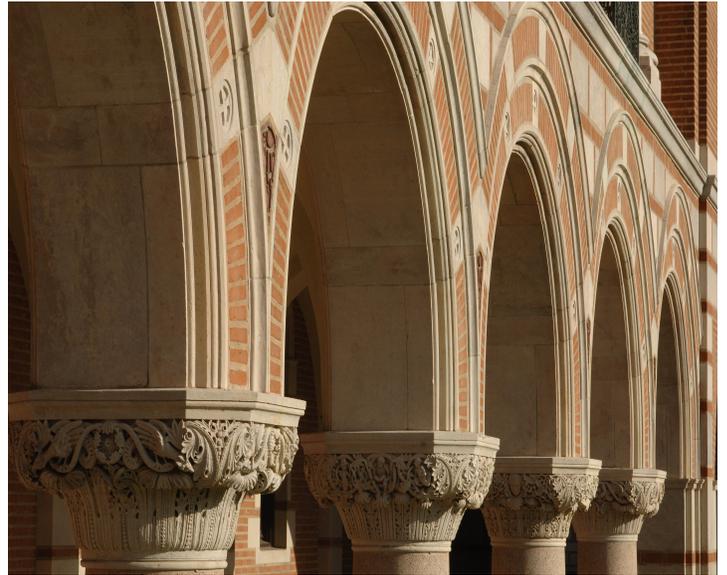


# CLIC

Center for Languages & Intercultural  
Communication



## The 2<sup>nd</sup> International Conference on Spoken Chinese Corpora: From Theory to Pedagogical Applications

**August 10-11, 2016**

# TABLE OF CONTENTS

*The 2<sup>nd</sup> International Conference on Spoken Chinese Corpora:  
From Theory to Pedagogical Applications*

*Hosted by: Rice University*

*Co-Sponsors: Nanjing University & Beijing Language and Culture University*

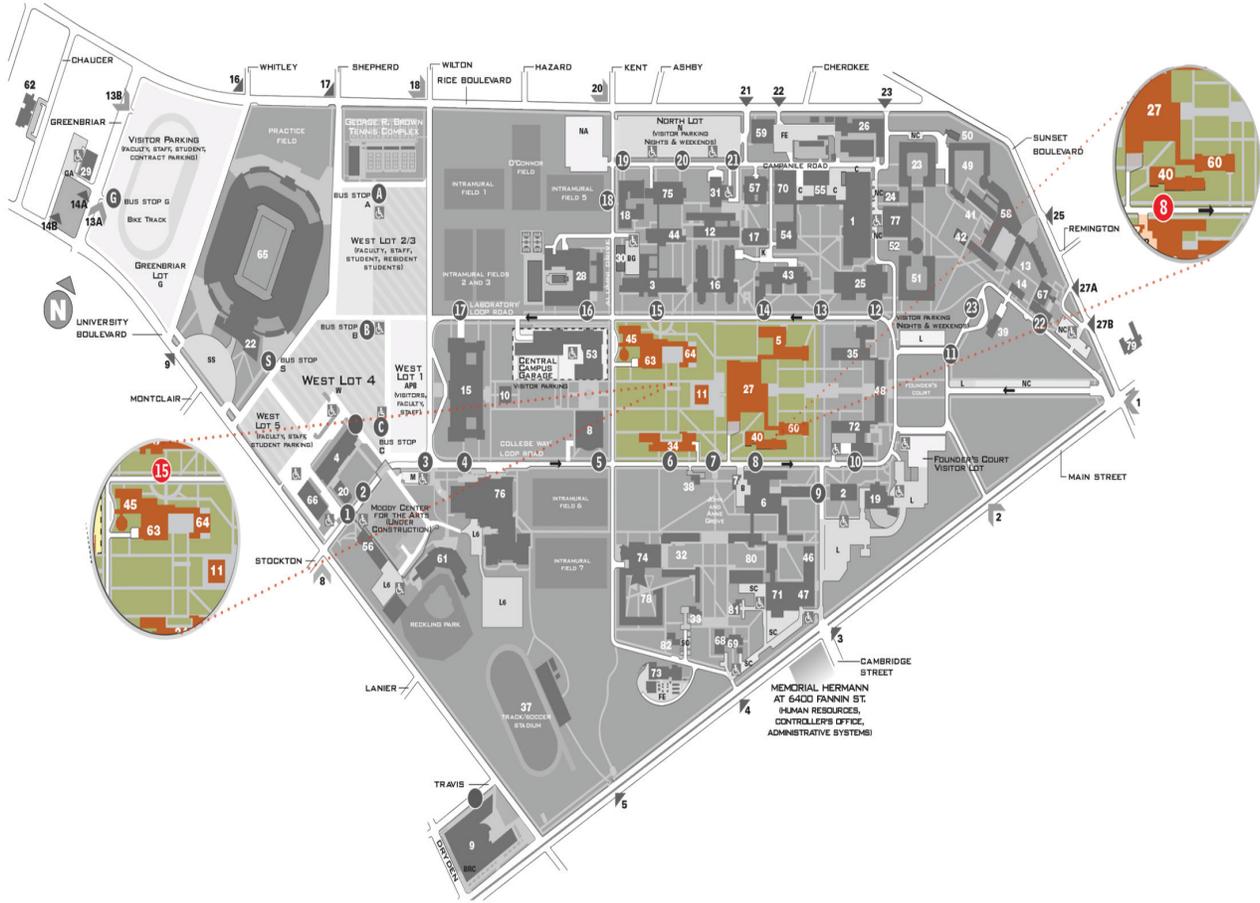
Campus Map.....	3
Conference Schedule.....	4-6
Plenary Speakers and Bios.....	8-9
Presenters and Abstracts.....	9-24
Notes.....	25-27



1<sup>st</sup> International Conference on Spoken Chinese Corpora, Nanjing University  
July 10-11, 2015

# RICE UNIVERSITY CAMPUS MAP

8/28/2015



## MAP KEY

- Entrance Gates
- Visitor's Entrances
- Bus Stops
- One-way Road

## PARKING KEY

- Faculty/Staff Parking
- Resident Student Parking
- Commuter Parking
- Visitor Parking
- Accessible Parking

## PARKING LOTS:

- APB** Alice Pratt Brown Hall Lot
- B** Baker College-Housing & Dining Lot
- BG** Biology-Geology Lot
- BRC** BioScience Research Collaborative Garage
- C** Campanile Lot
- CG** Central Campus Garage
- FE** Facilities, Engineering and Planning Lot
- G** Greenbriar Lot
- GA** Greenbriar Annex
- K** Keck Lot
- L** Lovett Lot
- L6** Lot Six
- M** Moody Lot
- N** North Lot
- NA** North Annex Lot
- NC** North Colleges Residents Lot
- SC** South Colleges Residents Lot
- SS** South Stadium Lot
- W** West Lot

