Language Evolution from a Simulation Perspective

Tao Gong
tgong@ee.cuhk.edu.hk

Language Engineering Laboratory,
http://www.ee.cuhk.edu.hk/~lel/
Dept. of EE, The Chinese University of Hong Kong
Feb. 2, 2007;

Outline

- The simulation perspective
  - Contemporary linguistic research methods;
  - The simulation perspective and some simulation methods;
- A compositionality-regularity coevolution model
  - Linguistic universals: compositionality and regularity;
  - The computational framework;
  - Representation and acquisition of linguistic knowledge;
  - Communication scenario;
- Simulation results
  - Coevolution of compositionality and regularity;
  - “Bottom-up” syntactic developmental process;
  - Cultural influence on language evolution;
- Discussion and some collaboration aspects
  - Neural basis of language-related abilities;

The simulation perspective

Contemporary linguistic research methods

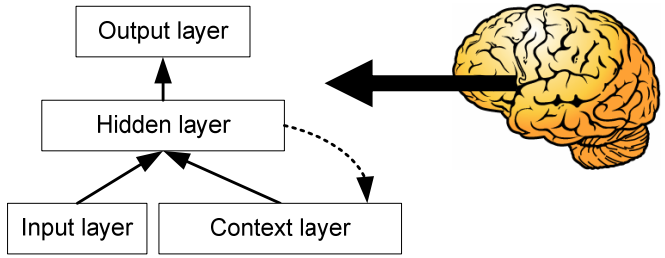
- Empirical data collections:
  - Corpus linguistics (Kučera and Francis 1967): e.g., *Child Language Data Exchange System* (CHILDES, http://childes.psy.cmu.edu), *Linguistic Data Consortium* (LDC, http://www.ldc.upenn.edu);
- Behavioral study on subjects' performance in linguistic tasks:
  - Categorical perception: Zheng Hongying's work;
  - Anticipatory and carryover effects: Wong Yingwei's work;
  - Neural activities during production or comprehension related linguistic behaviors: Susan Shuai's, Yang jing's and Alice Chan's work:
- Computational simulation:
  - Artificial neural network model: Francis' work;

Kučera, H. and Francis, N. 1967. *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

The simulation perspective

- The simulation perspective: to adopt computational simulations to recapitulate the evolution of human language, reconstruct the language history, and reconsider the effects of linguistic and nonlinguistic factors on this developmental process.

- Behaviourist simulations
  - Artificial neural network:

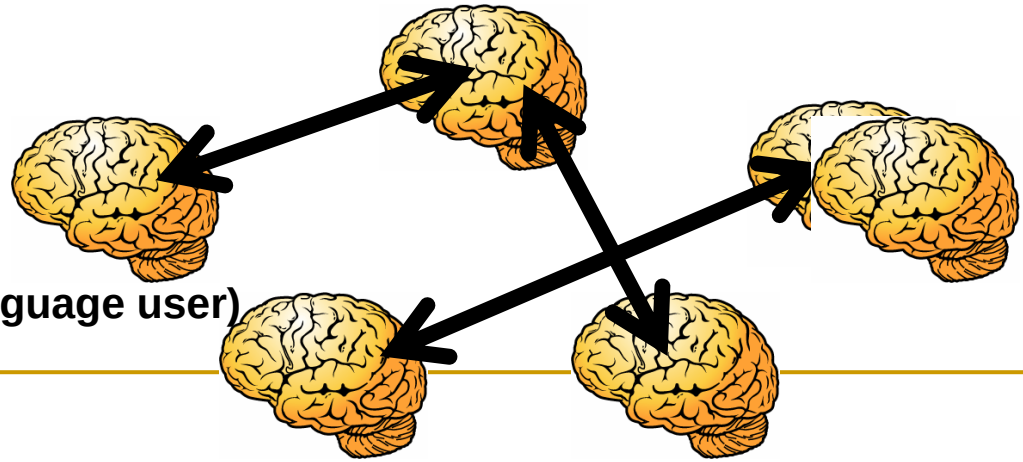


  - Rule-based system:

“cat”←→*/cat/*; “run”←→*/run/*;

/cat/ << /run/

  - Multi-agent system:



Agent (language user)

A compositionality-regularity coevolution model

Linguistic universals: compositionality and regularity

- **Compositional expressions:** the meaning of a complex expression is determined by its components (lexical items)' meanings.
  - E.g., **English:** "eat<dog, meat>"←→/DOG EAT MEAT/
    Chinese: "吃<狗,肉>"←→/狗 吃 肉/

- **Holistic expressions:** the meaning of a complex expression is determined as a whole, inseparable.
  - E.g., **English:** "to be angry"←→/HIT THE CEILING/
    Chinese: "to miss the appoint"←→/放鴿子、放飛幾/

