

Language Acquisition Corpus and Applications: Issues Related to Spoken Data

By

Weiping WU

The Chinese University of Hong Kong

For

SLATUFS2019 (Tokyo, Japan, June 1-2, 2019)



Outline

- L2 Learners and spoken corpus
- Language structure vs language use
- Key issues in LAC/spoken
- Samples of applications based on LAC/sc
- Discussion and concluding remarks

L2 Learners and spoken corpus

- L2 in different contexts
- ESL, CSL, JSL and xSL
 - who you are
 - where you are
- Spoken vs written data
 - dialogues in movies and drama
 - words in social media/wechat
- SLA for CSL learners: Order&frequency

LS vs LU

- LS: Language as a system of systems
- LU: Language as a communication tool

Hymes: communicative competence

Kramersch: language and culture

Wu: Pragmatic Framework

Ellis: TBLT

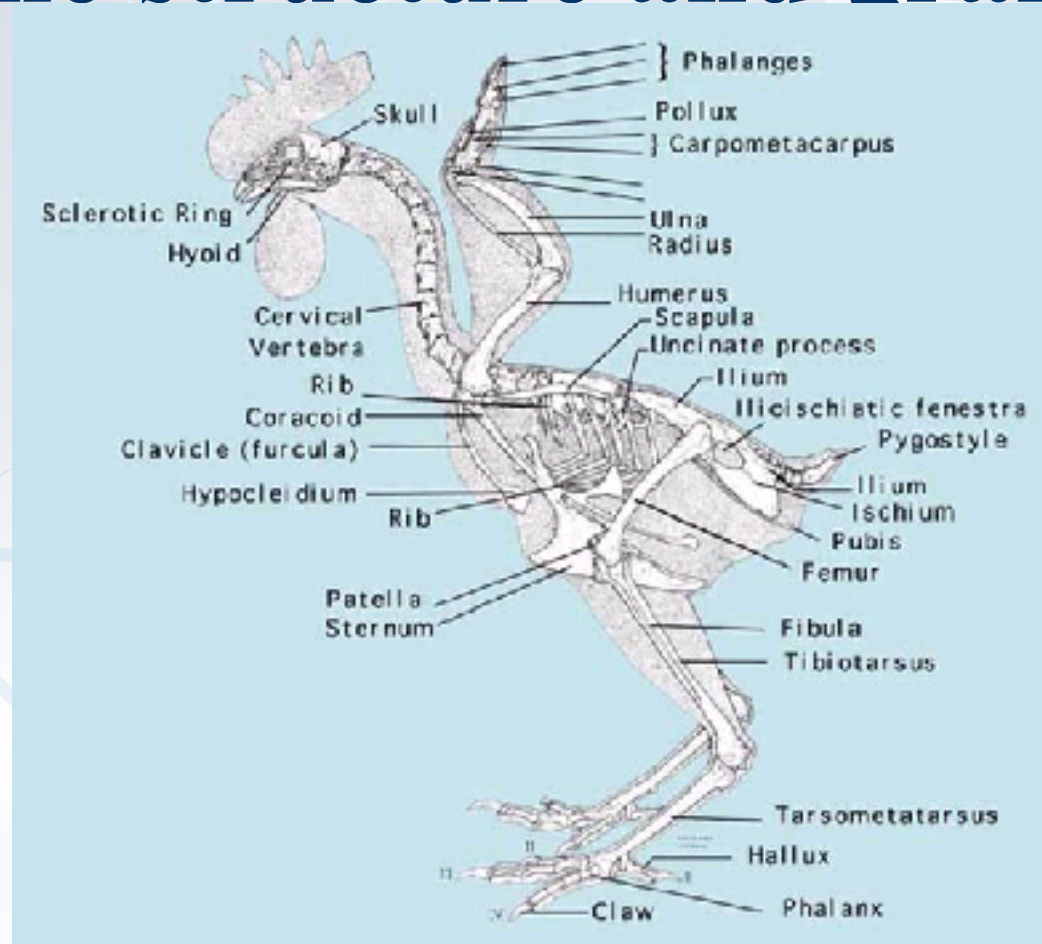
Wang: LTUT

Tao/Feng etc. : Yuti grammar

.....

