coevolution of compositionality and regularity**.
 - The close connections among linguistic components during language evolution.
 - ✓ □ The syntactic development follows a **“bottom-up” approach**.
 - Reconsideration of the innateness of syntax as in Innatism: whether syntax is determined once for all, or it can be gradually developed following a constructivist manner.
 - ✓ □ **Some domain-general abilities**, such as **pattern extraction** and **sequential learning**, can be precursors for the linguistic abilities to handle compositionality and regularity.
 - The close connections between linguistic and other cognitive abilities;
 - Reconsideration of the domain-specificity of linguistic competences;

- **Extensions of the current model on other topics in evolutionary linguistics:**
 - The effect of different linguistic components on language evolution (Minett et al. 2006);
 - The role of cultural transmission (Gong et al. 2007) and social factors on language evolution (Gong et al. 2008);

Minett, J. W., Gong, T., & Wang, W. S-Y. (2006). A language emergence model predicts word order bias. In: A. Cangelosi, A. D. M. Smith, & K. Smith, eds., *The Evolution of Language: Proceedings of the 6th International Conference (Evolang6)*, Singapore: World Scientific, 206–213.

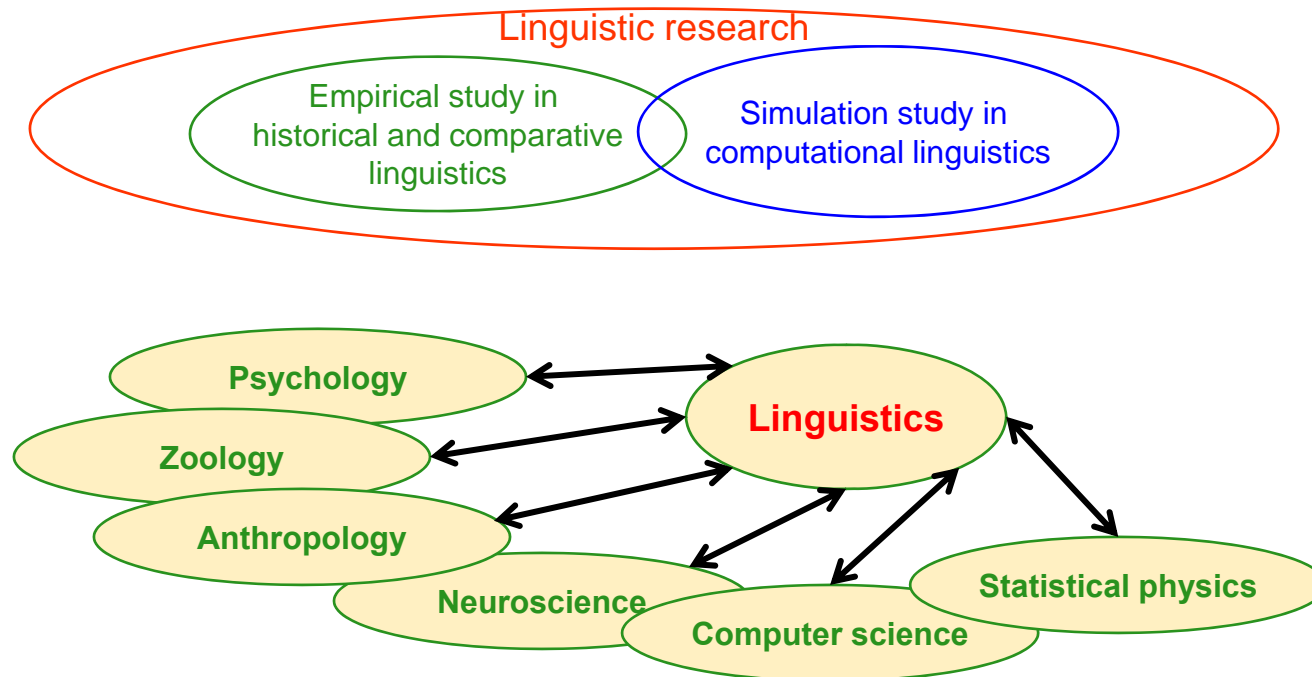
Gong, T., Minett, J. W., & Wang, W. S-Y. (2007). A Simulative Study of the Roles of Cultural Transmission in Language Evolution. *Proceedings of IEEE Congress on Evolutionary Computation 2007*, Singapore, pp. 843–850.

Gong, T., Minett, J. W., & Wang, W. S-Y. (2008). Exploring social structure effect on language evolution based on a computational model. *Connection Science*, 20(2-3): 135-153.



The multi-disciplinary nature of research in linguistics

- Provide language data;
- Inspire and verify linguistic theories;
- Present empirical bases for simulated behaviors and test simulation results;
- Verify the available theories and empirical findings;
- Inspire further empirical studies and new theories;





MAX-PLANCK-GESELLSCHAFT

Max Planck Institute
for Evolutionary Anthropology
Department of Linguistics

Danke schön!

Only humans have language

My work!



Have human languages gradually
evolved from domain-general abilities?

Domain-general abilities are shared among many animals