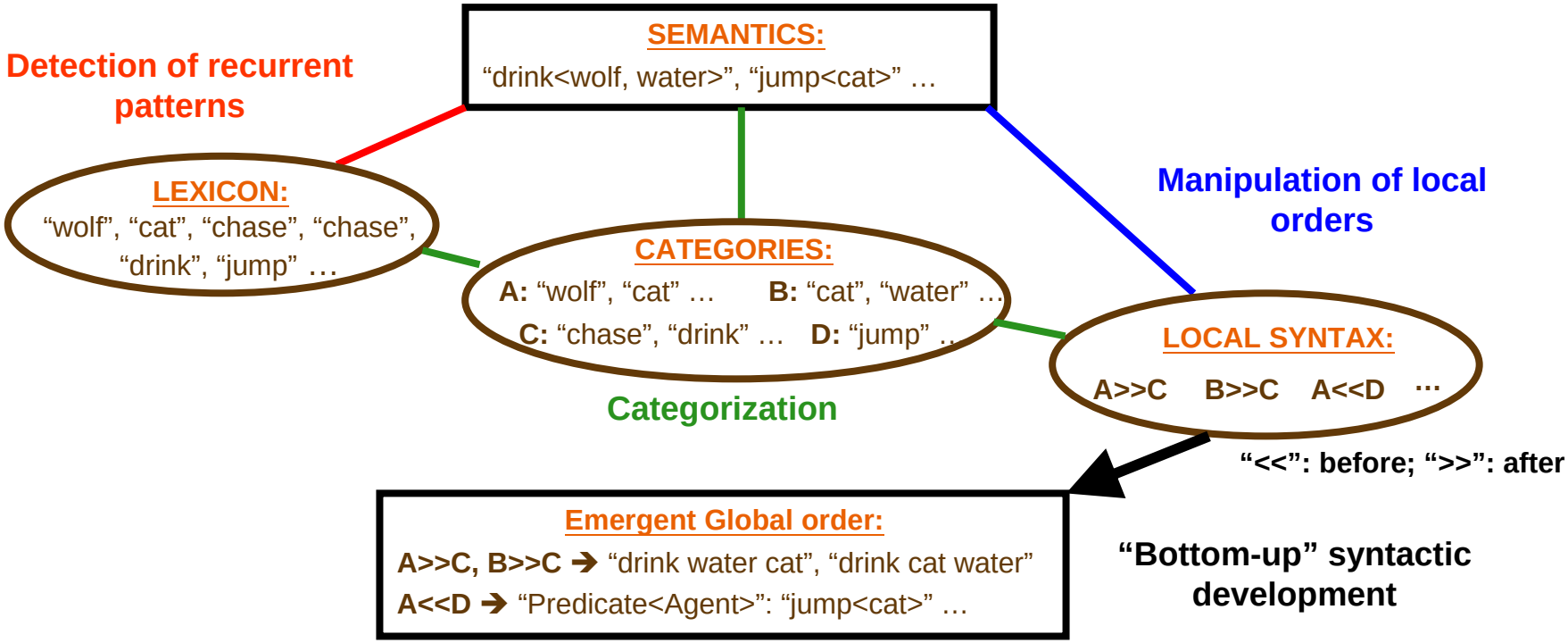
- **Regularity:** the meaning of a complex expression is also determined by its structures (e.g., word order).
  - E.g., different word order affects the meaning;
    English (Subject Verb Object):
    "bite<dog, man>"←→/DOG BITE MAN/;
    "bite<man, dog>"←→/MAN BITE DOG/

How did compositionality (lexical items) and regularity (word order) emerge in human language?

A compositionality-regularity coevolution model

Our coevolution scenario of developing compositionality and regularity

- **Our hypothesis: coevolution of compositionality and regularity out of a holistic signaling system**



Local syntax: binary sequential relation (before or after) between 2 lexical items, e.g., SV, OV, SO;

Global word order (e.g., SVO, SOV) results from local syntax;

A compositionality-regularity coevolution model

Representation and acquisition of linguistic knowledge

- Linguistic rule: condition (M-U mapping or local syntax) + (strength);
  - Integrated meanings:

    Type1: "Pr_1<Ag>": e.g., "run<dog>";

    Type2: "Pr_2<Ag, Pat>": e.g., "chase<fox, wolf>";

  - Lexical rule: how to map semantic components to utterance syllables;

    Holistic rules

    (1) "chase<wolf, bear>"←→/a d/ (0.5)

    (2) "hop<deer>"←→/a/ (0.4)

    Compositional rules

    (3) "wolf"←→/d/ (0.6)

    (4) "chase<#, bear>"←→/a b * d/ (0.7)

  - Syntactic rule: local syntax (BEFORE / AFTER) on regulating 2 lexical rules' syllables.

