
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>

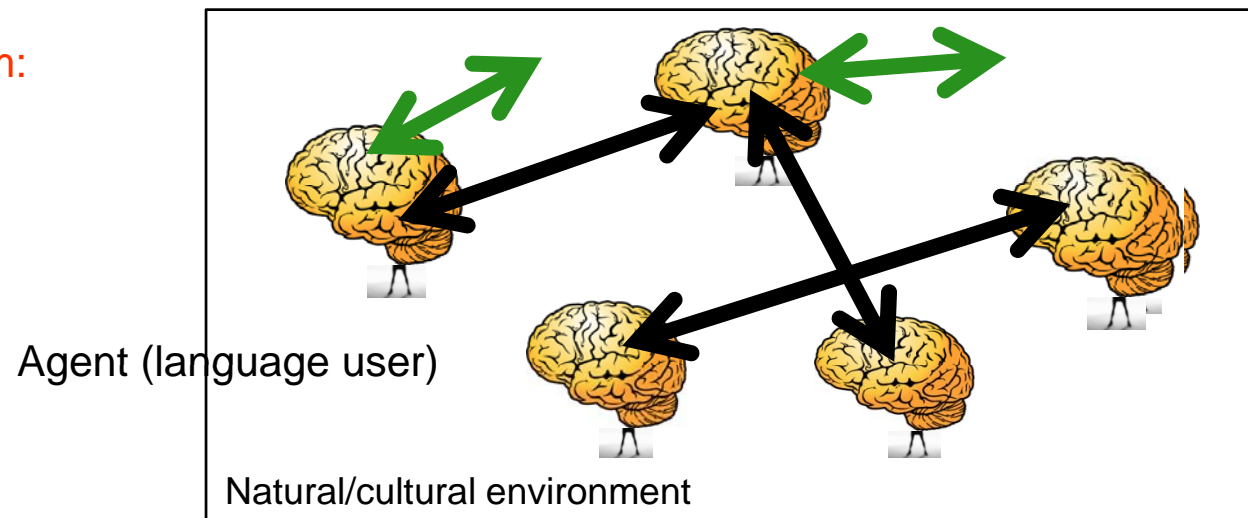
Oct. 29, 2008



Behavioral models

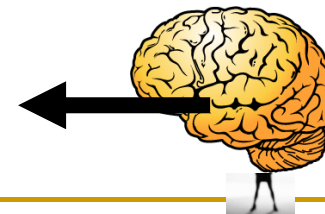
- Simulate both **synchronic development of idiolects** and **diachronic evolution of communal language**, as well as the **social-cultural environment of language users**
- Explore the minimum prerequisites (certain language-related behaviors) for certain linguistic universals (compositionality or regularity), and evaluate the effect of linguistic or non-linguistic factors on language evolution.