LS: The rooster analogy

(bone structure and grammar)

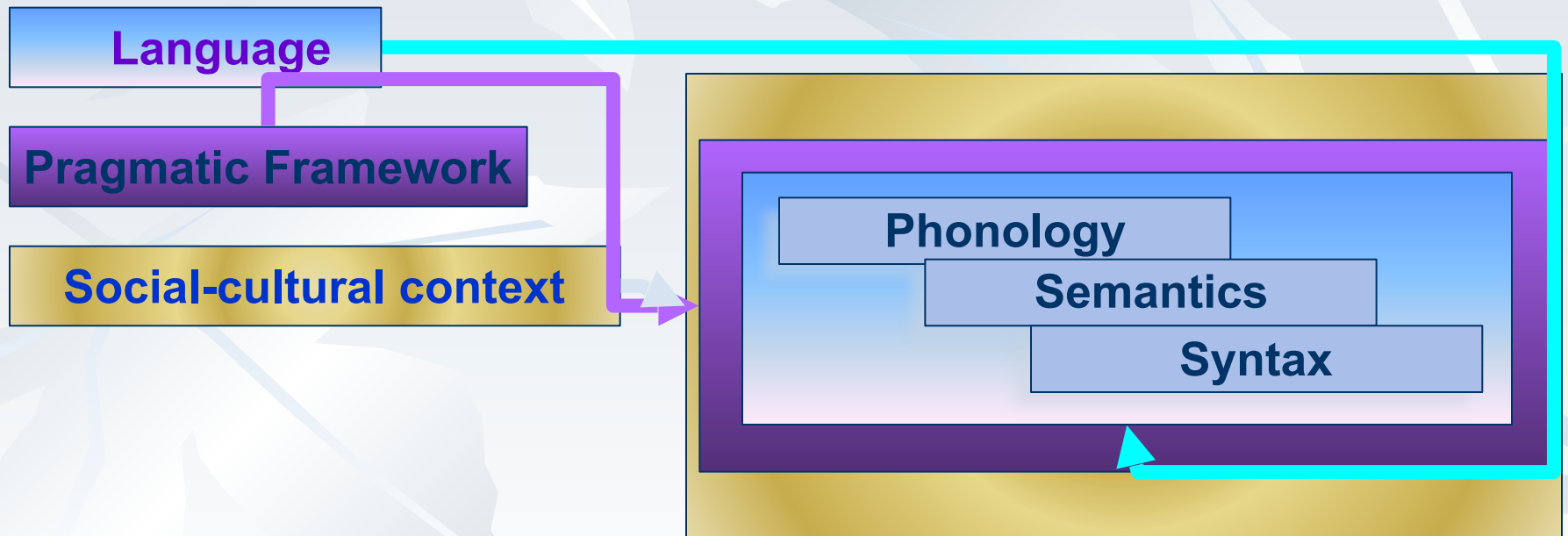


LU: The rooster analogy (live) (WU/2006)