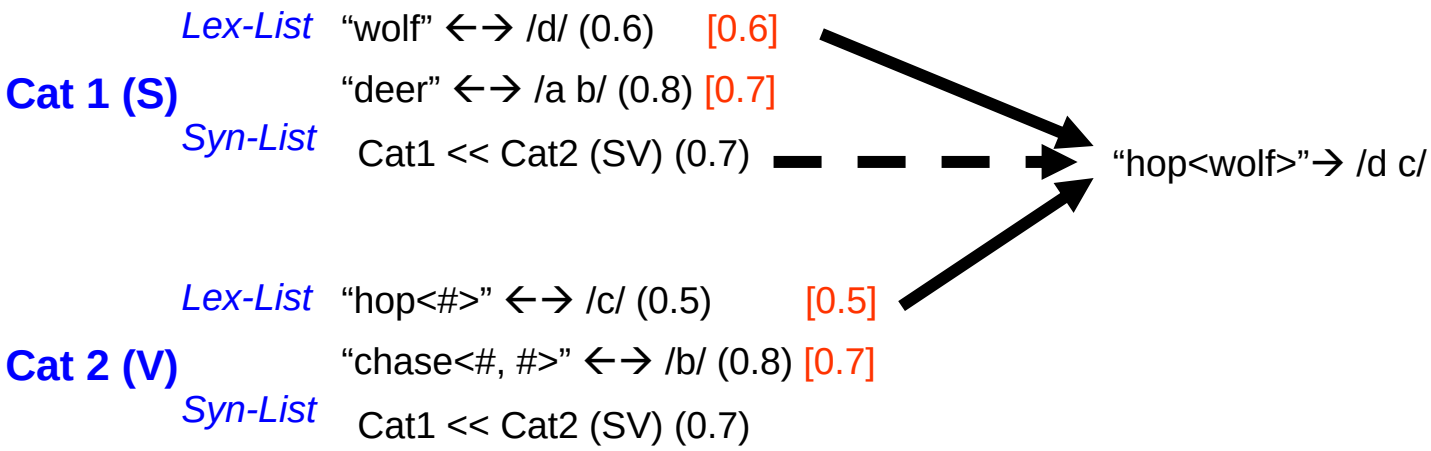
    (I) Cat1 << Cat2 (SV) (0.5)
    BEFORE

    (II) Cat1 >> "fight<#, #>" ←→ /d/ (0.4)
    AFTER

A compositionality-regularity coevolution model

Representation and acquisition of linguistic knowledge

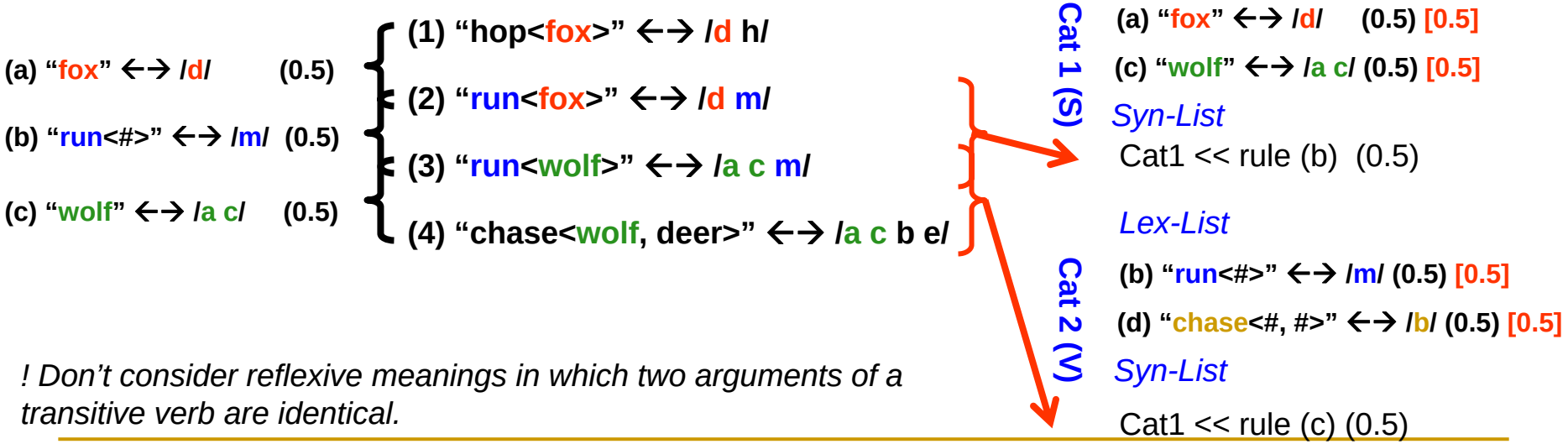
- *Syntactic categories:*
  - Syntactic role (S, O or V), Lex-List and Syn-List;
  - [Association weights]: the strength to associate a lexical rule into a category;



