

中国語チュートリアルシステムの国際共同開発

課題番号 14310136

平成 14 年度～平成 16 年度科学研究費補助金 基盤研究 (B)(2)

研究成果報告書

平成 17 年 5 月

研究代表者 砂岡和子

(早稲田大学・政治経済学術院・政治経済学部教授)

目次

研究の経緯		5
為提升口語能力的中文教學項目量化	砂岡和子 村上公一 劉松	11
Computerized Adaptive Testing (CAT) 方式的網路中文口語測驗的開發	村上公一 砂岡和子 劉松	17
與語料庫相結合的中文 CALL 練習系統的開發	劉松 砂岡和子 浦野義頼	23
與評獎和 Self-Editing 相結合進行的網絡華文寫作教學	砂岡和子 村上公一 劉松	30
日本的手機在中文教育上應用的問題與實踐	劉松 砂岡和子 浦野義頼	37
Mobile CALL-drill and Web-test Tools for the Communicative Chinese Conversation Lessons	Song LIU、Kazuko SUNAOKA、Yoshiyori URANO	44
日中台遠隔中国語口語レッスン管理プラットフォームの開發 -早稲田大学 Tutorial Chinese E-Learning Platform-	劉松 砂岡和子 浦野義頼	50
PC、携帯端末に対応するコールドリルシステムの開發	劉松 砂岡和子 浦野 義頼	55
汉语语料库的建设及其在汉语教学中的应用	俞士汶 段慧明 朱学锋	60
基於社會互動語言觀點的遠距教學原則	信世昌	71
合成音声利用による中国語声調の学習システム	比企静雄 砂岡和子 楊立明 今泉一哉	78

研究の経緯

北京大学および台湾師範大学と共同で、中国語口語能力に焦点をあてた教学プログラムの国際開発を行った。平成 14 年はインターネット経由による両校からの授業配信実験を行い、平成 15 年より授業管理プラットフォームの構築 (<http://www.tcwaseda.net/tc/>)、インストラクター養成法、テキスト共同編集、平成 16 年には Web 作文添削、自動クラス分け Web 試験の開発、授業と連動した Call ドリルなど教学プログラムの開発を行った。

研究は教育面でも目に見える成果を挙げ、平成 15 年より早稲田大学正規科目として、北京大学から 136 駒、台湾師範大学から 84 駒、計 220 駒の遠隔授業を配信し始め、平成 16 年度は、北京から 216 駒、台湾から 144 駒、計 360 駒を学生 250 名に配信完了した。

遠隔授業の他に、平成 15 年度前後期計 240 駒、16 年度前後期計 480 駒の対面授業も、平行して実施した。外部客観テストによる検証の結果、本受講生の語学力は、平成 15 年半期で平均 10%、16 年は 20% 向上、最高得点が 985 点（満点 1000 点）など、顕著な学習効果を挙げた。

研究課題期間中、国内外の研究者および教育者との教学研究の交流が活発となり、平成 14 年 12 月、北京大学計算言語学研究所、台湾師範大学・韓国高麗大学など国内外遠隔授業と研究者 40 名を招聘し、「中国語国際遠隔教育の実践と展望」を開催した。その成果は論文集にまとめられている。平成 15 年以降、中国語 corpus 研究会を 5 回主催した。平成 16 年度より早稲田大学会員企業と産学協同で教学配信プログラム・口語能力判定テスト・携帯電話端末対応 CALL ドリル・中国語声調自動矯正システムの開発を行った。

中国語口語能力プログラムの開発は、システム自体はほぼ課題を達成した。今後は、自動添削、高精度 Web 試験、インテリジェント Call ドリル、自動発音矯正プログラムなど、学習者の認知能力を生かし、自律的学習機能を兼ね備えた、各種教育アプリケーションソフトの開発に力をいれて研究を継続してゆく。

研究組織

- 研究代表者 : 砂岡和子 (早稲田大学・政治経済学術院・教授)
研究分担者 : 村上公一 (早稲田大学・教育総合科学学術院・教授)
研究分担者 : 楊立明 (早稲田大学・国際教養学術院・教授)
研究分担者 : 楊達 (早稲田大学・文学学術院・教授)
研究分担者 : 劉傑 (早稲田大学・社会科学総合学術院・教授)
研究分担者 : 青山瑠妙 (早稲田大学・教育総合科学学術院・助教授)
- 研究協力者 : 兪士汶 (北京大学計算語言学研究所・教授)
研究協力者 : 信世昌 (台湾師範大学華語教育研究所所長・教授)

交付決定額 (配分額)

(9,700 千円)

	直接経費	間接経費	合計 (円)
平成14年度	2,800,000	0	2,800,000
平成15年度	2,800,000	0	2,800,000
平成16年度	4,100,000	0	4,100,000
平成 年度			
平成 年度			
総計	9,700,000	0	9,700,000

研究発表

(1) 学会誌等

「国際ネットワーク型中国語教育 Chinese Online」『漢字文献情報処理』創刊号 2000 年 10 月好文出版
“Cyber Chinese language Learning by global computer network in Waseda University” The 2nd
International conference on the Teaching and Learning Chinese 論文選 2000 年 12 月

「国際ネットワーク型語学学習プログラムの授業導入—早稲田大学 Chinese Online 実践報告—」『
諸学教養』109 号 2000 年 12 月早稲田大学政治経済学部紀要

「発信・交流型中国語教育 Chinese Online」砂岡和子・村上公一・于洋 『語研フォーラム』17 号
2001 年 10 月早稲田大学語学研究所紀要

「Chinese Online2 年目課題と展望」『漢字文献情報処理』第 2 号 2001 年 10 月好文出版

「語彙&語法ヘルプ 機能付き中国語 CALL ドリル “ゴーツー(構詞)”」『諸学教養』111 号 2001 年 12 月
早稲田大学政治経済学部紀要

「日本の中国語教育—その現状と課題・2002—」、日本中国語学会中国語ソフトアカデミズム検討委員
会編所収、日本中国語学会発行、好文出版 2002 年 3 月

「中国語 Tutorial Method 開発と発話分析」早稲田大学政治経済学部紀要『諸学教養』112 号
2002 年 3 月

「早稲田大学漢語遠程教学的技術と理論」砂岡和子『E-Learning 与对外漢語教学』
清華大学出版社(中国)2002 年 8 月

「発信・交流型中国語教育 Chinese Online」砂岡和子, 村上公一, 于洋
早稲田大学語学研究所紀要『語研フォーラム』17 号 2002 年 10 月

「早稲田大学現代外語教育技術的理論和实践」(北京)『外国語言学及应用語学言』第 4 期
2002 年 12 月首都師範大学出版社

「コーパス利用による中国語教育(1) 中国語コーパス利用とデータ変換」砂岡・尹明共著
早稲田大学政治経済学部紀要『諸学教養』紀要 114 号 2003 年 3 月

「中国語チュートリアル教学とレッスン成果分析—2002 年度早稲田大学中国語チュートリアル実践報
告」砂岡和子、于洋早稲田大学語学研究所紀要『語研フォーラム』第 19 号、2003 年 10 月