- **Multi-agent system:**



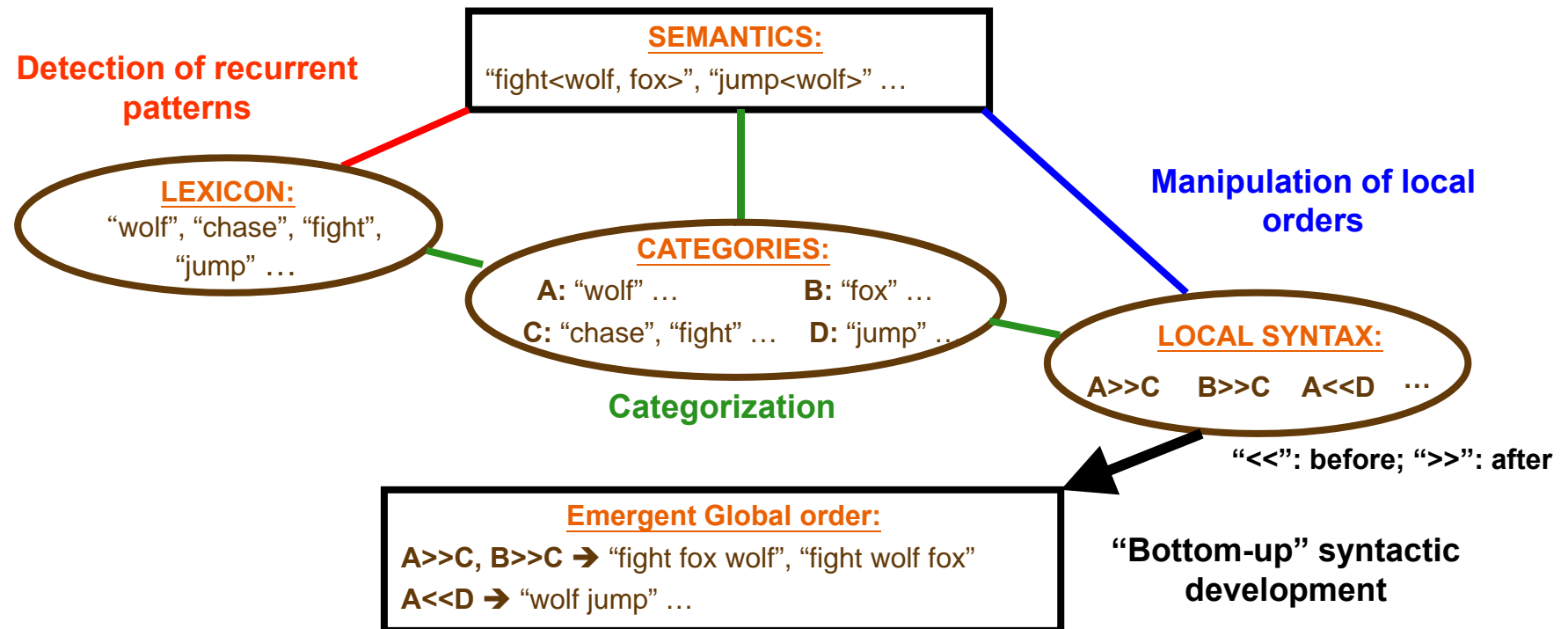
- **Rule-based system:**

“cat” ↔ /cat/; “run” ↔ /run/;
/cat/ << /run/



1) Set up the working hypotheses and the conceptual framework

- The conceptual framework: Alison Wray’s holistic protolanguage hypothesis;
- 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;



2) Define the artificial language and the linguistic knowledge

- Language is defined as a set of **M(eaning)-U(tterance) mappings**:
 - e.g., “fight<wolf, fox>” \leftrightarrow /a b c/;
 - The meaning part**: an integrated meaning or some semantic item(s);
 - The utterance part**: syllable(s) chosen from a syllable space;
 - The mapping is **bidirectional**: in production, to express “fight<wolf, fox>”, the speaker may use /a b c/; in comprehension, when hearing /a b c/, the listener may interpret it as “fight<wolf, fox>”;
- The linguistic knowledge is represented by different types of linguistic rules:

- Linguistic rules**: M-U mapping or local syntax + strength;
 - Lexical rules**: how to map semantic components to utterance syllables;
 - Syntactic rules**: local orders (BEFORE “<<“, AFTER “>>“, SURROUND “▼” and BETWEEN “▲”) to regulate syllables of two lexical rules.
 - A **rule strength** denotes the probability (within [0.0 1.0]) of successfully applying its M-U mapping or local order;

Lexical rules

Holistic rules:

- (a) “chase<wolf, bear>” \leftrightarrow /a b/ (0.5)
- (b) “hop<deer>” \leftrightarrow /c/ (0.4)
- (c) “hop<deer>” \leftrightarrow /d e/ (0.6)

Compositional rules:

- (d) “wolf” \leftrightarrow /f/ (0.6)
- (e) “run<#>” \leftrightarrow /c/ (0.7)
- (f) “chase<#, bear>” \leftrightarrow /e f * g/ (0.7)

Syntactic rules

- (1) {“chase<#, bear” \leftrightarrow /e f * g/ (0.7)} ▼ {“fox” \leftrightarrow /a/ (0.5)} (0.7)
- (2) {“wolf” \leftrightarrow /f/ (0.6)} ▲ {“chase<#, bear” \leftrightarrow /e f * g/ (0.7)} (0.5)
- (3) Category 1 (S) << Category 2 (V) (0.8)
- (4) Category 3 (O) >> Category 2 (V) (0.4)

“#” is unspecified semantic item; “*” is unspecified utterance syllable(s);