A compositionality-regularity coevolution model

Representation and cquisition of linguistic knowledge

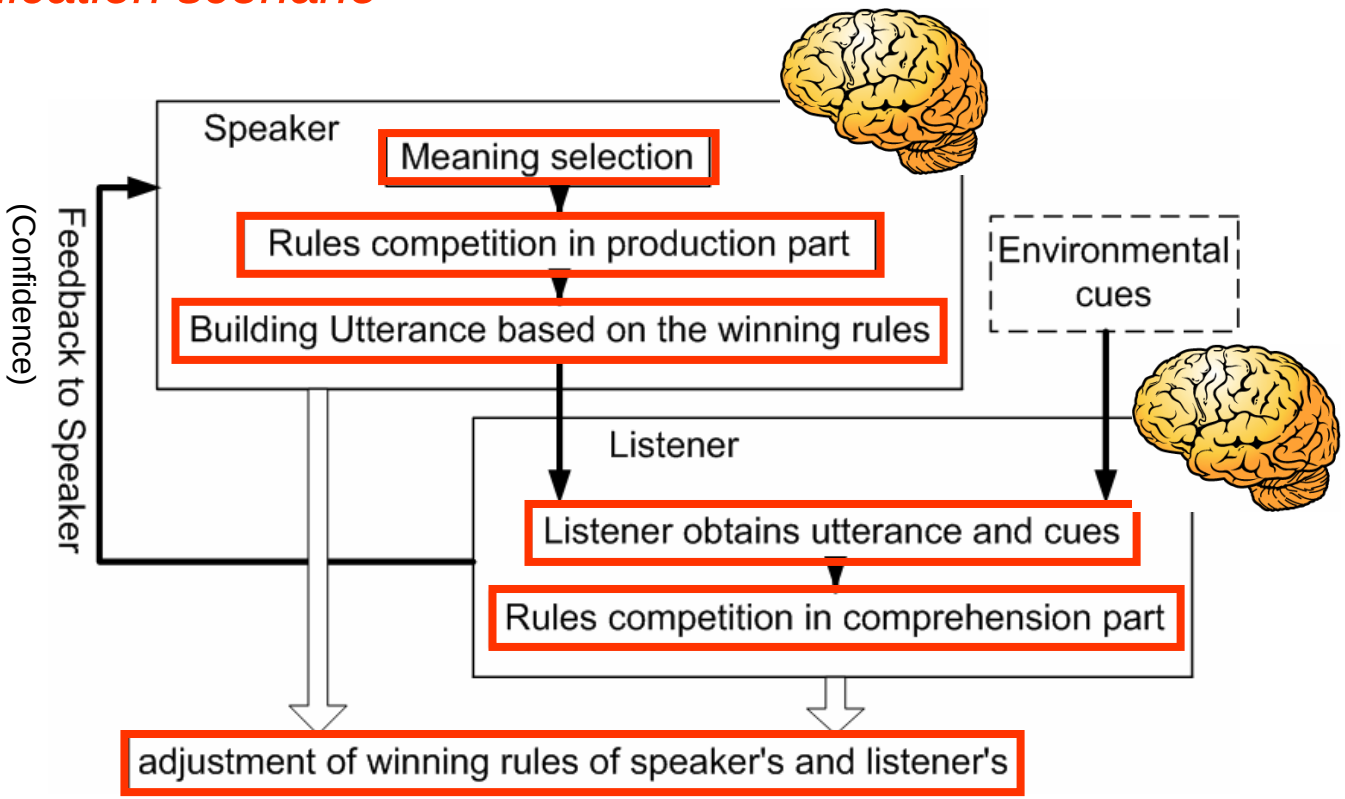
- **Acquiring lexical rules**
  - Random creation of holistic rules and **Detection of recurrent patterns;**
- **Acquiring local orders and syntactic categories**
  - Lexical rules with same semantic role and similarly used (local order) in M-U mappings can be associated into categories; **lexical rules**



! Don't consider reflexive meanings in which two arguments of a transitive verb are identical.