「基于 Active Database 与師生課本互動的國際漢語口語教学」砂岡和子・村上公一・浦野義頼・劉松・
張悦・倪萌『数字化对外漢語教学理論与方法研究』(清華大学出版社) 2004 年 7 月

「遠程網絡漢語口語教学管理平台的開發」劉松・張悦・倪萌・砂岡和子・村上公一・浦野義頼『数字化
对外漢語教学理論与方法研究』(清華大学出版社) 2004 年 7 月

(2) 口頭発表

「連結北京・台湾・漢城的国際遠隔教育報告」

早稲田大学 CCDL 中国語国際遠隔教育シンポジウム 2002 年 12 月

「早稲田大学中国語遠程教学演習 Chinese Online」

“The 2nd International Conference on Internet Chinese Education” (ICICE 2001)

「Digitalization for Global Chinese Education」 Proceedings 2001 年 12 月

「中国語国際遠隔語学演習 Chinese Online」梅田雅子、砂岡和子、村上公一 2002PC

Conference 論文集、2002 年 8 月

「国際遠隔中国語チャットの実践と語学教育効果」、日本中国語学会年次大会論文集、

砂岡和子、村上公一、梅田雅子 2002 年 11 月日本中国語学会

「当前日本漢語教学对言語信息处理技術的需求」 「Current Demanding of Teaching Chinese in Japan on Technologies of Language Information Processing」 北京大学計算語言学研究所主催、

第二回中日自然言語处理専門家委員会 (CJNLP2002) Proceedings、2002 年 11 月北京大学

「中国語レベル分けのための Web 上での中国語能力診断テスト」 村上公一・砂岡和子・劉松 日本中国語学会第 53 回全国大会予稿集 2003 年 10 月

「以提升交際能力為目的之漢語教学以及水平測驗難議-早稲田大学国際漢語遠距教学四年実線見証」 砂岡和子、村上公一 The 3rd International Conference on Internet Chinese Education (ICICE 2003)

『Dialogue between Technology and Chinese Language & Culture』 2003 年 10 月

「多国語 Windows 作業系統下基于 Corpus 的中国語教学課件的開發」 The 3rd International Conference on Internet Chinese Education (ICICE 2003) 尹明, 砂岡和子, 成田誠之助

「赴華中文短期研習与利用網路環境的事前、事後学习」 村上公一・砂岡和子・李利津・于洋・梅田雅子 第七屆世界華語文教学研討會論文集 第 4 冊 2003 年 12 月

「网络汉语口语课本和词汇分级」 早稲田大学遠隔教育中国語部会 漢語コーパス研究会第 3 回報告会配布資料 2004 年 1 月

“Occurrence frequency and transition probability of the Chinese tones”, Shizuo Hiki, Kazuko Sunaoka, Liming Yang and Yasuyo Tokuhiko; Proceedings, International Symposium on Tonal Aspects of Languages: With Emphasis on Tonal Languages, TAL 2004, 28 - 30 March 2004, Beijing, China, pp.73 - 76 (Poster P06).

“Statistical properties of the Chinese four tones derived from the Grammatical Knowledge - Based of Contemporary Chinese”, Shizuo Hiki, Kazuko Sunaoka and Liming Yang; Lecture meeting at the Institute of Computational Linguistics, Beijing University, March 30, 2004

“Statistical properties of the Chinese four tones: Occurrence frequency and transition probability”, Shizuo Hiki, Liming Yang, Kazuko Sunaoka and Yasuyo Tokuhiko; Proceedings of the 18th International Congress on Acoustics, ICA2004, 4 - 9 April 2004, Kyoto, Japan, IV-3317 - 3320 (Poster, Th.P3)

「現代漢語語法信息詞典による中国語声調の統計的性質の解析」 比企静雄, 砂岡和子, 楊立明、漢語コーパス研究会第 4 回報告会、早稲田大学遠隔教育中国語部会 2004 年 4 月

「コーパスを用いた中国語口語テキストの詞彙提示」砂岡和子・伊明・劉松、漢語コーパス研究会第4回報告会配布資料、2004年4月早稲田大学遠隔教育中国語部会

「日中台遠隔中国語口語レッスン管理プラットフォームの開発」劉松・砂岡和子・浦野義頼 2004 PC Conference 研究大会予稿集 2004年8月

「合成音声利用による中国語声調の学習システム」比企静雄・砂岡和子・楊立明・今泉一哉 2004 PC Conference 研究大会予稿集 2004年8月『Dialogue between Technology and Chinese Language & Culture』2003年10月

「発話能力を測定するための中国語 Web テストの開発」村上公一・砂岡和子・劉松
日本中国語学会第54回全国大会予稿集 2004年11月日本中国語学会

“Mobile CALL-drills and Web-test Tools for the Communicative Chinese Conversation Lessons” Song Liu、Kazuko SUNAOKA、Yoshiyori URANO CDROM The 18th Pacific Asia Conference on Language, Information and Computation 2004年12月

“Teaching discrimination between Tone-2 and Tone-3 based on statistical properties and acoustical characteristics of the Chinese Four Tones” Shizuo Hiki, Kazuya Imaizumi, Kazuko Sunaoka and Liming Yang、CDROM The 18th Pacific Asia Conference on Language, Information and Computation 2004年12月

「コーパス利用による中国語教育(2)可変的語彙教育を支援する中国語コーパス構築とその応用」砂岡和子 早稲田大学政治経済学部紀要『教養諸学研究』118号 2005年3月号

「PC、携帯端末に対応する中国語コーールドリルシステムの開発」劉松・砂岡和子・浦野義頼「ケータイ・カーナビの利用性と人間工学」20053月予稿集 31-32頁

「與評獎和 Self-Editing 相結合進行的網絡華文寫作教學」砂岡和子・村上公一・劉松
4th International Conference on Internet Chinese Education (ICICE 2005) 2005年6月台灣

「與語料庫相結合的中文 CALL 練習系統的開發」劉松・砂岡和子・浦野義頼 4th International Conference on Internet Chinese Education (ICICE 2005) 2005年6月台灣

「Computerized Adaptive Testing (CAT) 方式的網路中文口語測驗的開發」村上公一, 砂岡和子, 劉松 4th International Conference on Internet Chinese Education (ICICE 2005) 2005年6月台灣

「為口語能力提升的中文教學項目量化」砂岡和子・村上公一・劉松二十一世紀中文項目營運策略國際教學研討會 2005年6月台灣

「日本の手機在中文教育上應用的問題與實踐」劉松・砂岡和子・浦野義頼二十一世紀中文項目營運策略國際教學研討會 2005年6月台灣

(3) 出版物

- 『早稲田大学 CCDL 中国語国際遠隔教育シンポジウム論文集』2002年12月好文出版
- 『Tutorial 漢語課本』砂岡和子、于洋、蘇紅輝、楊志輝、李利津共著月東方出版 2003年4月
- 『Tutorial 漢語課本 2004年版』砂岡和子、李利津、王玉、劉敬華編東方書店 2004年4月
- 『プログレッシブ中国語辞典（袖珍版）』武信彰・砂岡和子他共著小学館 2004年11月
- 『Tutorial Chinese Basic(2005年版)』砂岡和子、李利津、王玉、劉敬華編
早稲田大学 CCDL 中国語部会 2005年4月
- 『Tutorial Chinese Advanced(2005年版)』砂岡和子編
早稲田大学 CCDL 中国語部会 2005年4月
- 『数字化校園』砂岡和子、村上公一、李均洋、張立新（松岡一郎原著東洋經濟新報社
『デジタル・キャンパス』中国語版 北京首都師範出版社 2004年9月