## BUILDINGS

Abercrombie Engineering Laboratory.....	1	Greenbriar Building.....	29
Allen Business Center.....	2	Greenhouse.....	30
President's Office		Hamman Hall.....	31
Anderson Biological Laboratories, M.D. ...	3	Hanszen College.....	32
Anderson-Clarke Center.....	4	Hansen College Masters House.....	33
Dean of Susanne M. Glascock School of Continuing Studies		Herring Hall, Robert R.....	34
Anderson Hall, M.D.....	5	Herzstein Hall.....	35
Dean of Architecture		Holloway Field.....	37
Baker College.....	6	Housing and Dining.....	38
Baker College Masters House.....	7	Hudspeth Auditorium.....	20
Baker Hall, James A. III.....	8	Huff House, Peter and Nancy.....	39
Dean of Social Sciences, James A. Baker III		Alumni Affairs, Center for Student Professional Development	
Institute for Public Policy		Humanities Building.....	40
BioScience Research Collaborative.....	9	Dean of Humanities	
Booth Centennial Pavilion, Suzanne Deal...10		Jones College, Mary Gibbs.....	41
James Turrell's "Twilight Epiphany" Skyspace		Jones College Masters House.....	42
Brochstein Pavilion, Raymond and Susan.....11		Keck Hall.....	43
Brockman Hall for Physics.....	12	Keith-Wiess Geological Laboratories.....	44
Brown College, Margaret Root.....	13	Ley Student Center.....	45
Brown College Masters House.....	14	Lovett College, Edgar Odell.....	46
Brown Hall, Alice Pratt.....	15	Lovett College Masters House.....	47
Dean of Shepherd School of Music		Lovett Hall.....	48
Brown Hall, George R.....	16	Admission Office, Welcome Center	
Dean of Wiess School of Natural Sciences		Martel College.....	49
Brown Hall for Mathematical Sciences, ...17		Martel College Masters House.....	50
Herman		McMurtry College, Burton and Deedee..51	
Butcher Hall, Dell.....	18	McMurtry College Masters House.....	52
Cohen House, Robert and Agnes.....	19	McNair Hall, Janice and Robert.....	53
Cox Fitness Center, John L.....	22	Dean of Jesse H. Jones Graduate School of Business	
Duncan College.....	23	Mechanical Engineering Building.....	54
Duncan College Masters House.....	24	Mechanical Laboratory.....	55
Duncan Hall, Anne and Charles.....	25	Media Center.....	56
Dean of George R. Brown School of Engineering		Mudd Computer Science Building.....	57
Facilities Engineering and Planning.....26		North Servery.....	58
Building		Oshman Engineering Design Kitchen.....	59
Fondren Library.....	27	Rayzor Hall.....	60
Gibbs Recreation and Wellness Center, ...28		Reckling Park at Cameron Field.....	61
Barbara and David		Rice Children's Campus.....	62

Rice Memorial Center.....	63
University Bookstore	
Rice Memorial Chapel.....	64
Rice Stadium.....	65
Rice University Police Department.....	66
Rich Student Health Center, Morton L.....	67
Richardson College, Sid W.....	68
Richardson College Masters House.....	69
Ryon Engineering Laboratory.....	70
Seibel Servery, Abe and Annie.....	71
Sewall Hall.....	72
Rice Art Gallery	
South Plant.....	73
South Servery.....	74
Space Science and Technology Building..75	
Tudor Fieldhouse and Youngkin Center..76	
Autry Court, Fox Gymnasium	
West Servery.....	77
Wiess College.....	78
Wiess President's House.....	79
Will Rice College.....	80
Will Rice College Masters House.....	81
Wilson House.....	82
Wiess College Masters	

## ACADEMIC SCHOOLS

Architecture, School of.....	5
Business, Jesse H. Jones.....	53
Graduate School of Continuing Studies, Susanne M.....	21/4
Glascock School of Engineering, George R. Brown.....	25
School of Humanities, School of.....	40
Music, The Shepherd School of.....	15
Natural Sciences, Wiess School of.....	16
Social Sciences, School of.....	8

## PARKING RATES:

West of Entrance 18: \$1 each 20 minutes, \$11 daily maximum  
 East of Entrance 18 and BioScience Research Collaborative Garage: \$1 each 12 minutes, \$11 daily maximum  
 Greenbriar Lot: \$1 per entry, per day

## PAYMENT METHODS:

BioScience Research Collaborative Garage and Central Campus Garage: cash or credit card  
 Greenbriar Lot, Founder's Court, North and West Lots Visitor Section: credit card

## OFF-CAMPUS FACILITIES

Rice Graduate Apartments  
 Rice Village Apartments

# CONFERENCE SCHEDULE

## WEDNESDAY, AUGUST 10<sup>TH</sup> – Rayzor Hall 123

- 8:30-9:00 Breakfast in Rayzor Hall Lobby
- 9:00-9:30 Welcome Remarks and Group Photos  
*Dr. Rafael Salaberry, Director of Center for Languages & Intercultural Communication, Rice University*  
*Dr. Lora Wildenthal, Associate Dean of Humanities, Rice University*  
*Dr. Aimin Cheng, former Dean of Institute for International Students, Nanjing University*
- 9:30-10:45 Plenary Speaker  
*Hongyin Tao (陶红印), UCLA*  
Exploiting Spoken Corpora for Chinese Language Research and Teaching
- 10:45-11:00 Break
- 11:00-12:00 Plenary Speaker  
*Baolin Zhang (张宝林), Beijing Language and Culture University*  
为汉语课堂教学服务—汉语中介语语料库建设与发展的新思路
- 12:00-1:30 Lunch at Sammy's in the Rice Memorial Center  
Use the lunch coupons provided

1:30-3:00

Panel A (RH123)

Session Chair: Marshall McArthur,  
University of Houston

“The Taiwanese Accent of Mandarin:  
A Corpus Analysis of Marking  
Influence on Accented L2”

*YUCHAU HSIAO*  
*National Chengchi University*

“Are Mandarin Tones Optimized for  
Efficient Communication?”