A compositionality-regularity coevolution model

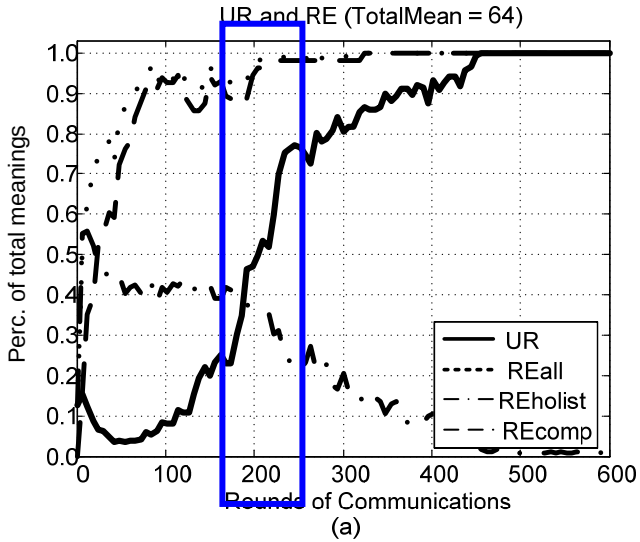
Communication scenario

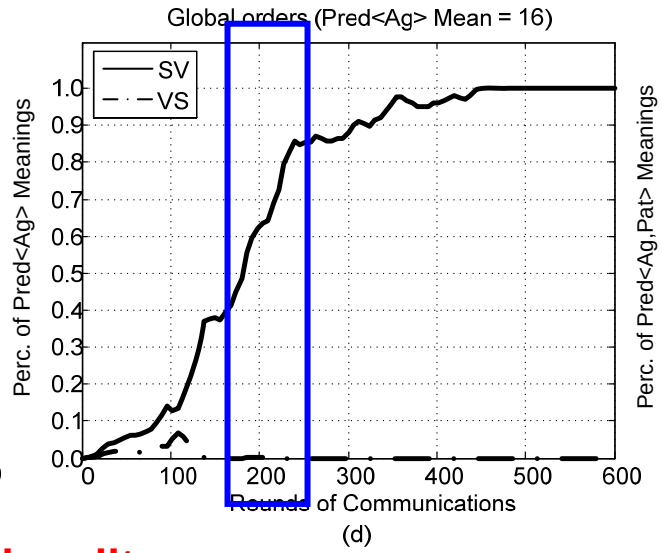


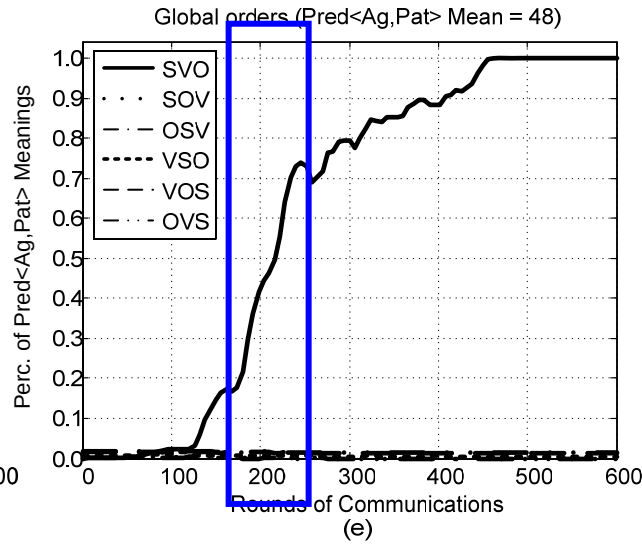
- Indirect Meaning Transference:
  - Independent linguistic knowledge;
  - Confidence feedback based on the combined strength of the used rules;

Simulation results

Coevolution of compositionality and regularity







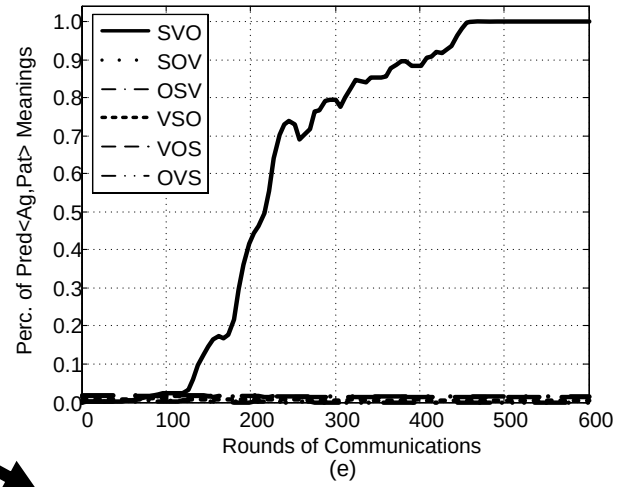
- **Emergence of compositionality:**
  - **RE** ➔ Holistic signals → compositional expressions;
  - **UR** ➔ the emergence of a common compositional language.
- **Emergence of regularity:** some global word orders becoming **prevalent**;
  - **Prevalent order:** the global / local orders that are frequently used in comprehensions;
  - e.g., **SV** for "Pr_1<Ag>" meanings; **SVO** for "Pr_2<Ag, Pat>" meanings;

- **Coevolution of compositionality and regularity:**
  - The co-occurrence of the sharp increase of the UR and the sharp increase of the understandability of the prevalent global word orders;