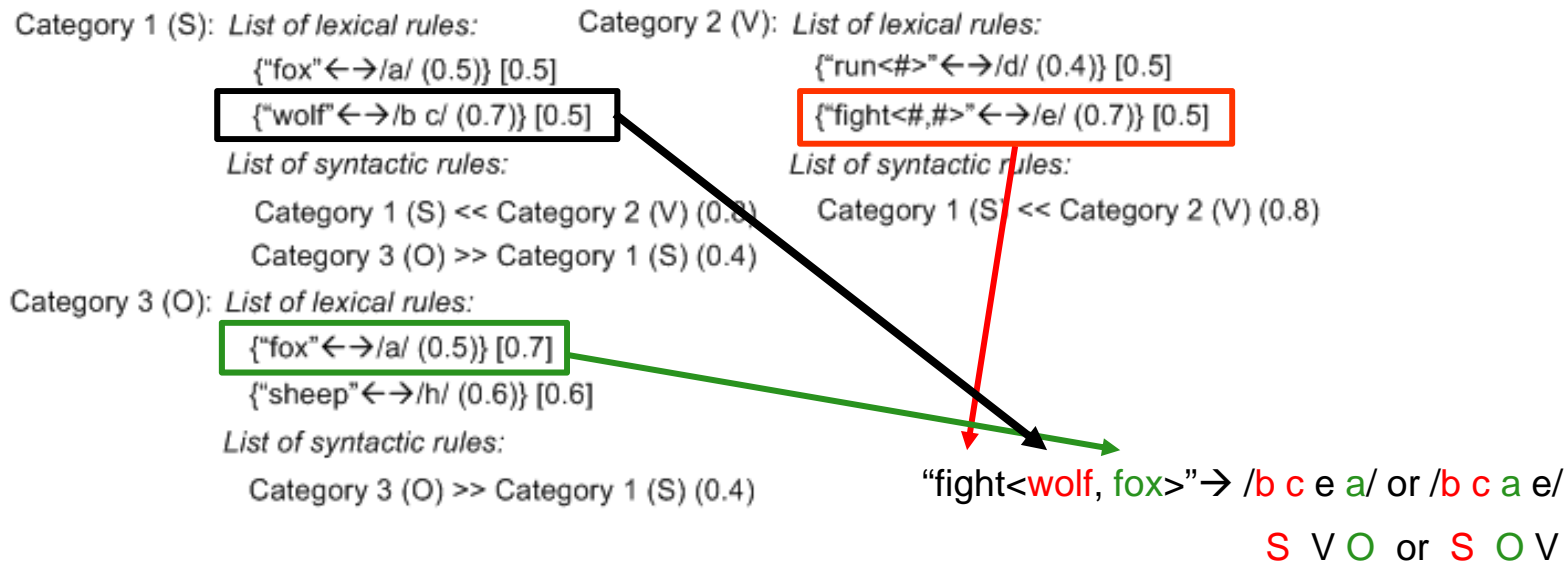


2) Define the artificial language and the linguistic knowledge (cont.)

Categories:

- Inspired from the “Verb-islands” hypothesis (Tomasello, 2003);
- Syntactic role (S, O or V), a list of lexical rules and a list of syntactic rules;
- 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 categories

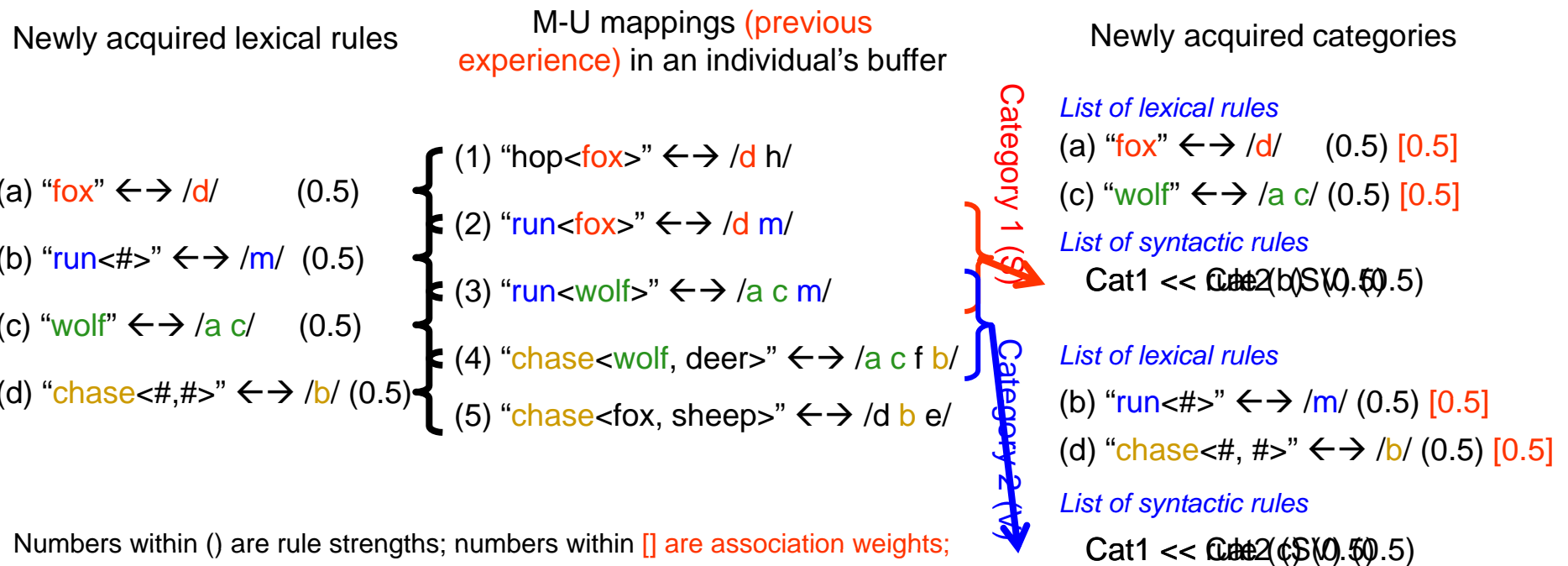


Numbers inside () are rule strengths; numbers inside [] are association weights (the membership value for a lexical rule to be associated into a category)