*YUAN-LU CHEN*  
*University of Arizona*

Panel B (RH113)

Session Chair: Edie Furniss,  
University of Houston

“Using Chinese Spoken Corpus to  
Teach Chinese for Medical Purposes”

*LIANG FU(傅亮)*  
*Rice University*

“The Effects of Learners’ Anxiety on  
Oral Proficiency of Foreign Learners  
of Chinese”

*LIJIE QIN*  
*Brandeis University*

CENTER FOR LANGUAGES & INTERCULTURAL COMMUNICATION PRESENTS

**STUDY  
ABROAD  
CONFERENCE**

**RICE UNIVERSITY FEBRUARY 10-11, 2017**

# CONFERENCE SCHEDULE

3:15-4:45

## Panel C (RH123)

Session Chair: Weidong Shi, Texas A&M University

“Listener Responses as Stance Display: Reactive expressions and their conversational uses in daily Mandarin conversation”

*I-NI TSAI*

*National Taiwan University*

“来华留学生预科汉语拼音学习研究”

*MIN WEI(魏敏)*

*KAIWEN YANG(杨恺雯)*

*Wuhan University of Technology*

## Panel D (RH113)

Session Chair: Sarah Tsai, University of St. Thomas

“基於語料庫探討量詞「支」的泛化使用現象”

*YUMING CHEN(陳玉明) National Taichung University of Education*

“浅议汉语口语语料库的语料选取问题”

*SHAN HE(何珊)*

*Beijing Normal University, Iowa University*

*CHEN MENGtian(陈梦恬)*

*Iowa University*

“Integrating revised Bloom’s taxonomy with authentic materials to improve teaching, learning, and assessing in CFL contexts”

*JING YU(俞菁)*

*University of California- Santa Barbara*

4:45-7:00

Welcome Reception with Vietnamese Cuisine

# CONFERENCE SCHEDULE

## THURSDAY, AUGUST 11<sup>TH</sup> – Rayzor Hall 123

8:30-8:50 Breakfast in Rayzor Hall Lobby

8:50-9:30 Plenary Speaker

*Aimin Cheng, Nanjing University*

“南京大学学习者口语语料库建设与使用”

9:30-11:00 Panel E

Session Chair: Aiming Cheng, Nanjing University

“A Database of Native Mandarin Speakers’ Lexical and Sublexical Frequency Judgements”

*SETH WEINER*

*Carnegie Mellon University*

“The Embodiment of UNDERSTANDING and Its Implication in Teaching Chinese as a Second Language”

*ZI-YU LIN*

*Macao Polytechnic Institute*

“Building spoken interlanguage corpora from online interaction for L2 Chinese teaching and learning”

*CHIN-CHIN TSENG(曾金金)*

*Taiwan Normal University*

11:00-11:10 Break

11:10-12:00 Plenary Speaker

*Meng Yeh (葉萌), Rice University*

“Teaching Interactional Practices through Building Corpora of Spoken Languages”

12:00 Lunch at Salento in Brochstein Pavilion

Use the lunch coupons provided

# PLENARY SPEAKERS



XIANWEN CAO (Nanjing University Institute for International Students)

## 汉语二语习得研究与学习者口语语料库建设

Professor Cao's research focuses on applied linguistics and Chinese as a Second Language. He has built the "Nanjing University Oral Corpus of Chinese Interlanguage." Currently, he is working on the project "Variability in Chinese as L2 Development: A Dynamic Systems Perspective" supported by the National Social Science Foundation of the PRC.



HONGYIN TAO (UCLA, Asian Languages and Cultures Department)

## Exploiting Spoken Corpora for Chinese Language Research and Teaching

Professor Tao specializes in discourse analysis, applied linguistics, and corpus linguistics. He has developed the UCLA Corpus of Written Chinese and is co-directing the US component of the International Corpus of English. His projects on advanced spoken Chinese & academic Chinese language research/teaching are funded by the US Department of Education.



MENG YEH (Rice University, CLIC)

### Teaching Interactional Practices through Building Corpora of Spoken Languages

Professor Yeh's current research focuses on language pedagogy and curriculum design to develop learners' interactional competence. Present projects include: "Corpus-based Pedagogy: Teaching and Learning the Temporal Connective JIU" and "Developing and Assessing Interactional Competence: Beginning Level."



BAOLIN ZHANG (Beijing Language and Culture University, Faculty of Linguistic Science)

### 为汉语课堂教学服务——汉语中介语语料库建设与发展新思路

Professor Zhang specializes in applied linguistics and language policy/planning. He has worked on the project "Corpus of Chinese Interlanguage: A Global Perspective" funded by the Ministry of Education of the PRC. At present, he is in charge of a BLUC research project "Policies on Building Language Database" in various countries.

# ABSTRACTS

## **Exploiting Spoken Corpora for Chinese Language Research and Teaching**

Plenary Speaker

Hongyin Tao

University of California, Los Angeles  
[tao@humnet.ucla.edu](mailto:tao@humnet.ucla.edu)

Audio-video-recorded spoken corpora, due to the rich information they contain, are widely being exploited for research and teaching. Yet questions remain as to how spoken corpora can be constructed and exploited for specific goals, especially for language teaching. In this talk, I use examples from a number of spoken corpora being developed at UCLA and elsewhere to discuss ways in which these data can be utilized for discourse analysis and for Chinese language education. Emphases will be placed on theoretical issues that corpus data raise and ways in which spoken corpora can be used to develop teaching materials for both classroom use and self-study in the context of Chinese as a second/foreign language.

## **Teaching Interactional Practices for beginning students through Building Corpora of Spoken Languages**

Plenary Speaker

Meng Yeh

Rice University  
[myeh@rice.edu](mailto:myeh@rice.edu)

Corpora of naturally occurring conversation provides rich and necessary material for developing L2 learners' interactional competence (IC). IC is the ability to employ mechanisms (such as turn-taking, repairing, or sequence organization) to jointly accomplish recognizable actions through the sequential, turn-by-turn unfolding of the interaction. However, textbook dialogues are artificially constructed to highlight language features and tend not to reflect the interactional architecture of ordinary conversation. In this presentation, I will first demonstrate the corpora of natural spoken language built at the Center of Languages & Intercultural Communication at Rice University. Second, I will present how to use naturally occurring conversations as teaching material with a pedagogical approach that fosters the development of beginners' IC. What I will describe is part of a research project which compares two groups of students enrolled in the first-semester Chinese courses; only the students in the experimental group uses natural conversation and received IC instruction. The data reveals that the students in the experimental group, when engaging in conversations with native speakers, enacted the interactional practices that were the target of instruction, thereby proving the effectiveness of IC-based pedagogy.