研究成果による工業所有権の出願・取得状況

なし

為提升口語能力的中文教學項目量化

砂岡和子 村上公一 劉松

摘要

近年來，隨著認知學習理論和功能法教學的發展，外語教學出現新的改革。認知學習理論把學習者看作是資訊加工主體，此外，新的改革強調要發揮多媒體電腦和網際網路所具有的特性。然而，華語語言研究的速度趕不上外語教學改革的腳步。如何把理論具體應用到外語教學實踐上，目前還沒有很好的答案，因此造成口語教學缺乏系統性和持續性。

動態教學需要網站的支援和管理體制，才能充分發揮師生互動的交際作用。本文將介紹早稻田大學與北京大學和台灣師範大學合作進行的“Tutorial 漢語”教學3年的實踐情況，具體分析海內外對外華語老師們提交到我們網路平臺上的幾百份教學報告，提出以人為本，電腦輔助，課本為綱，語料庫支援的口語教學項目量化方案。

I 早稻田大學“Tutorial 漢語”教學的特點

早稻田大學 Tutorial 漢語教學是在只能使用漢語的環境下，〔4名學習者+1名教學者〕的小班制會話課程。分為遠距型（圖1）和面授型（圖2）兩種授課方式，皆是以提升華語口語能力為目標的強化課程。教學者的培訓來源為以漢語作為第二語文進行研究的高等教育單位（北京大學對外漢語教育學院，臺灣師範大學華語文教學研究所），以及早稻田大學早大國際公司 Tutorial 漢語語言培訓中心。



圖1 遠距型教學



圖2 面授型教學

遠距教學在上課時利用網路電視設備，由在臺灣師範大學或北京大學的Tutor直接為身處日本的學生上漢語口語課，把國際間的交流實踐帶進課堂裡。隨著網絡技術日益進步，遠距教學已經幾乎不受物理環境的負面因素影響。參加電視會議時，師生在分離隔兩地的情況下，仍舊能夠充分發揮自己所有的語言能力、溝通能力、異文化認知能力以及網路電視會議的操作技術等，努力互動交流，進而完成開會目的這項任務。遠距教學創造了必須使用漢語的環境，使學生掌握了面對不同文化時的應對能力，提高了他們學習外語的動力¹。

面授型教學則在日本早稻田大學校內進行。由早大的 Tutor 對學生進行面對面授課，強調教師與學生之間的互動，用啟發、推想和引導等多樣教學技巧，激發學習者自我發展中文口

¹ 郑艳群 2004

語表達之潛力，完成語言交際任務。

我們經過 3 年多的實踐，2004 年共進行 480 節即時面授型教學，360 節即時遠距漢語口語課，參加人數約 250 名。2005 年進而擴大師生陣容，教學上向“聽說讀寫”全面發展，並加強了教學研究以及網絡的開發，目前已經取得了初步的成果。

II 基於漢語語料庫的教學管理平臺的開發

早大 Tutorial 漢語教學的首要目的是提高學生的口頭表達能力。“說”和“寫”屬於復用性技能，比被動的“聽”和“讀”這兩種接受性技能要困難得多。Tutorial 漢語教學主要依據認知法(Cognitive Approach)和功能法(Functional Approach)法的教學理論。認知法和功能法教學根據交際需要和目的，向學生展示真實和自然的交際情境，因此具有綜合性訓練語言技能，以及讓學生自由地表達個人意見的優點。這種教學理念適用於我們連接 3 個地區師生互動進行的動態教學。

我們鼓勵老師多與學生互動，用啟發、推想和引導等教學方法來協助學生完成語言交際任務。反對有錯必糾，用呆板的方式照本宣科。此外，不計較語音、語法等不影響交際的一些錯誤，以免學生失去學習信心。當然，從學習者的需求來看，交際的迫切性大於掌握結構的必要性。但從對初中級學生進行教學的效果看，掌握結構的必要性則大於掌握交際功能的迫切性²。許多口語教學因人而宜，因材施教，然而如果缺乏規範性，教學體系就會失去系統性。

2003 年我們根據“結構(structural)－意念(notional)－主題(topical)”相結合的教學理念，開發了支援這種動態教學的管理平臺。“結構”反映語言自身的規律，以結構作為教學或編寫教材的大綱，相對穩定，不會有太大的變動。而“意念”反映語言功能，如何把語言結構和功能結合起來是一個難題。“主題”則根據說話人所關心的事情不同，會不斷湧現新的話題，更難以和“結構”“意念”結合在一塊。於是我們利用語料庫，用數量統計的方法，把語言的結構和功能掛勾。同時，與基於語料庫的電腦輔助教學相結合，讓學生按自己的學習順序逐步加深對於語言規則的理解以及提高語言運用的能力。這種 Tutorial 漢語教學是以人為本，電腦輔助，課本為綱，語料庫支援的教學。下面具體介紹我們建設基於語料庫的漢語網路教學平台的過程。