3) Define the acquisition mechanisms and the communication scenario

- **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)

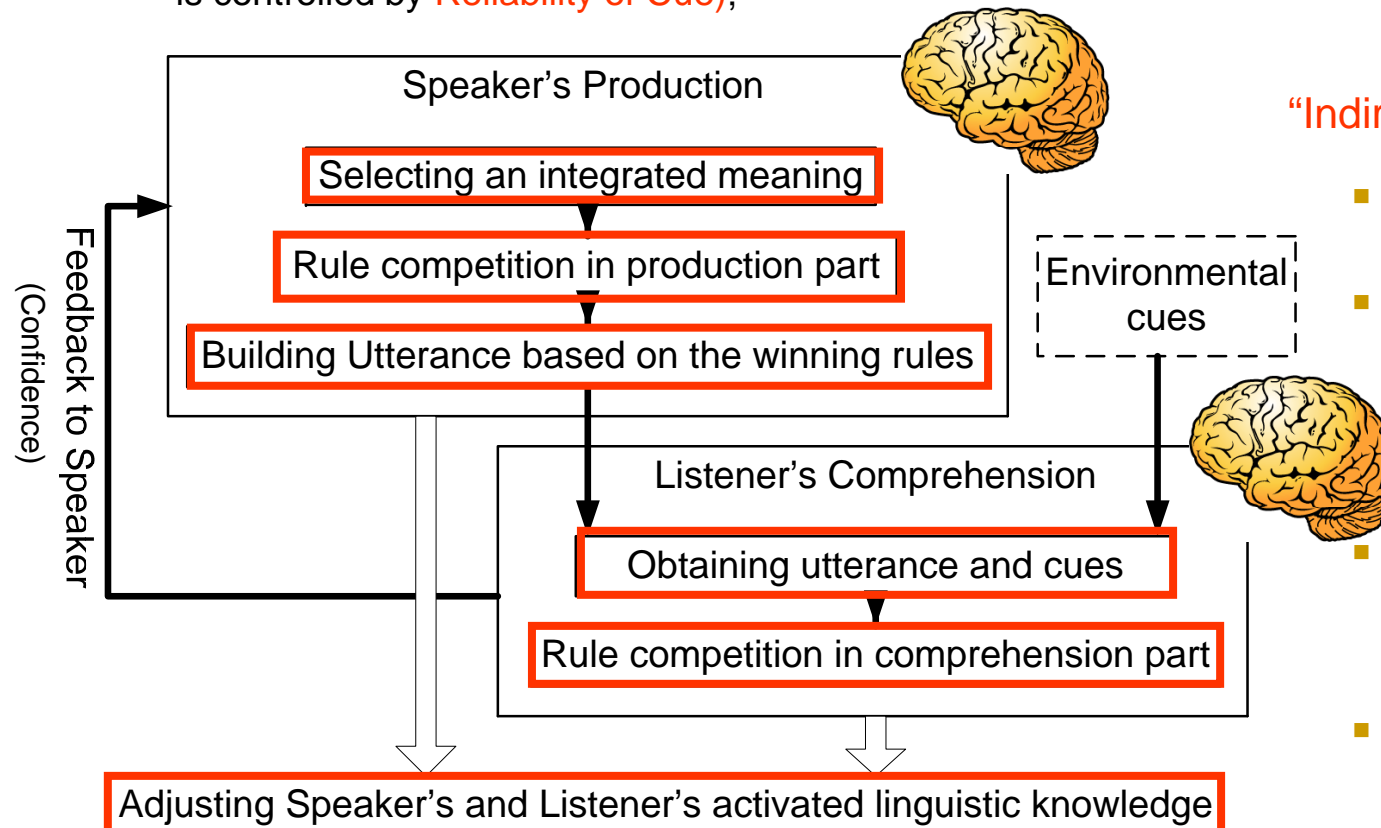
(III) Category 1 (S) << Lexical rule (d) (0.5)

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)!

3) Define the acquisition mechanism and 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**);



"Indirect meaning transference"

- Independent linguistic knowledge;
- No telepathic connection between individuals' minds: no direct check between the speaker's intended meaning and the listener's comprehended one;
- Confidence feedback based on the combined strength of the listener's winning rules;
- Both linguistic and nonlinguistic information take part in comprehension.