## 为汉语课堂教学服务——汉语中介语语料库建设发展的新思路

Plenary Speaker

Baolin Zhang (张宝林)

Beijing Language and Culture University  
[zhangbl@blcu.edu.cn](mailto:zhangbl@blcu.edu.cn)

近 20 年来，汉语中介语语料库建设的持续发展，及其在对外汉语教学研究中，特别是汉语习得研究中所起到的巨大作用，引起了海内外汉语学界的广泛关注，基于语料库的对外汉语教学研究和汉语中介语研究取得了很大成就。这种研究以定量分析和客观性为其显著特征，结论令人信服，大大提高了对外汉语教学研究的水平。因而语料库及其相关研究得到了学界的认同与欢迎。另一方面，汉语中介语语料库建设仍处于初级阶段，建设水平还不高，语料库的本体研究还十分缺乏，还不能完全满足汉语教学与研究的多方面需求。同时，基于语料库的字、词、语法研究固然对教学有促进作用，但这种作用是间接的，需要教师们在学习、理解相关的研究成果之后，才能落实到教学当中。从目前的现实情况看，汉语中介语语料库并没有对课堂教学产生直接的促进作用，与课堂教学尚处于游离状态。

其实，语料库与课堂教学的结合早已有之，即数据驱动的学习（data-driven learning，缩写 DDL）。这是一种基于语料库的计算机辅助教学形式，它为外语/第二语言学习提供真实的语言环境，强调探索和发现的学习过程，重视培养学生的自主学习能力。而课堂数据驱动学习(hands-off DDL)则将语料库资源引入常规课堂，它一方面保留了数据驱动学习的特点，即学习者基于语料库资源进行发现式学习；又把大量原始语料的整理工作交由教师完成，从而降低了对学习者语言水平的要求；加之采用纸质媒介进行教学，摆脱了对计算机的直接依赖，使语料库资源的课堂应用成为可能。这种教学方式本世纪初即在中国大陆的英语教学领域加以引进、介绍并进行教学实验，但在对外汉语教学界却尚无尝试。这需要汉语教师们了解新的语言教学理论，学习新的教学模式与方法，并积极地在自己的教学实践中进行尝试。

另一个把语料库用于教学的机遇是翻转课堂（the Flipped Classroom）模式。所谓“翻转课堂”模式是指将传统的“课堂教学，课下练习”转变为“课下学习，课堂练习”。其显著效益是可以妥善的解决高级汉语口语教学面临的四对矛盾中的三对，即(1)有限课时与人均口语练习时间之间的矛盾；(2)教师讲解与学生口语操练量之间的矛盾；(3)教学进度与学生个体差异之间的矛盾。采用这种教学模式，教师要给学生提供合适的课前学习资源。而这种资源不仅指教材、课文和生词的 MP3 录音、教师对重点词句的讲解视频等学习资料，也包括语料库，包括母语者的口语语料库和中介语口语语料库：前者可以使学习者了解并模仿学习正确的发音，后者可以使他们认识与避免错误的发音。

语料库还可以从其他方面为教学提供有益的服务。例如基于语料库研发不同母语背景的汉语学习者使用汉语汉字的汉字频率表、口语词和书面语词对照表、语法项目表，口语话题表、以及字、词、语法、语音、语调、停顿的偏误表；附加在语料库系统中的、帮助教师备课的教案辅助生成器，帮助教师考试命题的试题辅助生成系统、作文自动批改器等。

从语料库建设的角度看，确实尚未能满足汉语教学与相关研究的需求。例如口语语料库，目前虽然有几个宣称已建成并可以使用的口语语料库，但由于其并不开放，情况究竟如何，完全无法证明。唯一开放的汉语口语中介语语料库是暨南大学华文学院的外国留学生语料库，但它只是一个生语料库，文字转写没有体现外国人学汉语的口语特征，例如不正常的停顿、声韵调的偏误等都被略去了。因而无法使学习者通过它了解和避免错误的发音。

显而易见，教学对语料库是有很高的要求的。语料库要想更好地发挥自己的作用，为教学服务，必须了解教学，贴近教学，积极主动、全心全意地为教学服务；并在为教学服务的过程中不断完善自己，提高自己的建设水平和服务能力。

## 汉语二语习得研究与学习者口语语料库建设

Plenary Speaker

Xianwen Cao (曹贤文)

Nanjing Univeristy

[xcao@nju.edu.cn](mailto:xcao@nju.edu.cn)

(Unable to attend the Conference)

尽管在实际生活中人们说的比写的更多，但由于建设口语语料库代价更高，如，花费更多的资金和时间，需要更加复杂的处理技术等，结果是语料库口语数据大大少于书面语数据。据统计，在世界上 140 个学习者语料库中，只有 38 个与口语语料有关 (Granger et al. 2015)。即使在大规模、综合型的学习者语料库中，口语语料也只占到很小的比重，例如，英国国家语料库 (British National Corpus) 中只有 10% 是口语语料。

在屈指可数的汉语中介语语料库中，口语语料更是凤毛麟角。近年来，汉语学习者口语语料库建设日益受到重视，除了在汉语教学和语言处理等方面的现实需要以外，汉语二语习得研究也对建设高质量的学习者口语语料库提出了迫切需求。新型学习者口语语料库对于促进汉语二语习得研究可以起到诸多促进作用，本文暂列几项如下：第一，二十多年的汉语二语习得研究已经取得了不少研究成果，例如，大量成果涉及到汉语二语发展模式（习得顺序和发展序列），这些成果绝大部分是基于书面语语料的研究，其结论既有一致性的部分也有互相冲突之处，迫切需要采用大规模语料，特别是实时口语语料加以重新检验。第二，已有研究大都建立在单一的横向跨层语料的静态分析之上，缺少不同学习者、不同语言环境、不同任务条件、不同产出方式等能够反映各种语言使用变体的语料数据，因此，得出的结论通常只是某种固定的静态习得模式，无法反映二语发展中丰富复杂的现象和规律，例如，汉语学习者语言中的变异性和二语非线性发展规律等长期被严重忽视。第三，新的二语习得理论、分析框架和研究方法的不断发展，需要建设新型汉语学习者口语语料库作为基础平台以支撑相关研究，例如，动态系统理论、多维语言表现分析框架、以及“移动极值图表”、“相变”等分析方法，均需要精致的纵向语料库才能支撑多维动态分析。

除了对基于语料库的汉语二语习得研究进行总结梳理以外，本文也打算探讨汉语二语习得研究和学习者口语语料库研究之间的互动关系。一方面，讨论学习者语料库对于深入揭示二语发展规律的作用，详细分析如何利用学习者口语语料库来促进二语习得研究；另一方面，讨论二语习得理论对学习者的口语语料库设计和分析的重要性，即我们可以建设什么样的学习者语料库来解答当前二语习得理论需要回答的核心问题。在语料库语言学研究中，有“语料库指引” (corpus-informed)、 “基于语料库” (corpus-based)、 “语料库驱动” (corpus-driven) 等研究路径，这是一种从“供给侧”视角的分类。本文将从当前二语发展研究的“需求侧”视角，结合南京大学汉语学习者口语语料库的建设实践，讨论需求驱动的学习者口语语料库建设的相关问题。