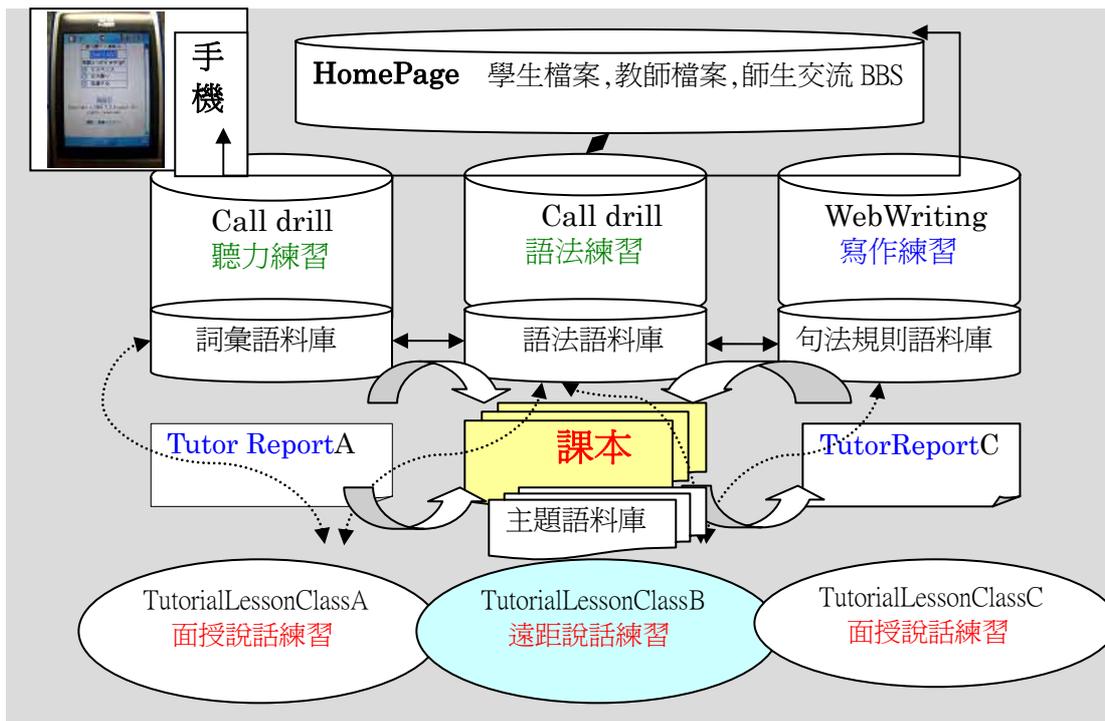
III 基於語料庫的 Tutorial 漢語網路教學平台

為了發揮網路教學的特點，系統地學習漢語結構和功能，同時把動態課堂上的語言資訊及時地回饋給學生，我們建立了連接 3 個地區師生的強大用戶聯盟網路。不僅已經實現了網上提交[教學報告]、[批改作業]、[記錄學習履歷]、[BBS]以及[網上考試]和[問卷調查]等功能，而且開發了基於各種語料庫的【CALL練習系統】³。這些電腦輔助練習同時支援PC和手機兩種終端裝置（圖 3）。在指定的網站（需要密碼）或在手機上，可以進行和課本相關的[聽力][語法]以及[寫作]等練習。基本上能輔助實際的交際口語教學。

預習(復習) ⇒ 上課 ⇒ 作業 ⇒ 復習(預習)

² 趙賢洲 1988

³ 劉松,砂岡和子,浦野義賴 2005



(圖 3) Tutorial 漢語教學流程圖

IV 設計漢語語料庫，統計結構和功能的相關度

如何在教學上有系統地結合語義表達和形式規則是一個尚未解決的難題。不管是“主題”與“結構”還是“主題”與“意念”都涉及到語言的表達和功能的關係。中國國家漢辦制定了“辭彙大綱”“漢字等級大綱”和“語法大綱”，“功能大綱”則沒能寫成，只好以“交際任務項目表”來代替現代漢語教學大綱⁴。我們爲了實現“結構—意念—主題”三項結合的理念，試編了奠基於漢語語料庫的教材⁵。儘管我們的工作未能徹底實行，下面仍列舉一些具體的工作情況供各位參考。

1) “主題語料庫”的設計

“Tutorial 漢語課本”的編寫參照了中國國家漢辦爲漢語短期教學制定的“交際任務項目表”、“現代漢語基本句模”⁶以及我們平台“教學報告”的上課討論題目匯報等 3 種原始資料，並且將這些資料合併、簡化，優選日本國內學生使用頻率高、覆蓋率大的 20 件交際任務項目。然後依據北京大學計算語言研究所《現代漢語短語結構規則庫》《現代漢語語法信息語料庫》《現代漢語標注語料庫》等結構分類以及語法標注，來嘗試編寫能夠反映出“意念”的具體課文和語句。

中國國家漢辦“交際任務項目表”奠基於交際任務教學法，適於讓學習者用漢語去完成一系列任務，具有明確的情景性和話題性。此外，可以根據多等級學生的特點，動態地設計教學實施方案，具有針對性和靈活性⁷。《Tutorial 漢語課本》按照“初級任務”的基本交際類、生存類、社會活動類和綜合資訊類等 5 大類來具體描述交際任務專案所涉及的 20 件項目，並以這 20 件項目作爲課本的主題。每課圍繞一個話題編寫，會話場景從大學的日常生活開始，

⁴魯川他 2000 年

⁵砂岡和子、李利津、王玉、劉敬華 2005 年

⁶同注 4

⁷中國國家對外漢語教學領帶小組辦公室 2002 年

逐漸發展為和中國社會和文化相關的話題。

以這 20 項主題作為核心資料，我們製作了“主題語料庫”，以備就“主題—語法(句型)—詞彙”的相關度進行統計。將來可以用這些數據編纂各種教材。

同時我們對“主題”專案本身進行調查，用 3 個地區師生每次下課後向平台提交的“教學報告”來定期調查在上課中“學生關心的討論問題”是否與課本的“主題”一致。比如 2004 年我們共收集了 652 份“教學報告”，其中第 15 課有關“出事”的 24 份教學報告中，有下列各式各樣的“學生關心的討論問題”。

A 班	講講自己身邊發生的大大小小的事故
B 班	遇到地震、火災時應該怎麼做，碰到交通事故時，怎樣處理（報警還是私了）
C 班	你看見過車禍現場嗎？是怎麼樣的情況？
D 班	發生地震的時候，應該怎麼辦？
E 班	自己曾經遇到過的麻煩
F 班	事故的后果
G 班	如果你是車禍的受害人，你有什麼要求？ 如果你是車禍的責任人，你怎麼處理這場車禍？

(圖 4) 第 15 課“出事”的“學生關心的討論問題”

從上面報告中我們得知以 15 課“出事”的主題來說，學生對自己所經歷過的體驗以及處理方法最感興趣，除了我們已經在此課列出的“發生、地震、火災、車禍”等跟“出事”密切相關的常用詞以外，“應該怎麼辦(怎麼做、怎麼處理)、V 過、報警、后果”等句型和詞語也出現得相當頻繁。我們把這些資訊輸入“主題語料庫”裡，以備統計“主題—語法(句型)—詞彙”的相關度。

另外，我們分析 Tutorial 漢語上課錄像，進一步研究“出事”這一個主題跟哪些句型和詞語相關度較大。(圖 5)是 2004 年 10 月 24 日台灣師範大學給早大遠距授課用的語言調查。可證明“出事”這主題的討論裡，〔存在句〕〔被動句〕〔助動詞〕〔形容詞〕〔程度副詞〕等句型以及相關詞彙的出現率較高。這些資訊我們也輸入“主題語料庫”裡。

台灣遠距 高級 E 班	存在句	發生了地震 停水停電 有很大的颱風 失火了
	被動句	車子被房子壓倒了 受災
	經驗態	自己曾經遇到過什麼麻煩? 我老家受到了火災
	助動詞	警察會來的
	形容詞	災情很嚴重 很恐怖 比較安全 非常危險
	程度副詞	真恐怖! 要立刻藏好!
	感情動詞	感覺得到嗎?
	感歎詞	哇! 厲害!
	傳聞詞句	據說 傳說
	程度補語	搖得很厲害
	方向補語	東西掉到地下 跑到外面去 往高地走去

(圖 5) 第 15 課“出事”的課堂語言分析(台灣師範大學遠距課)

2) “語法(句型)語料庫”的設計

我們的“語法(句型)語料庫”參照北大計算語言研究所的《現代漢語短語結構規則庫》進行基本設計。它既反映了詞語間的聚合關係，又更多地描述每個詞語同其他詞類的組合關係，是一個抽象程度較高的漢語句法語料庫。它將大部分的詞歸入了 18 個基本詞類。然後把這些詞類相組合，構成合法短語。然後再用短語替代合法短語中的詞類，構造出更複雜的短語，如此得到 600 多條短語結構規則，將短語的構造規則描寫清楚⁸。我們把上述“主題語料庫”裡的句型和詞語跟《現代漢語短語結構規則庫》的“結構表”和其“例句”相比較，製作出一個Tutorial漢語的“語法(句型)語料庫”(圖 6)。