If the feedback is positive, acquisition of linguistic knowledge in listener



4) Define the indices to evaluate the simulation results

For compositionality:

- **Rule expressivity (RE):** $RE = \frac{\sum_i \text{No. integrated meanings that agent } i \text{ can express}}{\text{No. agents} \times \text{No. integrated meanings}}$
- **Understanding rate (UR):** $UR = \frac{\sum_{i,j} \text{No. understandable integrated meanings between agents } i \text{ and } j}{\text{No. pairs of } i, j \times \text{No. integrated meanings}}$

For regularity:

- **Global word orders' understandability ($UR_{Global(i)}$, SVO, SOV, OSV, VSO, VOS, OVS):**

$$UR_{GloOrder(i)} = \frac{\sum_{j,k} \text{No. Type2 meanings understandable using Global order } i \text{ between agents } j \text{ and } k}{\text{No. pairs of } j, k \times \text{No. integrated meanings}}$$

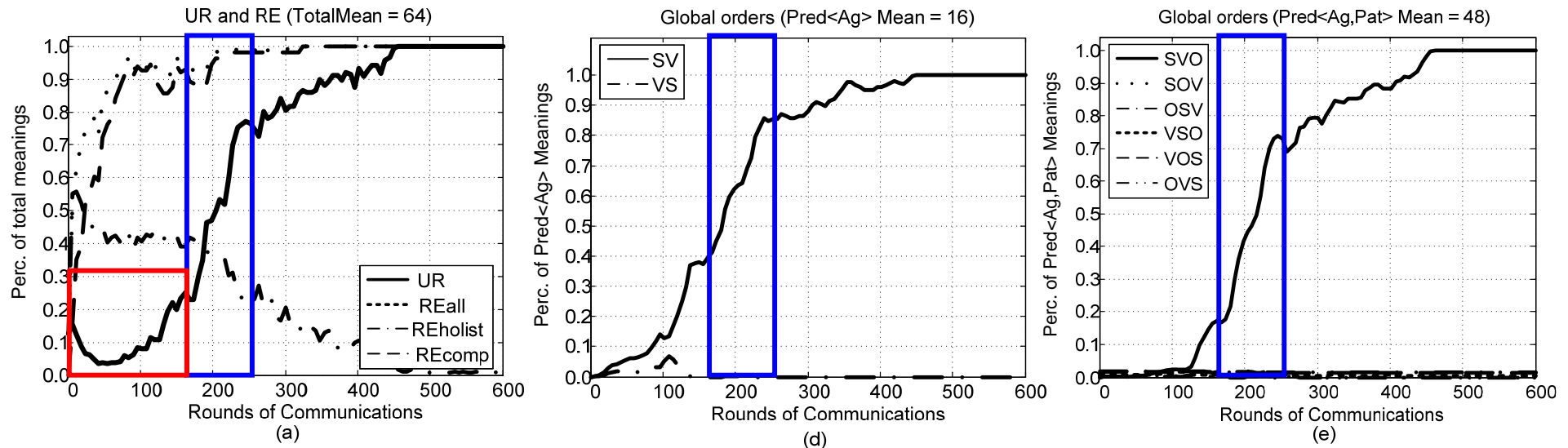
- **Local order's understandability ($UR_{Local(i)}$, SV, VS, VO, OV, SO, OS):**

$$UR_{LocOrder(i)} = \frac{\sum_{j,k} \text{No. Type2 meanings understandable using Local order } i \text{ between agents } j \text{ and } k}{\text{No. pairs of } j, k \times \text{No. integrated meanings}}$$

■ Simulation setup

- No. agents = 10, No. communications = 6000 (600 rounds). No. utterance exchange = 20.
- **Semantic space:** 4 "Ag", 4 "Pr₁", 4 "Pr₂", 4 "Pat" (= "Ag"); 16 "Pr₁<Ag>" and 48 "Pr₂<Ag, Pat>";
- **Initial holistic signaling system:** limited holistic rules (8) are shared;
- Reliability of cue = 0.6; Buffer size = 40;