## Are Mandarin Tones Optimized for Efficient Communication?

Yuan-Lu Chen

University of Arizona  
[cheny@email.arizona.edu](mailto:cheny@email.arizona.edu)

The lexical system of human languages evolves for the optimization of communication pressure: frequent words tend to be shorter (Sigurd et. al. 2004, Piantadosi et al. 2011). Specifically, Piantadosi et al. (2011) argue that cross-linguistically words with less average information content tend to be shorter. For Mandarin Chinese, segmental length is not a possible dimension to accommodate the amount of information content, because most of the morphemes in MC are single syllable with phonotactic restrictions (i.e. no complex onset and only sonorant codas allowed); instead, the super-segmental dimension (i.e. tone) is used. The current paper investigates how the tone system responds to communication pressure. It is expected that the complexity of tones of MC should be optimized just like the length of lexicon of non-tonal languages (Sigurd et. al. 2004). The complexity of tones is defined using the acquisition order found in Li and Thompson (1977), which argue that Tone1 and Tone4 are first acquired and later Tone2 and Tone3. I hypothesized that Tone2 and Tone3 are complex while Tone1 and Tone4 are simple. Along with the assumption of the optimization, the working hypotheses is in (1):

- (1) Tone2 and Tone3 are less frequent and have low information content. Tone1 and Tone4 are more frequent and have high information content.

A corpus (Glenn and Linguistic Data Consortium 2013) is used to test the hypotheses, which is a news corpus transcribed using MC characters. The characters are converted into tone-annotated pinyin using Pypinyin (2014, January 1). Two factors are focused: a word's frequency and its average information content, which is defined in Piantadosi et al. (2011):

$$(2) \quad -\frac{1}{N} \sum_{i=1}^N \log P(W = w | C = c_i) \quad (w \text{ is the target words, and } c \text{ is the context.})$$

Assuming a bigram-model, if a word is totally predicated by the preceding word of it, the word will not have any information. On the other hand, if the target word is a surprise in that context, the word carries high information content. The finding is NOT consistent with the working hypothesis. The current study found that in MC, Tone1 has the most information content and least frequent; Tone4 has least information content and most frequent. Tone3 and Tone2 are in the middle (log-freq~ tone, ANOVA (F(3,1126) = 6.99, p <.001; log freq~ tone, ANOVA (F(3,1126) = 6.99, p <.001). To see how communication pressure effects the evolution of tonal lexicon, it is required to define the complexity/cost of each tone. This is open for future research.

## 基於語料庫探討量詞「支」的泛化使用現象

Yuming Chen (陳玉明)

National Taipei University of Education

[Jteacher.yam@gmail.com](mailto:Jteacher.yam@gmail.com)

漢語中的量詞有分類的功能，特別是個體量詞中的形狀量詞。搭配量詞「張」、「條」、「顆」的名詞，明顯說明是不同形狀的物體，此規律性使形狀量詞成為外籍漢語學習者最容易掌握的一種量詞。

然而近年來，研究者觀察到母語人士越來越常使用量詞「支」，例如：「一支遙控器」、「一支手機」、「一支蓮蓬頭」、「一支鑰匙」，包括鄧守信教授最新主編出版的教材——當代中文課程第一冊，也已收錄「一支手機」，這令研究者在教學時有分類上的困難，也推測量詞「支」已逐漸泛化使用了。那麼，泛化使用的範圍為何、是否受方言影響、是否與社會發展、新事物增多有關，成為亟須了解的問題，以便面對學生可能的疑問。

研究者查考相關現代漢語語料庫，如：中研院「現代漢語語料庫一詞泛讀」、政治大學「漢語口語語料庫」，初步得到：許多桿狀物或有把手的物品已泛化使用量詞「支」了，特別是現代的新物品由於沒有過去的量詞規範，便大量使用量詞「支」，致使「支」、「把」、「根」等量詞使用更難以歸類，是個值得漢語教學界重視的現象。

## Using Chinese Spoken Corpus to Teach Chinese for Medical Purposes

Liang Fu (傅亮)

Rice University  
[Liang.fu@rice.edu](mailto:Liang.fu@rice.edu)

The use of corpora is becoming increasingly important in language classes to promote active learning of the language. It is even more so for developing courses for Specific Purposes. While many written corpora have been built and are readily accessible for these courses, spoken corpora of spontaneous conversations for these purposes are not easy to find. This paper will present how authentic Chinese spoken data is used to develop class activities and materials for a Chinese for Medical Purposes course. It will first explore the need to collect the spoken data as the textbooks and other course materials fall short to meet the goals of developing the interactional competence of the learners. While textbooks provide good resources for learners to study written and formal forms of language, they lack the interactional features of the language. By exploring the authentic spoken data in medical settings, learners are able to gain understanding of the linguistic and socio-cultural strategies that are used in conversations and learn to adopt them when similar life situations arise. It then will showcase a Chinese spoken corpus that has been built to provide materials for data-driven learning and teaching. The spoken corpus is accessible online. It has been transcribed and categorized by topics, interactional and socio-linguistic features, and so on. Lastly, it will demonstrate how these spoken data are used to develop lesson plans which teach not only the linguistic, but also the socio-cultural aspects of the language. The corpus-based lesson plans are developed in the way that they allow students to conduct conversation analyses, test hypotheses on language rules and uses, and understand the interactional strategies used by speakers of different ages, genders, occupations and social statuses. These lesson plans also enable students to develop awareness of the many cultural factors that are displayed throughout the conversations such as the constant concern for face saving of the Chinese speakers, the quietness of the Chinese listeners when receiving information, etc.