北大《現代漢語短語結構規則庫》			《Tutorial 課本》主題語料庫	
分組	結構	動賓結構表	常用句型和詞語	主題分類
v-v	述賓	v(助動="助")+v	想跳舞 會游泳 可以唱歌	興趣
v-v	述賓	vf+v(準謂賓="準")	觀看表演 參加足球比賽	興趣愛好
v-d-v-v	述賓	<lhvuvp>+v	會不會來 能不能來 該不該報警 可以不可以修理	出事
v-r-a	述賓	vb+<zwrap>	覺得她漂亮 希望你成功	服裝
v-v	述賓	"有"+v	有安排 有照顧	旅行
v-m-q	述賓	v+<數量>	買二三十斤 看三次 找了十塊錢	購物

(圖 6) 語法(句型)語料庫的設計

我們以這種方式為每個主題挑選出 100 多條句子，全書共 20 課，一共登記 2000 句，建立了一個“語法(句型)語料庫”。北大《現代漢語短語結構規則庫》有計畫要編一部《現代漢語語法資訊詞典》，進一步描述詞類與詞類之間的組合關係⁹。我們的努力方向是要構建一個專供漢語口語教學使用的“語法(句型)語料庫”。

3) “詞彙語料庫”的設計

我們“詞彙語料庫”的收詞標準和範圍，是根據北京大學計算語言研究所《現代漢語語料庫》2750 萬條詞的使用頻率 (Frequency)以及通用性 (Usuality)的統計結果選定高頻詞¹⁰，和中國國家漢辦“漢語水平考試(HSK)”的約 80000 詞，以及約 2000 條日本教科書的常用詞彙。我們先把這些詞彙收集好，然後根據難度分 3 級，再加上其他的屬性一起給這些詞標注。

《Tutorial漢語課本》收錄大約 10000 個常用詞句，相當於《漢語水平等級》五級的水準。其中最常用詞句重複出現兩三次，因此實際上能學到將近 3000 左右的詞句。學好本書後，可以與中國人進行基本交際。同時採用看圖說話的學習方法，讓華語老師能夠直接用中文進行教學。

詞彙以口語和短句為主，比如第 15 課收錄了“安全、車禍、地震”等有關“出事”常見的實詞 300 多條，此外還有“被”“常”“突然”“出來”等虛詞，以便學生說出完整的句子。因為我們上課注重動態教學，使用的語句隨著交際主題而變化，不同的交際情境產生不同的詞彙及用語。海內外華語老師們向我們網路平臺提交的教學報告中(圖 7)，幾乎每個班都超出既定詞彙的範圍外。在 2004 年 10 月 26 日—11 月 1 日之間，有 25 個水平不同的班級上了 Tutorial 漢語課。這段期間每個班都討論第 15 課“出事”，用的課本都一樣，然而上課用的詞語卻是五花八門(圖 8)。

⁸俞士汶,朱學鋒,李峰 1998

⁹同注 8

¹⁰張化瑞 2004,朱學峰、張化瑞、段慧明、俞士汶 2004 年



(圖 7) 網頁教學報告的課堂用生詞

班級	Tutor	等級	課堂用的詞句 (依據 Tutor 的教學報告)
面授 L	B	初級	乾燥, 燃燒, 倒塌, 道路, 預告, 根據, 震災, 備用, 墜毀, 空難, 裂開, 情味, 扒手, 酷暑, 出海, 洪水, 做篷車, 麵包車, 小臥車, 小貨車, 猜測, 波浪, 海嘯, 劫機
面授 G	E	中級	爬到樹上, 強盜, 一艘輪船, 倒塌 (房屋倒塌了), 劫持, 乘客, 燃燒 (火燃燒起來了), 報警, 架 (一架飛機), ~級 (日本發生了 6 級地震), 人質 (把乘客作為人質), 救命啊!, 損失, 察覺 (他察覺到有人在偷他的錢包)
面授 M	H	上級	墜毀, 劫機, 空難, 規劃, 海嘯, 扒手, 天乾物燥, 天災人禍, 趁火打劫, 雪上加霜, 雪中送炭, 傾斜, 傾倒, 便道, 非機動車,
遠距 台灣 C	L	中級	搖晃, 電線桿, 洪水, 意外, 消防隊員, 勇敢, 技術, 水桶
遠距 北京 L	O	初級	地震, 發生, 出事, 員警, 立刻, 安全, 危險, 妥善, 失事, 失火, 救火, 滅火, 防火, 受災, 人數, 統計, 水災, 丟, 馬虎, 注意, 爆炸, 恐怖, 出現, 小偷, 車禍, 處理, 事故, 騎, 撞, 曾經, 手錶, 碩士, 研究生, 厲害, 體育館, 藏, 跟, 一樣

(圖 8) 課堂用詞表

網頁教學報告上的課堂生詞會自動登記到“詞彙語料庫”中，如此設計的原因是為了將這些資訊用半自動化的方式登錄，以供課本修訂之用，或者用來為電腦輔助練習和分班測試等出題。另一方面可以及時反映實際動態教學資訊，大大節省教師的時間和精力¹¹。

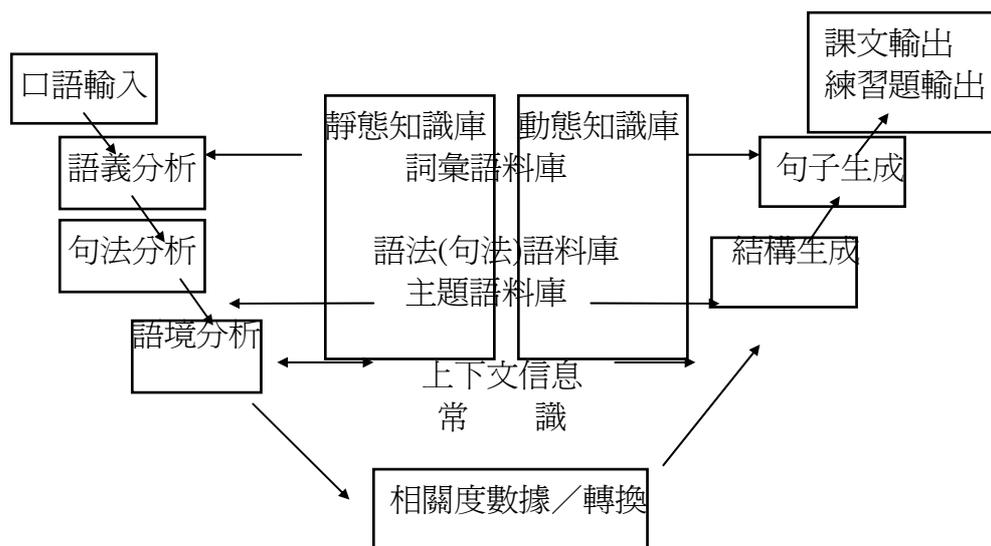
V 依據口語生成規律的中文教學項目量化

我們製作“主題”“語法(句型)”以及“詞彙”等語料庫，以這3種口語語料庫進行漢語口語教學量化規則的實驗研究。因為靜態和動態語料庫裡都包含了豐富的語言學知識，並隱含了語言中的一些普遍規律，因此如何最大限度地發揮這個語言知識庫的作用，是提高漢語口語教學品質的關鍵。基於這種認識，我們建立了基於語料庫的網絡教學管理組織(圖3)。