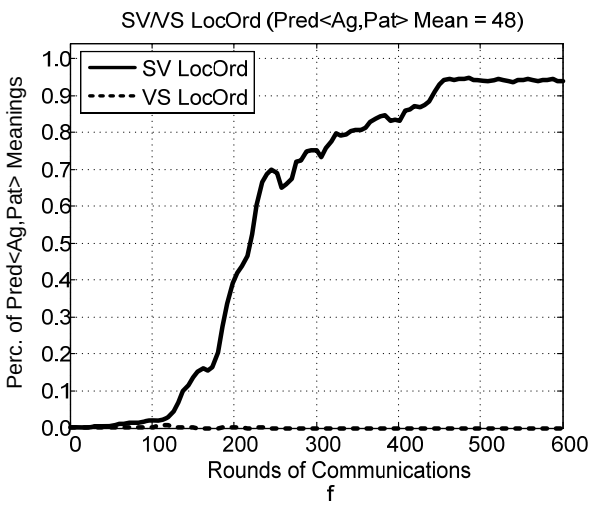
Simulation results

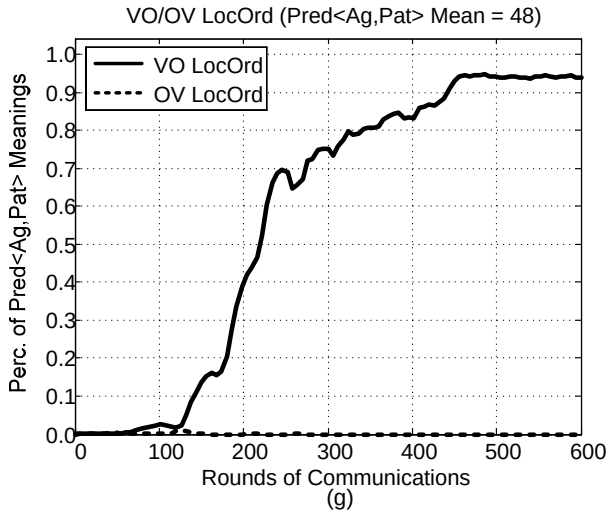
"Bottom-up" syntactic developmental process

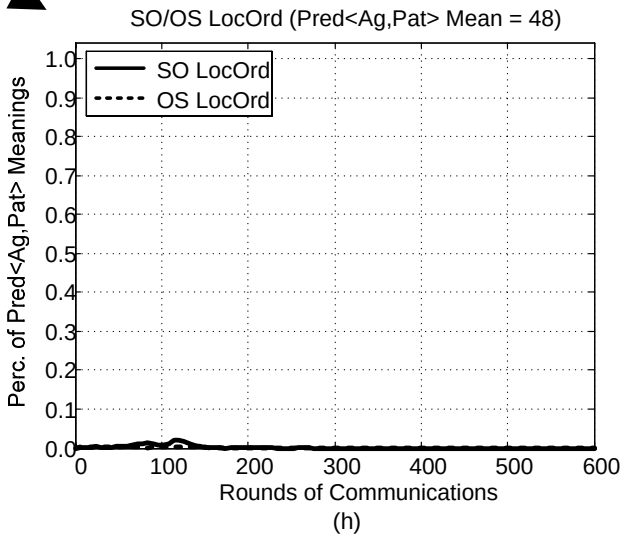
Global order's understandability



Local order's understandability





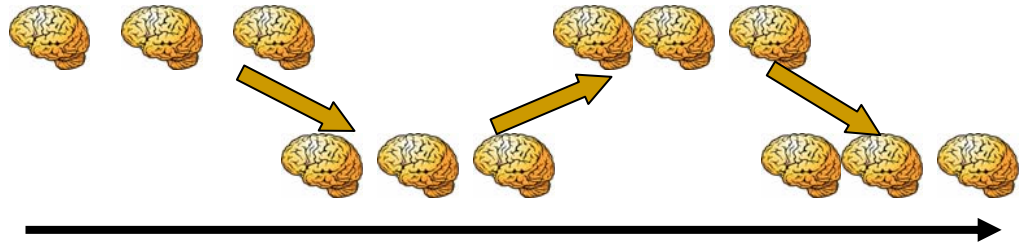


- The prevalent global word order reflects the local sequential information specified by the prevalent local syntax;