## 浅议汉语口语语料库的语料选取问题

Shan He (何姗)  
Beijing Normal University, University of Iowa  
[shan-he@uiowa.edu](mailto:shan-he@uiowa.edu)

Mengtian Chen (陈梦恬)  
University of Iowa  
[mengtian-chen@uiowa.edu](mailto:mengtian-chen@uiowa.edu)

了解语域分布情况和语体色彩是汉语学习者成功掌握汉语，正确且得体地使用汉语的关键之一，也是汉语作为第二语言教学中的重点和难点之一。通过中介语语料库可以对汉语学习者的语法词汇偏误、语用失误进行分析，但只有依靠口语语料库才能对学习者的语音面貌和语言动态使用情况进行较为全面的分析。对汉语学习者的口语进行分析必须要以现代汉语口语的充分研究为基础，但目前相关研究还比较缺乏，可供使用的大型现代汉语口语语料库也亟待建设。

建设现代汉语口语语料库对于现代汉语口语研究有极大帮助，但首先要解决的问题是：“现代汉语普通话”的操作定义是什么？电台、电视台的主持人的普通话发音标准，但媒体语料不是真实的口语；其余大部分汉语母语者的普通话又或多或少都带有一些方言的特点，如前后鼻音不分、鼻音边音不分、轻声等发音问题，不同地区的居民在用词用语上也有地区性特征，这些特征是否应该在语料库中显示出来？不够标准的普通话能不能作为语料？

我们建议广泛选取具有代表性的语料，一方面能使口语词汇、用法的标注更为准确，另一方面也能将这些语料作为教学材料，让海外学习者有机会接触到真实自然的汉语口语语料，有助于提高学习者的听说交际能力。

## The Taiwanese Accent of Mandarin:

### A Corpus Analysis of Marking Influences on Accented L2

Yuchau Hsiao

National Chengchi University

[ychsiao@nccu.edu.tw](mailto:ychsiao@nccu.edu.tw)

This paper addresses the accent arising from Taiwanese speakers' acquisition of Mandarin. In this research, I established a corpus containing 1260 Mandarin words produced by Taiwanese speakers. The data were collected from eight sessions of Mandarin conversation with four Taiwanese speakers. Each session lasted about one hour, with a total of eight hours recorded. The Taiwanese speakers included two males and two females, who were aged between 71 and 73, and are of primary education. As education was not commonly available to people in early years, most of them received only primary education or below. The informants thus represent the typical "old Taiwanese." Six accented patterns are observed from the corpus. First, the apical vowels, [ɿ] and [ʅ], are rounded as the high back rounded vowel, [u]. Second, the high front rounded vowel, [y], and glide, [ɥ], are derounded as the high front unrounded vowel, [i], and glide, [j]. Third, the vocalic sequences, [ei] and [je], are monophthongized as [e], and [ou] and [wo] as [o]. Fourth, nasal vowels are denasalized before nasal codas. Fifth, the retroflex onsets, [tʂ], [tʂh], [ʂ] and [ʐ], are deretroflexed as the dentals, [ts], [tsh], [s] and [z], and the retroflex coda, [r], is deleted. Finally, the labiodental fricative [f] is either reconfigured as [hw] or delabialized as [h].

Accents are known as different manners of pronunciation of a language, which basically fall in two domains. One domain of accent is defined among native speakers of various regions, or from diverse social classes, etc. People from Texas would speak with an accent from the point of view of a New Yorker, and vice versa. The other domain of accent refers to the production of nonnative speakers, commonly characterized as a foreign accent. A Frenchman may speak English with a French accent. In this case, the accented English is influenced by the sounds of French, and the emerging phonological system is an interlanguage, which is interim and often idiosyncratic. The Taiwanese accent of Mandarin in question is somewhere between these two domains; it involves systematic adjustments of segments and syllable structures. In fact, phonological influences in language acquisition inevitably involve systematic adjustments of segments and syllable structures. Such adjustments often result in the emergence of the universally unmarked. This paper observes that some marked segments and syllable structures of L2 are omitted in the accented L2, where unmarked forms occur as substitutes. Cognate correlation may somehow affect the accented forms, but it is subsumed under the marking relation of segments and syllables; precisely, if  $\alpha$  is the cognate of  $\beta$ , and  $\alpha$  is more marked than  $\beta$ ,  $\alpha$  would be interpreted as  $\beta$  in the accented version, but the reverse interpretation would be infeasible. I propose a marking-based model of accent formation and argue that universal marking is the key to the appearance of accented forms.

## The Embodiment of UNDERSTANDING and Its Implications in

### Teaching Chinese as a Second Language

Zi-yu Lin

Macao Polytechnic Institute

[zylin@ipm.edu.mo](mailto:zylin@ipm.edu.mo)

This presentation is part of a larger research based on the utilization of diachronic and synchronic corpora and open-source dictionaries, such as *Chinese Text Project*, *Chinese-English Sentence Aligned Bilingual Corpus*, and *CC-CEDICT*, in an effort to address issues on how abstract concepts are expressed within and across languages and why it happens the way that has been. An interesting example is the abstract meaning of UNDERSTANDING in English and Chinese. Data from the corpora unmistakably demonstrate that it is more than often conveyed by terms that are related to human embodiment experiences such as interaction with physical light or physical disentanglement. For the former case, in fact, about 20% of the Chinese equivalents of the English word "understand" is rendered into 明白 ("bright" + "while light"). Moreover, semantic connections between LIGHT and UNDERSTANDING can also be seen through the Chinese translations of the English term "clear" which too is related to LIGHT. A key Chinese counterpart of English "clear" is again 明白.

Now, in the light of the recent developments in cognitive and neural linguistics, which have successfully explained the neural circuitry mechanisms for metaphors like "THINKING IS SEEING," and further "UNDERSTANDING IS SEEING," the author argues for a chain of metonymical relations between LIGHT and UNDERSTANDING, as follows:

LIGHT → SEEING → THINKING → UNDERSTANDING

The relationship between LIGHT and SEEING is a PART AND WHOLE metonymical relationship, because under normal circumstances, light is part of the conditions for a naked eye to be able to see. Then, in the LIGHT → SEEING metonymic development, the frame F is the vision frame, since 明 is an inalienable condition of 看 (SEEING), 明 then is able to take on the function of 看, as in, for instance, 失明 (BLIND), 复明 (GAINING EYESIGHT). In this way, 明, together with 白, has eventually completed its journey to become a word that means to understand.

Observations from the semantic and etymological analysis above can be readily applied to a Chinese language class to answer questions about why Chinese lexical items that originally mean LIGHT are used to mean UNDERSTANDING. For Chinese learners, they are curious about the semantic contents of each character that appears individually or in combination with others. As a matter of fact, the verb 明白 in Chinese is not a loner. In English, for instance, other words like "brilliant," and "bright," are linked to physical light to different degrees and can also mean "understanding" and "intelligent." This shows that because we humans have similar physiological characters and live on the same planet, we share a broad cognitive basis in our languages and thought, which constitutes the foundation of communication across languages and 2nd language acquisition. The data from corpora simply lend us more compelling evidence to make such an argument.