5) Analyze the simulation results and verify the working hypotheses



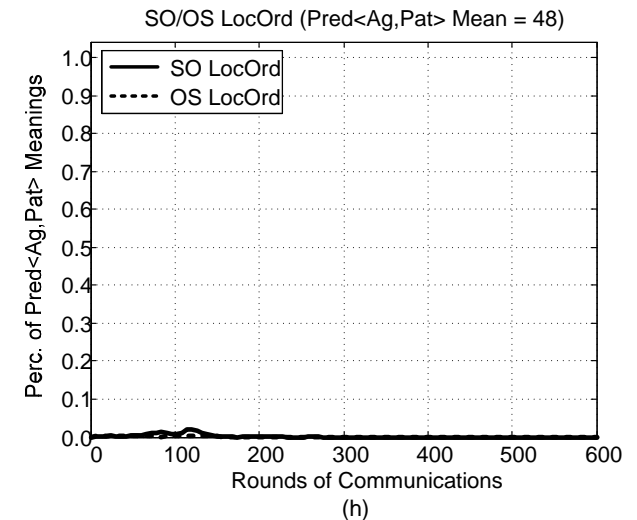
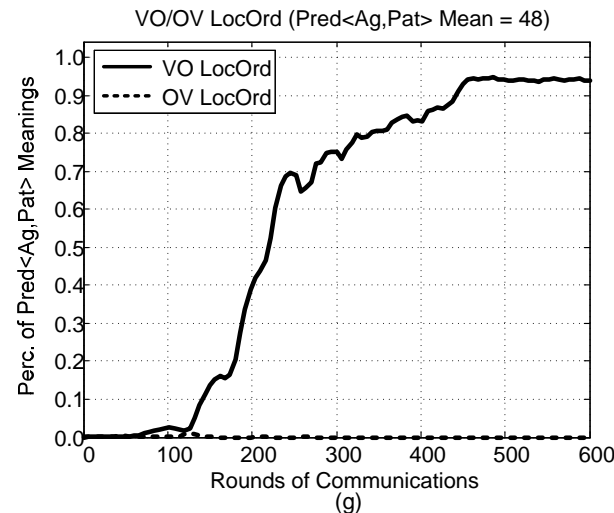
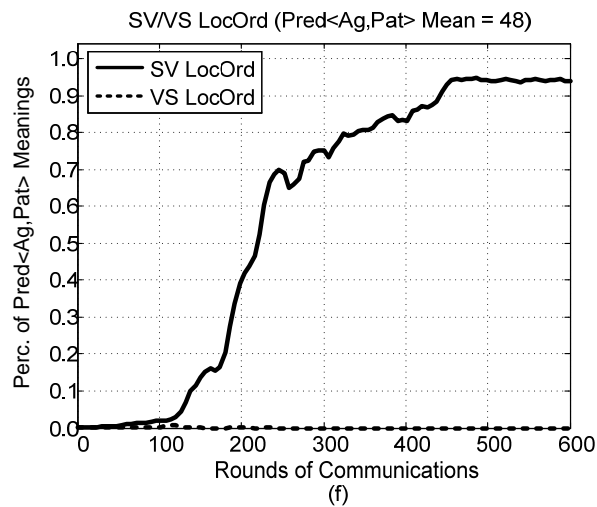
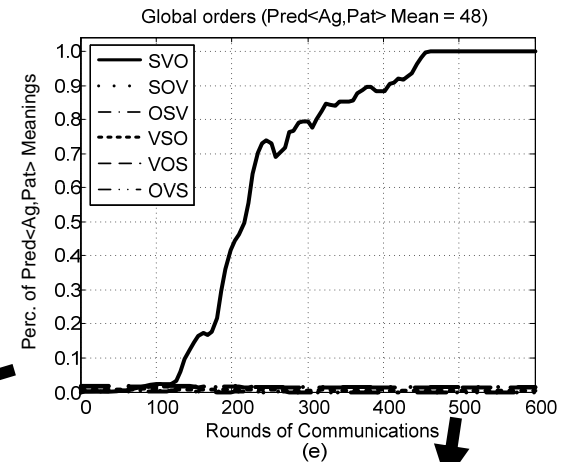
- **Emergence of compositionality:**
 - **RE** → Holistic signals → compositional expressions;
 - **UR** → The emergence of a common compositional language (**the S-shape sharp transition**).
- **Emergence of regularity:** some global word orders become **prevalent**;
 - **Prevalent order:** the global / 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.



5) Analyze the simulation results and verify the working hypotheses (cont.)

Global order's understandability

Local order's understandability



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



The advantage of computational simulation

- **Complementary** to empirical studies or theoretical analyses on language evolution.
 - Test the internal coherence of linguistic theories
 - Provide quantitative/qualitative analyses to evaluate verbally expressed linguistic theories
 - Generalize the available theories that are built upon incomplete data or trigger new ones
- Clearly studying the nature of language as a **Complex Adaptive System** (CAS, **Steels 2000; Wang 2006**):
 - **Language is a CAS**: involve various components and interactions among them, and some hierarchical structures;
 - Through manipulating parameters and repeating in similar conditions, computational simulation can **systematically examine** the collective effects of various components on language evolution.
- **Reliability** of computational simulation:
 - Adopt **plausible** assumptions verified by empirical findings in linguistics as well as other disciplines.
 - Use **objective** mechanisms, and follow **traceable** procedures to obtain **replicable** results.