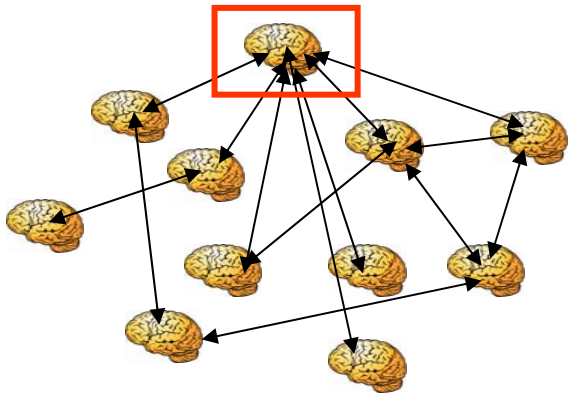
Simulation results

Cultural influence on language evolution

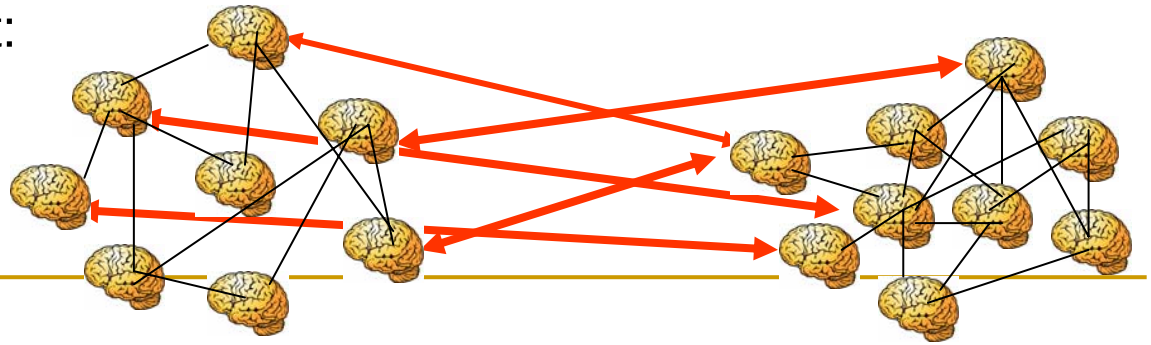
- Language is developed in a cultural environment:
  - Cultural transmission:



  - Social structure's effects:



  - Language contact:

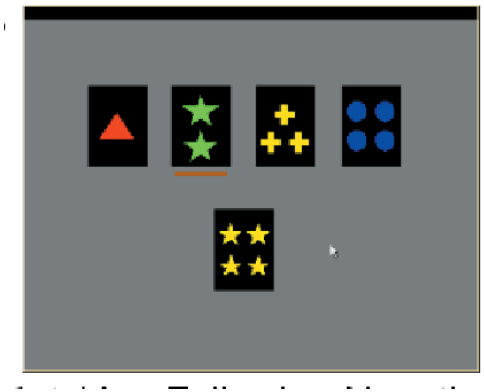


Discussion and some collaboration aspects

Neural basis of language-related abilities

- Most language-related abilities built in this model are domain-general, not specific for humans or language;
  - Empirical support from animal studies;

- The neural basis of language-related abilities:
  - The pattern extraction ability for compositionality:



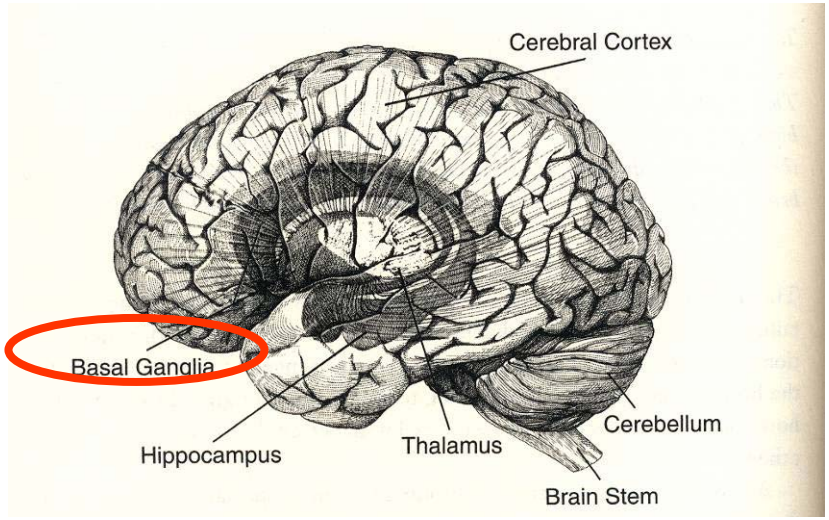
The mid-dorsolateral prefrontal cortex (BA 9/46) increased activity at the point when the current information must be related to earlier events stored in working memory.

Monchi et al. 2001. Wisconsin card sorting revisited. *The Journal of Neuroscience*, 21(19): 7733-7741.

Discussion and some collaboration aspects

Neural basis of language-related abilities

- The sequential learning ability for regularity:
  - The subcortical structure, Basal ganglia, plays the sequencing role of in motor control as well as syntactic abilities in human language (Lieberman 2006)