目前，我們只有主題—語法—詞彙的動態語料和它們之間相關度的數據。今後我們期待進一步獲得教學實踐中的語料，等標注語料的數據達到一定規模以後，利用統計方法得到的分布數據能夠近似地反映出語言中的一些普遍規律。將這些數據運用於3種語料的相關度自動標注處理，可盼獲得口語教學量化大綱。如此，將語料庫得到的普遍性知識和面授上課中學到的特殊性知識相結合，可以大大提高口語教學的效果¹²。

¹¹劉松·砂岡和子·浦野義賴 2005 年

¹²俞士汶,朱學鋒,李峰 1998



(圖 9)口語生成的基本模型

我們的中文教學量化方案試圖為漢語信息處理的研究找出路,建立口語語料庫,並採取基於統計和數字進行口語教學的模式。希望將靜態教學轉化為動態教學,將教學研究從演繹性轉化到歸納性教學,如此一來可以協助學生從被控制的受限者轉變為自律的學習者,協助口語教學的老師從具有個人技巧的教學者轉化到掌握口語生成規律的設計者。對母語非華語的中文教師來說,學會這種技能特別重要。

附記:本稿發表得到自 2002 到 2004 年度日本文化科學省學術研究費(基盤研究 B-2:研究題目“漢語 TutorialSystem 的國際開發” 課題號碼 14310136:研究代表 砂岡和子)以及 2001 年早稻田大學特定課題研究(國際共同研究)“中国語チュートリアル遠隔学習システムの国際共同開発—学習プログラムの企画とデジタル教材作成—” 課題號碼 20003C-002:研究代表 砂岡和子)的撥款,為該項研究成果的一部分。

參考文獻

- [1]魯川他 2000 年 現代漢語句模《世界漢語教學》第四期
- [2]中國國家對外漢語教學領帶小組辦公室編 2002 年《高等學校外國留學生漢語教學大綱》北京語言文化大學出版社
- [3]馬箭飛 2000 年 以“交際任務”為基礎的漢語短期教學新模式《世界漢語教學》第 4 期(總第 54 期)
- [4]馬箭飛 2003 年 對外漢語教學模式化研究《第七屆國際漢語教學討論會論文選》
- [5]祖人植 2003 年基於交際能力的對外漢語教學研究芻議《第七屆國際漢語教學討論會論文選》
- [6]趙賢洲 1988 建國以來對外漢語教學初級階段教學大綱,第二屆國際漢語教學討論會論文選
- [7] 史有為、周剛 2004 年 日本第二外國語漢語教育之探討—以教學大綱和教材為中心— 明海大學外國語學部論集第 16 集
- [8]俞士汶、朱學鋒、李峰 1998 現代漢語詞語的語法知識描述《漢語計量與計算研究》國際會議論文集,香港城市大學
- [9]俞士汶、朱學鋒、王惠、張化瑞、張芸芸 2003 年《現代漢語語法資訊詞典詳解(第二版)》清華大學出版社
- [10]俞士汶、朱學鋒、李峰 現代漢語詞語的語法知識描述
朱學鋒、張化瑞、段慧明、俞士汶 2004 年 漢語高頻詞語法信息詞典的研制《語言文字應用》第 3 期
- [11]張化瑞 2004 年「語法タグ付大規模コーパスに拠る現代中国語の語彙統計」

早稻田大学遠隔教育中国語部会漢語コース研究会第2回報告会

- [12]史有爲、張延瑞 2004 年「若干大学教材の初步統計報告」早稻田大学遠隔教育中国語部会漢語コース研究会第3回報告会
- [13]周強、黃昌甯 2002 年 漢語句法規則的自動構造方法研究智慧技術與系統 清華大學電腦科學與技術系國家重點實驗室網站資料
- [14]鄭艷群 2004 日本早稻田大學 Tutorial 漢語遠端教學模式評析《世界漢語教學》2 期
- [15]李利津 2003 年《Tutorial 漢語課本》教師使用手冊,早稻田大学遠距電視會話教學 早稻田大学 Tutorial 漢語遠程教學師資培訓用資料
- [16]砂岡和子、李利津、王玉、劉敬華編 2005 年 “Tutorial Chinese Basic” 早稻田大学 CCDL 中国語部会出版
- [17]砂岡和子 2005 年 “Tutorial Chinese Advanced” 早稻田大学 CCDL 中国語部会出版
- [18]砂岡和子、村上公一 2003 年以提升交際能力爲目的之漢語教學以及水平測驗芻議-早稻田大學國際漢語遠距教學四年實踐見證-第三屆全球華文網路教育研討會論文集
- [19]劉松、張悅、倪萌、砂岡合子、村上公一、浦野義賴 日中台遠端網路漢語口語教學管理平臺的開發-早稻田大學 Tutorial Chinese 網路平臺-第四屆中文電化國際研討會予稿集 2004
- [20]砂岡合子、村上公一 “以提升交際能力爲目的之漢語教學以及水平測驗芻議—早稻田大學國際漢語遠距教學四年實踐見證”，第三屆全球華文網路教育研討會論文集，2003
- [21] 劉松、吳劍明、浦野義賴 “基於 PHP 與 MySQL 的互動型漢語教學支持平臺的研究與實踐” 第三屆全球華文網路教育研討會，2003.10
- [22] 劉松、砂岡和子、浦野義賴 2005 年 日本的手機在中文教育上應用的問題與實踐 台灣 ITCSL 予稿集印刷中
- [23] 劉松、砂岡和子、浦野義賴 2005 年 與語料庫相結合的中文 CALL 練習系統的開發 4th International Conference on Internet Chinese Education (ICICE 2005) 台灣予稿集
- [24]村上 公一、砂岡和子、劉松 2005 年 Computerized Adaptive Testing (CAT) 方式的網路中文口語測驗的開發 4th International Conference on Internet Chinese Education (ICICE 2005) 台灣予稿集

原載「二十一世紀中文項目營運策略國際教學研討會」2005 年 6 月台灣

Computerized Adaptive Testing (CAT) 方式的 網路中文口語測驗的開發

村上公一、砂岡和子、劉松

關鍵字：CAT，IRT，電腦化適性測驗，試題反應理論，網路測驗

摘要：

電腦化適性測驗 (computerized adaptive testing, CAT) 是以試題反應理論 (item response theory, IRT) 為基礎的測試。根據考生的答題正誤，調整下一題的難易度，從試題庫中選出對學生最適合的試題。其優點在於：根據考生的答題正誤，調整下一題的難易度，所以考生不用解答過難或過易的試題，每試題對考生中文能力估測的訊息量很大，考生只解答相當於一般的試卷的 10%-20% 的試題及時間，就可以精確估計考生的中文水平。