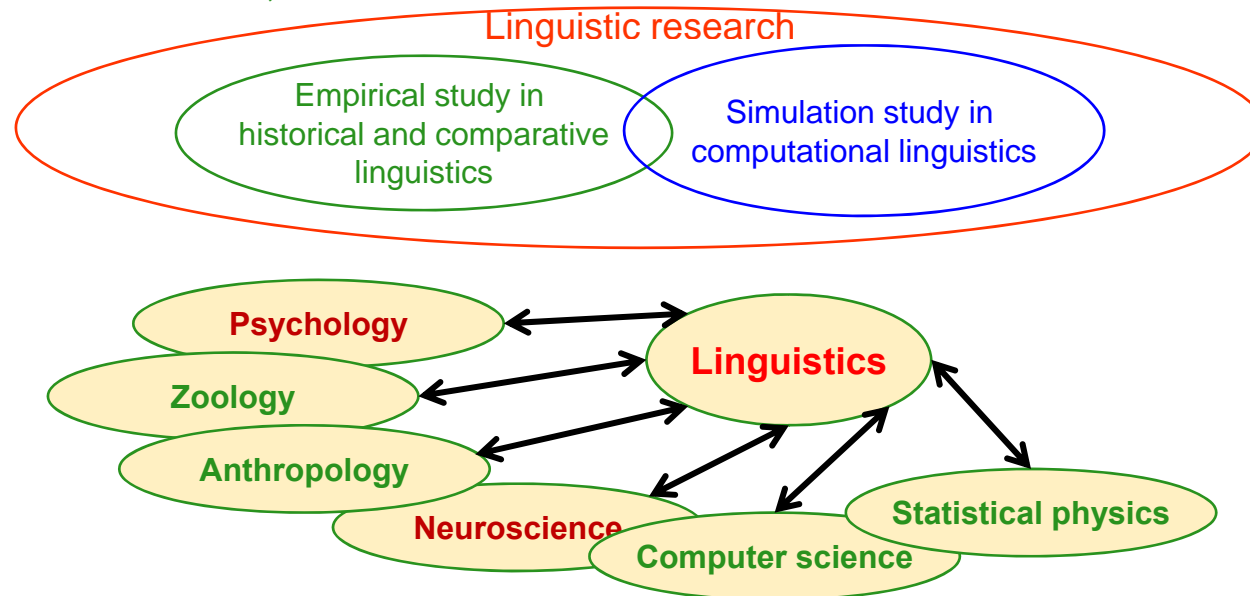
Steels, L. 2000. Language as a Complex Adaptive System, *Proceedings of the 6th International Conference on Parallel Problem Solving from Nature*: 17–26.

Wang, W. S-Y. 2006. 语言是一个复杂适应系统 (Language is a CAS). *Journal of Tsinghua University (Philosophy and Social Sciences)*, 21(6): 5-13.



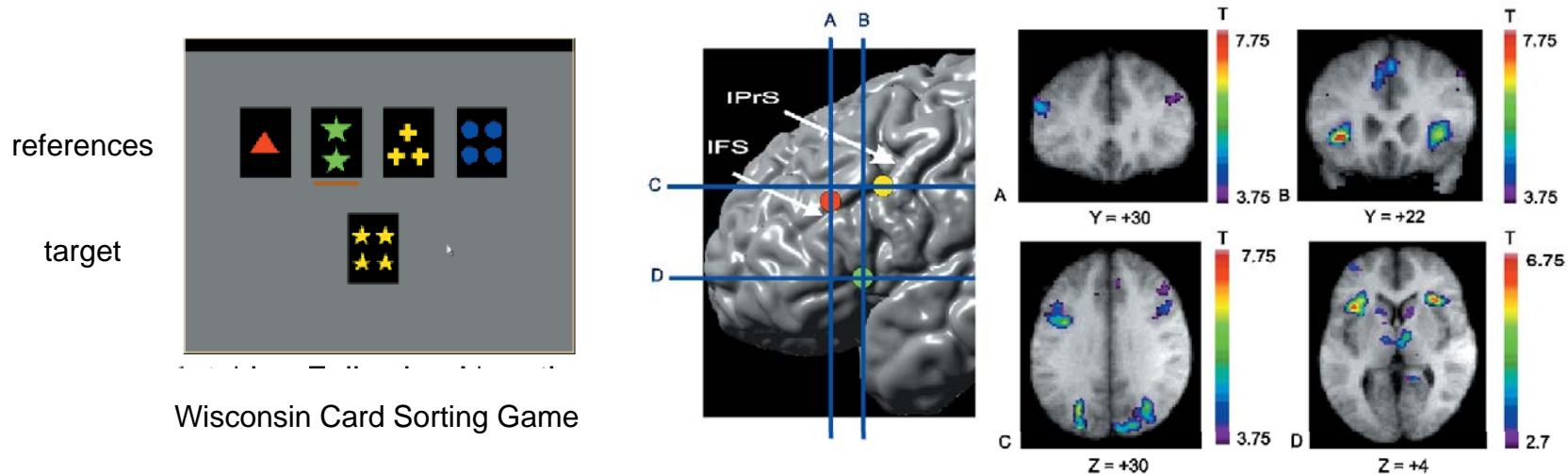
The multi-disciplinary nature of linguistic research

- The shortcomings of computational simulation: **Simplification and specification;**
- The difficulties of computational simulation
 - **Complexity and variety** of linguistic features and behaviors;
 - **Insufficient language data** to reconstruct prehistory;
 - **Inefficient techniques** on artificial intelligence and complex systems;
- 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;



Contribution of other subjects to computational simulation

- The neural bases of language-related abilities:
 - The pattern extraction ability for compositionality;
 - Wisconsin Card Sorting Game (Monchi et al. 2001): bi-lateralization activation in prefrontal cortex, basal ganglia, and thalamus for form and shift cognitive criteria.