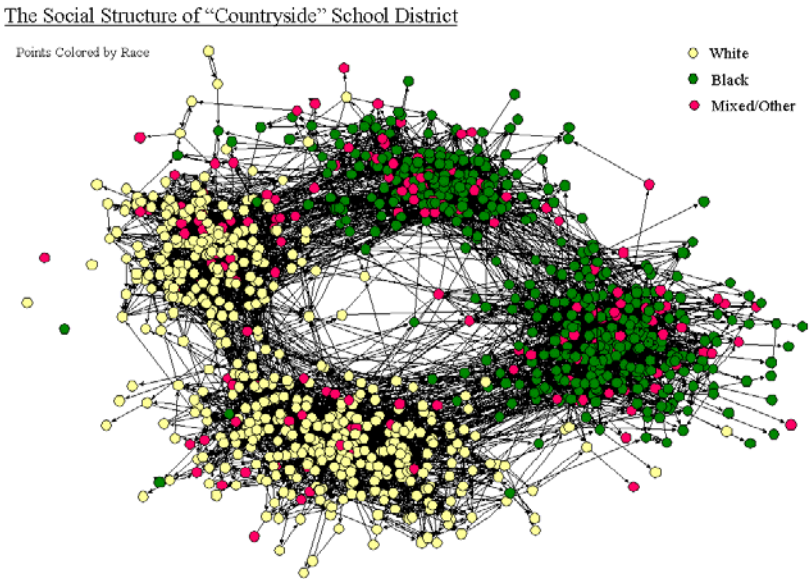
(from Edelman, G. M. and Tononi, G. 2000. *A universe of consciousness: How matters become imagination*. New York: Basic Books)

Lieberman, P. 2006. *Towards an evolutionary biology of language*. Cambridge, MA: Belknap Press of Harvard University Press.

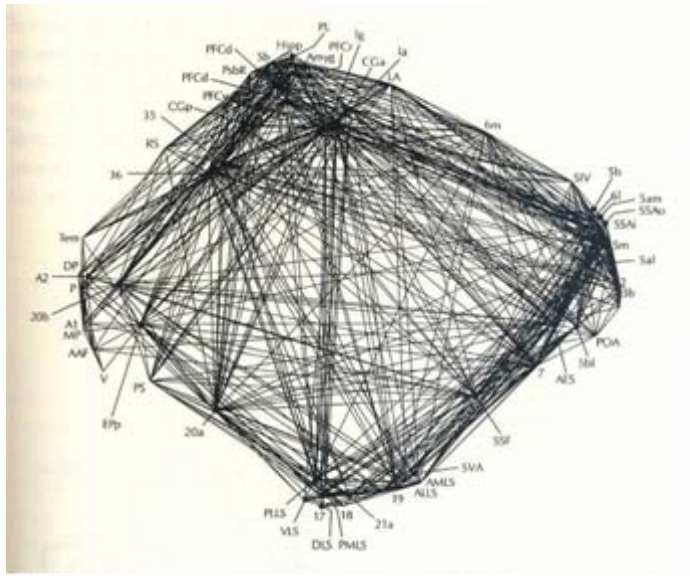
Discussion and some collaboration aspects

Neural basis of language-related abilities

- Language is a whole brain function:
  - Linguistic knowledge is stored as complex neural networks in human brains (Lieberman 2006);
  - Complex network theory provides guidelines for the study of linguistic activities in human brains.



Friendship networks of countryside school district in Columbia, Moody, http://www.sociology.ohio-state.edu/jwm/

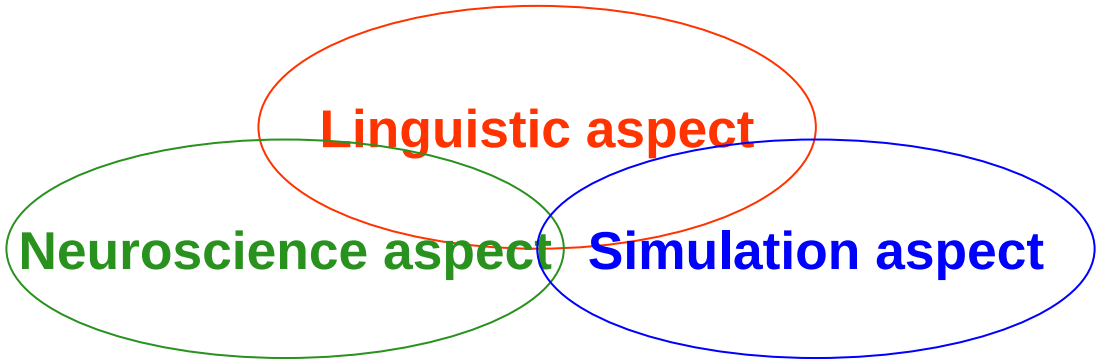


Edelman, G. M. and Tononi, G. 2000. *A universe of consciousness: How matters become imagination*. New York: Basic Books.

Discussion and some collaboration aspects

Multi-perspectives to tackle linguistic problems

- Analysis of linguistic database;
- Digest of linguistic theories;
- Psychological studies on linguistic behaviors;



- Empirical basis for simulated linguistic behaviors ;
- Suggest new related behaviors for simulations;

- Implement linguistic behaviors and study their long-term effects on language evolution;
- Suggest further behavioral studies about simulated linguistic behaviors;