## **The effects of learners' anxiety on oral proficiency of foreign learners of Chinese**

Lijie Qin

Nanjing University, Brandeis University  
[lqin@brandeis.edu](mailto:lqin@brandeis.edu)

Oral proficiency of foreign learners of Chinese is measured against four standards: accuracy, fluency, complexity, diversity. My research focuses on how learners' anxiety affects one or more aspects of the above named oral proficiency markers, and how each marker is affected by a certain type of learners' anxiety. From information collected from the Foreign Language Classroom Anxiety Scale survey and from the Personal Information Questionnaire, I can cross-reference the results from these against the oral corpus of Institute for International Students, Nanjing University. Of the 270 questionnaires collected from foreign students at Nanjing University, only 73 matched directly with the university's corpus, however all results were used in collating data for the research study. The data collected from the surveys are processed using the softwares SPSS and STATA, and Hycon was used to transcribe and analyze recordings from the Nanjing University corpus. My results show that there is a significant difference in fluency of oral proficiency between high-level anxiety students and low-level anxiety students. My research fills the gap between research on anxiety and on oral proficiency.

**Building spoken interlanguage corpora  
from online interaction for L2 Chinese teaching and learning**

Chin-Chin Tseng (曾金金)

National Taiwan Normal University

[tseng@ntnu.edu.tw](mailto:tseng@ntnu.edu.tw)

The present study investigates how online interaction video recordings of Chinese L2 learners can be collected and analyzed to develop a spoken interlanguage corpora for pedagogical applications. Selinker (1992) summarised that interlanguage study can never be replaced by other means of language analyses due to the very own nature of its subject: interlanguage is itself a structure with an inherent system that is constantly in the dynamic process of interacting and changing with learner's language background, phase of acquisition, teaching methods, and other non-linguistic factors. In this study, interlanguage analysis of spoken Mandarin Chinese by American English speaking learners is conducted. 165 video clips were collected from online interactions between practice teachers of Chinese as a Second Language (CSL) and CSL learners in University of Texas at Arlington (UTA). Common spoken Mandarin error patterns can be identified, such as the 'yu' sound in 面具、捐, 'qu/chu' distinction; 'fang/feng' distinction, phrase-initial /ts/ sound with a falling tone tend to become aspirated, tone combination 3+1→\*2+1, 可以 becomes\*克以, 不可以 becomes 不\*科以, 它很薄 becomes 它\*恨\*保, 臥室 becomes '\*wo2+shi3', 廚房在餐廳左邊→\*做邊, word order, function words, such as \*/過; 是/有; 不/沒 etc.

Interlanguage is best comprehended and processed in the interaction, on the account of dynamic factors for optimum evaluation and denotation. The use of video clips allows researchers to have a more comprehensive understanding of how interlanguage functions in context, supplementing learners' input from the practice teacher, facial expressions, and other visual teaching materials and contextual details that are involved in the interaction, which are crucial when conducting interlanguage analyses.

The model of corpus implementation proposed in this study has the following advantages:

- 1). The interlanguage data were collected in a meaningful interactional setting, which adds discourse context to the interlanguage data.
- 2). It is in the video format, so we can use visual cues to assist clarifying what the learner is trying to express.
- 3.) Learners' difficulties, interlanguage characteristics, and developmental sequences can be analyzed in this dynamic speech corpora. Longitudinal interlanguage data were also collected in this study. Some errors are corrected after one year's Chinese learning and practice. However, some errors are highly resistant to change, such as stressed syllables tend to be pronounced as a falling tone, sound like the 4th tone in Mandarin Chinese, and uncertainty rising intonation tends to cause lexical tone change. For example, 籃球 becomes 籃秋?, 'wang3+qiu2' becomes 'weng' (mid level tone) + 'qiu' (high level tone)?.
- 4). The database is subject to interlanguage analysis for pedagogical purposes. By systematic documenting learners' second language errors ecology, it will not only contribute to defining benchmark of each proficiency level, but also help L2 language teaching to be more effective.

## Listener responses as stance display:

### Reactive expressions and their conversational uses in daily Mandarin conversation

I-Ni Tsai

National Taiwan University

[initsai@ntu.edu.tw](mailto:initsai@ntu.edu.tw)

Conversation is a highly interactive and collaboratively co-constructed activity. Listeners contribute as much as speakers to ongoing conversations. As small as listener response tokens may seem (*mm* or *mhm*), listeners display their understanding of or stances toward the message they have just received and help to shape the trajectory of the ongoing conversation (Drummond & Hopper, 1993; Gardner, 2001; C. Goodwin, 1986; Jefferson, 1984; Schegloff, 1981; McCarthy, 2002, 2003).

Studies on listener responses in Mandarin Chinese are relatively scarce (Chui, 2002; Clancy et al., 1996; Deng, 2008; Tao & Thompson, 1991; Wang et al., 2010; Wu 2001). Although Tao and Thompson (1991) based their study on Chinese conversations, they dealt with English backchannels. Clancy et al. (1996) and Deng (2008) presented a general picture of listener responses in Mandarin conversations. However, these two studies emphasized cross-linguistic comparisons of overall numbers and the types of tokens being used, leaving the functions of each single response token unexplored. While Wu (2001) provided an overview of distribution and functions of listener feedback, Wang et al. (2010) narrowed down to focus particularly on *hao* and *dui* in combination with final particles (*a*, *le*, *la* etc.) and explored interactional functions of these lexical bundles.

Based on daily conversational data, this study aims to investigate listener responses, with a focus on reactive expressions, in casual conversations in Taiwan Mandarin. Reactive expressions are defined as non-minimal turns with two-syllable content words, lexical bundles or short phrases; they are so defined to be distinguished from the minimal, mostly non-lexical, reactive tokens and backchannels. The purpose of this study is twofold: (a) to provide an overview of frequent reactive expressions used in Taiwan Mandarin and (b) to analyze the stances these tokens/ expressions are used to display.

The data examined in the present study come from an online spoken corpus, “The NCCU Corpus of Spoken Chinese.” (The NCCU Corpus of Spoken Chinese <http://140.119.172.200>) Ten 20- to 30-minute clips were selected for examination and they ran for approximately four hours in total. These interactions are mostly daily, informal conversations. All the participants are native speakers of Mandarin Chinese from Taiwan.

The findings have pedagogical implications. It is argued that displaying features of spoken language, such as natural listener responses, in the teaching material makes it easier for learners to develop natural conversational and communicative skills and textbook dialogues should closely resemble authentic spoken language (McCarthy, 2002, 2003; McCarthy & Carter, 1994; Ur, 1984). This study can be seen as part of a growing effort to underscore the importance of integrating findings of the spoken language into communicative teaching projects (Hatch, 1992; McCarthy & Carter, 1994; Tao, 2005).