早稻田大學 “Tutorial Chinese” 自 2003 年 4 月以來已嘗試過 4 次一般性的網路分班測驗，2005 年 4 月嘗試了第 1 次電腦化適性測驗 (CAT) 方式的網路中文口語測驗。

早大網路中文測驗共分為 3 個部分：

- (一) 詞彙聽力：共 10 題 (聽詞彙選日譯)，3 項選擇，用來測試考生的詞彙量
- (二) 完成說話：共 10 題 (聽一段話，然後補充後續部分，完成全文)，3 項選擇，用來測試考生的聽、說能力 (基礎表達能力)
- (三) 復讀：共 10 題，4 項自報選擇 (包括復讀全部、復讀一半、只復讀一兩個詞、一點也不能復讀)；用來測試考生的聽、說能力 (聲音化能力)

本文介紹我們開發的早大網路中文測驗的具體內容，並提供給中文教學者作為參考。

0 · 前言

早稻田大學從 2002 年開始推展 Tutorial Method 漢語會話訓練。目前有 213 名學生分 60 個小組參加訓練。推展互動型小班漢語教學時，為了確保同一班上的同學口語水平一致，以利分班，所以需要測驗學生的口語能力。我們自 2003 年 4 月以來已嘗試過 4 次一般性的網路分班測驗，2005 年 4 月嘗試了第 1 次電腦化適性測驗 (CAT) 方式的網路中文口語測驗。本文介紹我們開發的早大網路中文測驗的具體內容，並提供給中文教學者作為參考。

1 · 一般性的網路分班測驗的概要

我們嘗試過 4 次一般性的網路分班測驗；第 1 次在 2003 年 4 月及 7 月，第 2 次在 2003 年 12 月，第 3 次在 2004 年 4 月，第 4 次在 2004 年 7 月舉行。各次測驗的具體出題形式如下(表 1)。

	(1)詞彙拼音	(2)詞彙聽力	(3)回答提問	(4)完成說話	(5)復讀
第 1			共 5 題	共 5 題	共 10 題
第 2			共 5 題	共 5 題	共 10 題
第 3			共 5 題	共 5 題	共 10 題
第 4	共 10 題	共 10 題		共 10 題	共 10 題

表 1(網路分班測驗出題形式)

已試過 5 種試題。(1)詞彙拼音；看詞彙選拼音，(2) 詞彙聽力；聽詞彙選日譯，(3)回答提問；聽一段話，然後回答提問，(4)完成說話；聽一段話，然後補充後續部分，完成全文，

(5) 復讀；聽一段話，然後復讀，自報結果（包括復讀全部、復讀一半、只復讀一兩個詞、一點也不能復讀）。

第 3 次測驗在 2004 年 4 月舉行，108 名學生參加。同時讓學生參加 TECC(Test of Communicative Chinese)團體考試（4 月 24 日舉行）。TECC 是由(日本)中文交際協會組織實施，主要檢定考生的中文運用、交際能力的“中文交際能力測驗”。從 1996 年研制開發，到現在共計舉辦 16 次公開測試，另舉辦好多次個別團體測試。考試時間為 80 分鐘，聽力試 70 題、筆記試 70 題，共 140 題。108 名中 78 名參加 TECC 團體考試。我們分析 2004 年 4 月份舉行的兩種測試的學生成績的相關性，確認分班測驗的可靠性。表 2、表 3 是第 3 次分班測驗和 TECC 團體測試的基本統計表和相關系數表。

	(3) (25)	(4) (25)	(5) (50)	總分 (100)	聽力 (500)	筆記 (500)	總分 (1000)
平均	9.598	10.920	28.080	48.598	227.310	168.092	395.402
標準偏差	5.868	4.733	11.182	18.286	69.046	57.744	121.477
最低	0	0	9	11	92	76	193
最高	0	25	50	90	436	396	832

表 2(基本統計表)

	(3)	(4)	(5)	總分	聽力	筆記	總分
(3)	1.000						
(4)	0.411	1.000					
(5)	0.609	0.470	1.000				
總分	0.800	0.678	0.929	1.000			
聽力	0.561	0.513	0.743	0.767	1.000		
筆記	0.456	0.443	0.645	0.656	0.835	1.000	
總分	0.536	0.502	0.729	0.748	0.965	0.950	1.000

表 3(相關系數表)

第 3 次網絡分班測驗總分與 TECC 團體測試的總分的相關系數(R)達 0.748，屬於高度相關。與聽力試得分的相關系數(R=0.767)比與筆記試得分的(R=0.656)高 0.111，表示網絡分班測驗的總分主要反映以聽力為主的中文交際能力。

網絡分班測驗的三種出題形式中與總分的相關性最大的是 (5) “復讀”，其次是(3) “回答提問”，(4) “完成說話” 最小。“復讀” 與 TECC 總分、TECC 聽力的相關性也相當高，可以認為是最合適的出題形式。因為“回答提問”和“完成說話”的出題數只有 5 題，所以與 TECC 總分、TECC 聽力的相關性都不

按分班測驗的結果分班後開始課。第一週的 Tutorial 以自我介紹為中時 Tutor 確認各各學生的實際的中文交的小班上的次序。表 4 是 Tutor 課上確生次序與按照各種測驗的分數推測的比較。

“0” 表示 Tutor 確認的次序與按照測驗的分數推測的次序一樣，“+”、字表示兩種次序之間的差異。按照

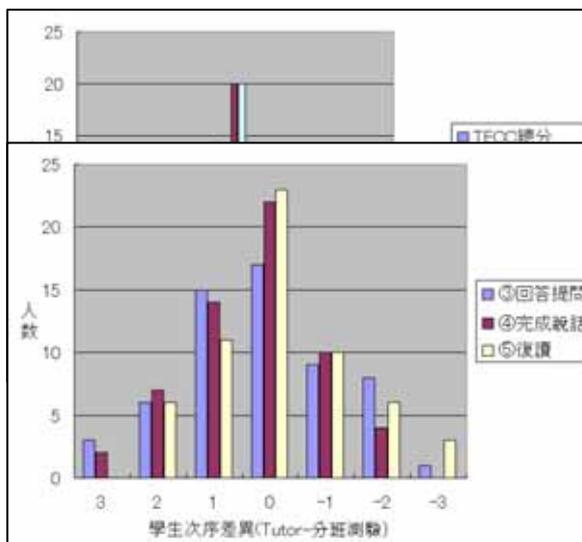


表 5(學生次序差異 2)

那麼高。Tutorial 心，上課際能力認的學次序的各種測“-”數 TECC 聽

力推測的次序和網絡分班測驗推測的次序與 Tutor 確認的次序比較接近，但不少學生的次序有±2 以上的差異。

表 5 是 Tutor 課上確認的學生次序與按照網絡分班測驗的各種出題形式的分數來推測的次序的比較。按照完成說話、復讀來推測的次序與 Tutor 確認的次序比較接近。回答提問差異比較大。

通過這些比較分析，我們了解“回答提問” 5 題、“完成說話” 5 題、“復讀” 10 題的一般性的網絡分班測驗已有和 TECC 差不多一樣的水平，但還不到我們所需要的精度。