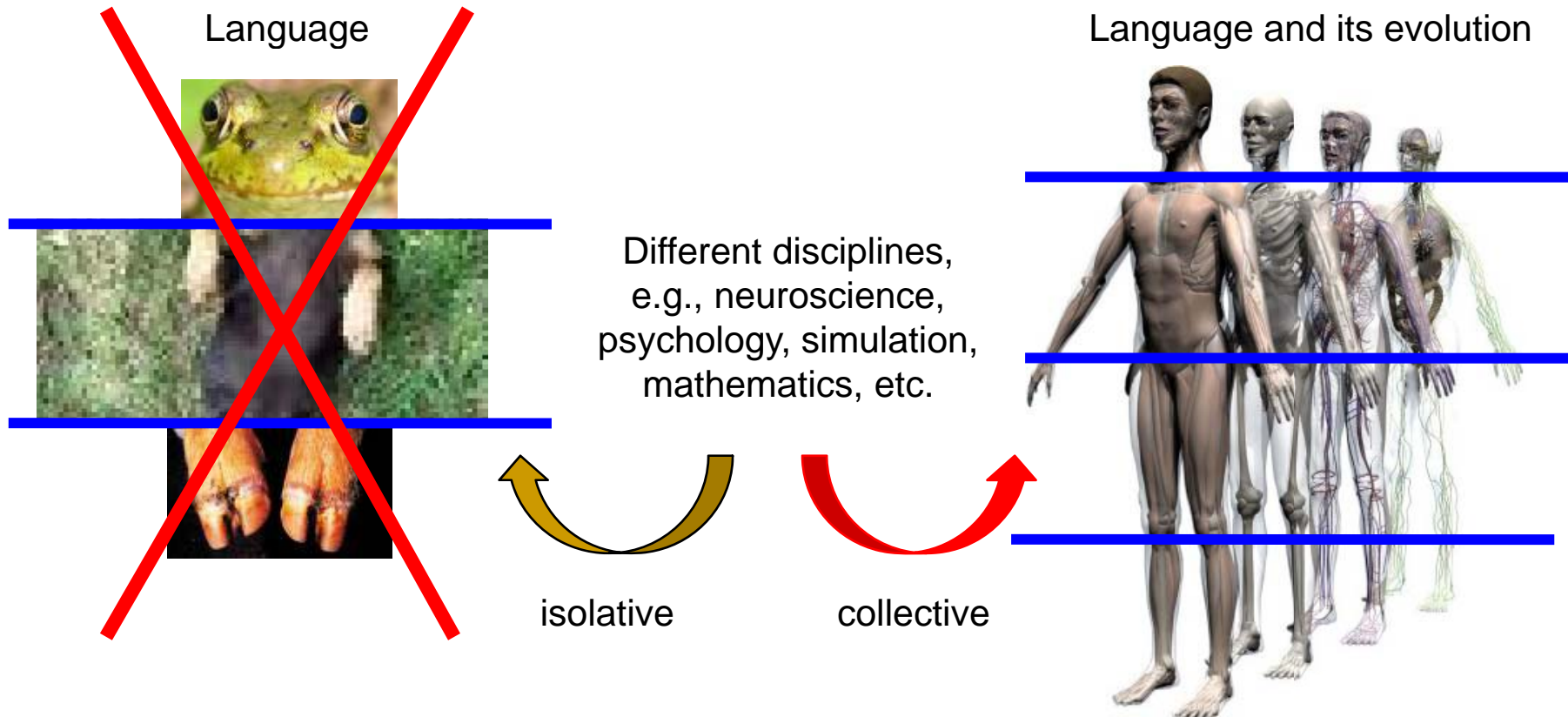


The mid-dorsolateral prefrontal cortex (BA 9/46) increased activity at the point when the current information must be related to earlier events stored in working memory.

- Sequential learning;
 - Parkinson's disease (PD), which damages the basal ganglia and mostly sparing cortex.
 - PD patients have difficulty in producing and comprehending sentences that have complex syntax (Grossman et al. 1993);

The multi-disciplinary nature of linguistic research

- Multi-disciplinary study is not “盲人摸象(blind people feeling an elephant)”



Different disciplines should contribute collectively to our understanding of human nature and language!



ありがとうございます!

Only humans have language

My work!



Have human languages gradually
evolved from domain-general abilities?

Domain-general abilities are shared among many animals