# 来华留学生预科汉语拼音学习研究

Min Wei (魏敏)

Wuhan University of Science and Technology  
[weixiaomin020@sina.com](mailto:weixiaomin020@sina.com)

Kaiwen Yang (杨恺雯)

Wuhan University of Technology  
[421501583@qq.com](mailto:421501583@qq.com)

**摘要:** 本文在参考前人有关问卷基础上, 自编问卷, 考察了预科留学生汉语拼音学习的研究, 受试者为华中师范大学 2014 级的 132 名预科留学生。通过均值及相关数据分析, 考察了拼音应学习多长时间、拼音学习的方式(是语流, 还是音素)、拼音学习的难点音及拼音的功能作用(是目的, 还是手段)。欧美组学生、非洲学生和亚洲组学生对拼音的学习呈现出不同的特点。亚洲学生语音学习在声母和韵母上都存在困难, 主要是很多“同形异音”所带来的困难, 其难点音主要是翘舌音。非洲学生主要在声调学习上存在较大困难, 平舌音相较亚洲和欧美学生有较大困难, 边鼻音偶有反映(均值接近“2.6”), 舌面音、前后鼻音不构成其发音难点。欧美学生主要集中在翘舌音和后鼻音, 而边鼻音、舌面音和前鼻音不构成其发音难点。

另外, 拼音教学应更具针对性。比如是以语流教学, 还是音素教学。根据调查显示, 亚洲学生更倾向于语流教学, 非洲和欧美学生更倾向于音素教学, 因此, 教学中就应注意这些倾向性。

在此基础上, 文章对汉语拼音教学提出了几点建议。

## A Database of Native Mandarin Speakers' Lexical and Sublexical Frequency Judgments

Seth Weiner

Carnegie Mellon University  
[sethw1@cmu.edu](mailto:sethw1@cmu.edu)

During spoken word recognition, listeners draw on previously learned information about the lexicon of their language, including each word's frequency of occurrence. High frequency words hold a privileged status: these words tend to be activated faster, activated from less acoustic input, and often probabilistically predicted in upcoming speech (Cutler, 2012). Native Mandarin speakers, for example, track and store lexical frequency information, as well as sublexical information (e.g., syllable frequency and syllable-conditioned tonal probabilities), the latter of which has been shown to affect the online processing of Mandarin speech (author's work). Because frequency effects are prevalent even in a second language (L2) (Ellis, 2002), exposure to L2 statistical information may help bootstrap early acquisition. Yet, a challenge adult L2 pedagogy faces is ensuring learners are exposed to highly frequent and probable sounds, while simultaneously teaching beneficial and practical vocabulary. Finding this balance requires a reliable source of Mandarin lexical and sublexical statistical information. The present study aims to provide such a source.

A crowd-sourced Mandarin database was constructed using native speakers' frequency judgments. Native speakers accessed a secure online survey, which prompted two types of judgments. Participants were presented with randomized Mandarin syllables in pinyin (e.g., *wō*) and asked to report the most frequent spoken word that begins with that syllable, irrespective of tone (e.g., 我 'I/me'). This provided a lexical frequency measure derived from the syllable. Participants were also presented with different randomized syllables written with all four lexical tones and the neutral tone (e.g., *nī, ní, nǐ, nì, nǐ*) and asked to rank the five possible combinations from most frequent to least frequent (or nonexistent). This provided a sublexical measure of syllable-conditioned tonal probabilities. Results from nearly 200 speakers confirm that explicit judgments of frequency are largely compatible with naturally occurring spoken word frequency corpora, e.g., SUBTLEX-CH (Cai & Brysbaert, 2010), though important differences emerged among certain syllable types. Similarly, sublexical judgments corroborated previously reported syllable-conditioned tonal probabilities, though once again differences emerged. Taken together, and benefits researchers interested in controlling lexical and sublexical frequencies of experimental Mandarin stimuli. More importantly, these native speaker judgments can be incorporated into L2 pedagogy to ensure learners are exposed to useful vocabulary as well as the most frequent forms and sounds.

**Integrating revised Bloom’s taxonomy with authentic materials  
to improve teaching, learning, and assessing in CFL contexts**

Jing Yu (俞菁)

University of California—Santa Barbara  
[jing02@umail.ucsb.edu](mailto:jing02@umail.ucsb.edu)

Currently, when teaching Chinese as a foreign language (CFL) in U.S. contexts, a majority of teachers are accustomed to introducing authentic materials, such as news, Talk Shows, to enhance students’ communicative competence and cultivate their cross-cultural awareness. Students do well in acquiring these knowledge under the concrete situation that teachers build up for them in the classroom. However, once these language learners are put in the real environment, it’s very difficult for them to transfer what they have communicated and learned to practical life (i.e. talking with a native speaker about the news happened yesterday). This gap precludes students’ learning process and discourage their interests in learning Chinese. The reason why they cannot apply to practice, especially in speech, is probably they are short of two key stages in cognitive process of knowing a new thing. In other words, they haven’t internalized the knowledge by absorbing and digesting. According to the revised Bloom’s taxonomy (RBT), there are six hierarchical levels of thinking and learning. They respectively are “remember”, “understand”, “apply”, “analyze”, “evaluate”, and “create”. Most Chinese teachers spend large amounts of time designing curriculum by using authentic materials to make students attain the first three learning objectives. Therefore, through instruction and repetitive drills, students indeed apprehend and even substitute newly-learnt words in the sentence patterns to complete the dialogue within short memory. However, remembering, understanding, and applying all belong to lower order thinking skills (LOTS). If students want to “create” in daily conversation, only mastering the preliminary skills is not enough for them to produce outcomes by reorganizing elements into a new pattern or structure. Therefore, when integrating the authentic reading or listening materials, teachers should not only treat them (lyrics, advertisements) as supplementary drill materials, but also take advantage of them to promote students’ higher order thinking skills (HOTS) and spread Chinese perceptual system. RBT, as an effective measure tool, can be widely utilized in Chinese language teaching, including planning of curriculum, delivery of instruction, as well as assessment standard. In this paper, I will explicitly explain how authentic materials are fully used in align with revised Bloom’s taxonomy as educational objectives to improve teaching efficacy and foster students’ intercultural communicative competence.





# CLIC SUMMER WORKSHOPS

**FOR L2 TEACHERS**

**Friday & Saturday**

**August 12-13, 2016**



Introducing the new *Bridging Research and Practice*

*Professional Development Resources for L2 Teaching and Learning*



**Classes**  
Samples of teaching from CLIC



**Workshops**  
Online workshop modules for professional development



**Conferences**  
Abridged versions of CLIC conferences and symposia



**Language Data**  
Authentic language samples and activities to use in class

Visit us at <http://CLICmaterials.rice.edu>

**Registration Fee**

**\$200**

**Includes:**

Up to 2.0 Continuing Education Units (CEUs)

Lunch

CLIC Faculty Guidance

**Register Now!**

Go to [l2teachingseminars.blogs.rice.edu](http://l2teachingseminars.blogs.rice.edu) for schedule and registration information