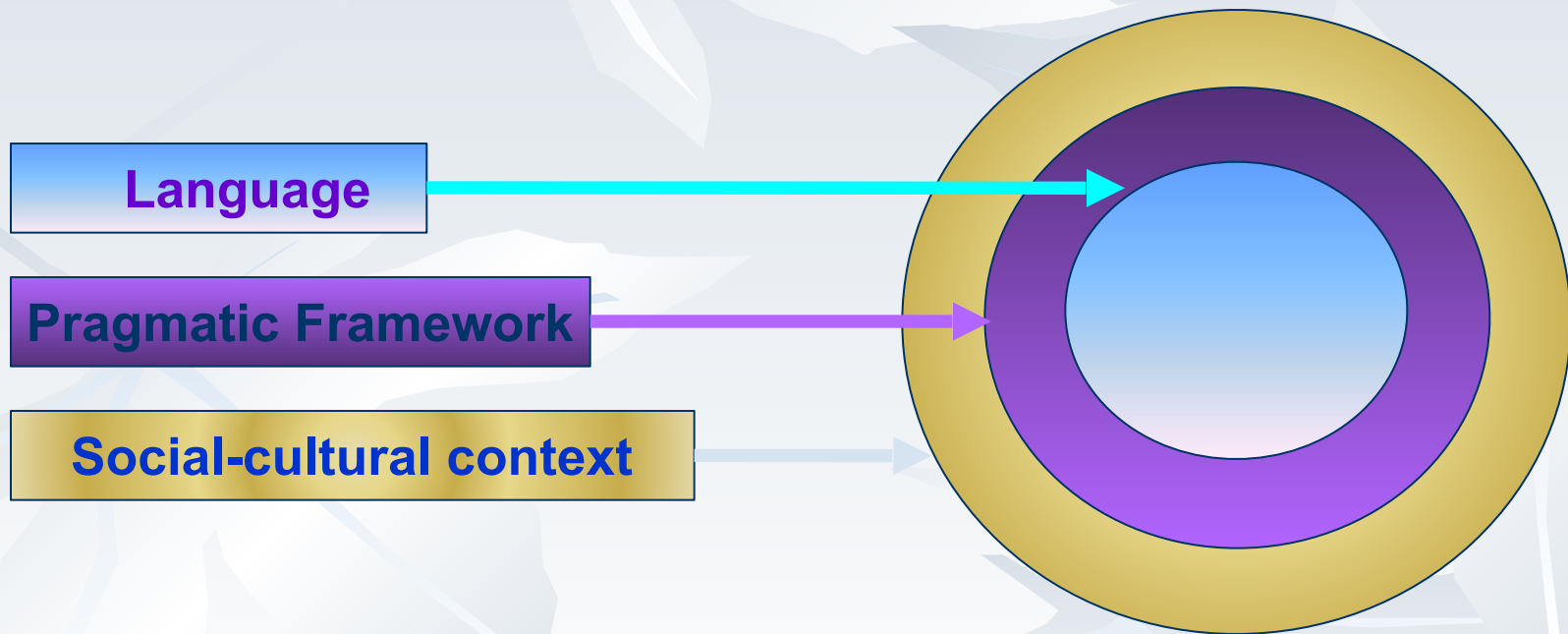
LU: Pragmatic Framework

- Figure 1A. Language, pragmatic framework and socio-cultural context (e.g. CSL)



LU: Pragmatic Framework

- Figure 1B. Language, pragmatic framework and socio-cultural context (e.g. ESL, etc.)



Key issues in LAC/spoken

- Language and cultural backgrounds
X4y (E4J/C4J/.....); X4x (J4J/P4P)
- Proficiency levels (for oral productions)
PSC/1A-3B; CEFR/A1-C2; ACTFL/I-S
- Authenticity of spoken data
L1: “by native for native”
L2: ??? (N2NN? NN2N? NN2NN?)
aware/not aware? test/non-test?

LAC/sc: A case study of CSL

- CSL learners with background in Japanese, Korean, English, and Cantonese: P4J/P4K/P4E/P4C
- Proficiency levels (for oral productions)
Modified version of the ACTFL Scale
- Possible structures of the Corpus
by native language/culture of learners
by proficiency levels

LAC/sc: Samples of applications

- Preparation of corpus for research
 - Pros and cons of tagging
 - Problems with transcription
- Synchronic and diachronic studies
 - Contrastive study in formal features between P4E and P4C CSL learners at the advanced level
 - Use of politeness forms in public speech by P4E learners with different proficiency levels

Discussion for spoken corpus

- Authentic vs semi-authentic data?
- Amount/level and accuracy of tagging?
- Basis for search: Written or oral features?
- The best vs the best-possible product?

Concluding remarks

- Facing the urgent needs for teaching and research:
Get started
- Improving in the process of building:
Start with a modular approach
- Doing corpus-based research:
Approach from your own interests

Thank you

Q&A

Weiping M. WU, Ph.D.

Director

Yale-China Chinese Language Center

The Chinese University of Hong Kong

Hong Kong SAR

Website: <http://www.cuhk.edu.hk/clc/staff/wwp.htm>

Email:

wwpclc@cuhk.edu.hk