爲了提高測驗精度，我們自 2004 年 4 月開始與早稻田大學 International 公司、教育測定研究所合作研制開發電腦化適性測驗（CAT）方式的新的網路測驗。

2· 電腦化適性測驗（computerized adaptive testing,CAT）

電腦化適性測驗（computerized adaptive testing,CAT）是以試題反應理論(item response theory, IRT)爲基礎的電腦測試。

試題反應理論與古典測驗理論的最大差異在於；後者依據各各試題的實分數，前者依據各各試題的參數。先實施大量預試，算出各各試題的難易度、鑑別度、猜測度等參數，建立題庫，然後自題庫中挑選試題以組成測驗。考試完後，處理預先算出好的參數推測考生的能力。因爲題庫裡的試題的參數都算出在同一個量尺上，一旦建立試庫我們就可以編製好幾種“同一個量尺、不一樣內容”的測驗。近幾年研制開發的外語測試一般都採用試題反應理論。上面所提的日本 TECC 也採用這些方式。

適性測驗(Adaptive testing)是利用依據試題反應理論建立的題庫組成測驗的一種方式。根據考生先前的答題情況，自題庫中挑選最能評量出考生能力的下一題進行施測。比方說如果考生上一題答對了，下一題就選難一點的試題，答錯了就選簡單一點的試題，每次都從試題庫中選出對學生最適合的試題。

傳統方式的測驗一般讓所有考生考同一張試卷，不考慮考生的個別差異，難題、易題都放在裡面，考生常碰到過難、過易的試題。這些試題對估測考生能力沒有積極作用，只是消耗考生的時間和精神而已。適性測驗的每道試題則都根據考生的答題正誤，調整下一題的難易度，所以考生不碰到過難、過易的試題，每試題對考生中文能力估測的訊息量很大，可以節省試題量及時間，而提高測驗的準確性。適性測驗由於選題、計分數需要大量的運算，因此適於電腦實施，這叫做“電腦化適性測驗（computerized adaptive testing,CAT）”。

下面介紹我們研制開發的電腦化適性測驗（CAT）方式網路測驗的建設流程及具體內容。建設流程如下。

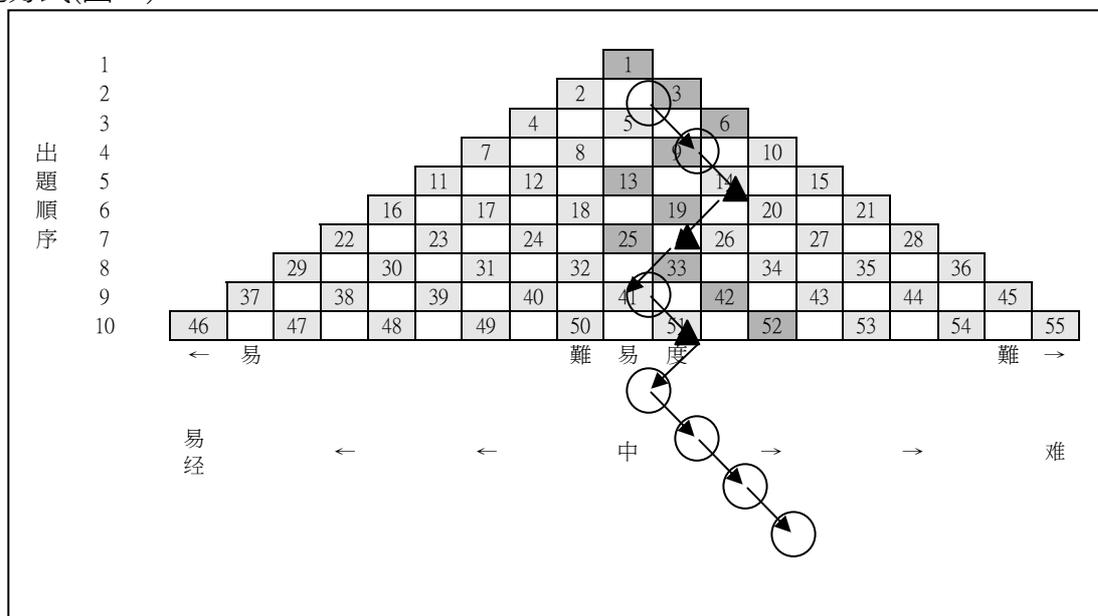
- （一）測驗總設計(確定測驗的目的、形式、出題數、精度等)
- （二）試題命題(爲了預試)
- （三）實施預試(利用一般性的網絡分班測驗系統)
- （四）建立試題庫(算出各各試題的難易度、鑑別度、猜測度等參數)
- （五）設計測驗系統

測驗的主要目的是精確估計以表達能力爲主的中文口語水平，作爲分班測驗、期末測驗之用。我們在第 1 到第 3 次一般性測驗已試過“回答提問”“完成說話”“復讀”的 3 種出題形式，第 4 次測驗再加上“詞彙拼音”及“詞彙聽力”兩種詞彙問題，我們採用比較反映考生的口語能力的“詞彙聽力”“完成說話”“復讀”的 3 種出題形式。

- （一）詞彙聽力：共 10 題（聽詞彙選日譯），3 項選擇，用來測試考生的詞彙量
- （二）完成說話：共 10 題（聽一段話，然後補充後續部分，完成全文），3 項選擇，用來測試考生的聽、說能力（基礎表達能力）

(三) 復讀：共 10 題，4 項自報選擇（包括復讀全部、復讀一半、只復讀一兩個詞、一點也不能復讀）；用來測試考生的聽、說能力（聲音化能力）

詞彙聽力及完成說話依據電腦化適性測驗方式，復讀採用傳統形式（但因隨機存取而出題，所以每次試題不一樣）。各 10 題，共 30 題。詞彙聽力及完成說話採用樹狀結構形式的出題方式(圖一)。



圖一(樹狀結構出題形式)

圖上○表示答對，▲表示答錯。考生答對了第 1 題（試題 1），就出題比試題 1 難一點的試題 3。答對了第 2 題（試題 3），就出題比試題 3 難一點的試題 6。答錯了第 3 題（試題 6），就出題比試題 6 簡單一點的試題 9。每道題都根據考生的上一題的答題正誤，調整難易度，所以考生不用解答過難或過易的試題。圖上的學生分為 19 層難易度的試題中只回答 4 層試題（〔1、13、25〕〔3、9、19、33〕〔6、42〕〔52〕）而已。除了這 4 層以外的 15 層試題對這個考生來說過難或者過易的試題，對估測考生能力沒有用處。

我們在 2004 年秋期實施總試題數 270（詞彙聽力 120、完成說話 120、復讀 30），400 多名學生參加的預試，算出各各試題的參數，建立好試題庫，2005 年 4 月嘗試了第 1 次電腦化適性測驗（CAT）方式的網路中文口語測驗稱謂“Test of Ability for Chinese Communication(TACC)”（試用版本）。



圖二（TACC 封面）

在 2005 年上學期開始的 4 月中旬，總共有 213 名學生考這個新網絡分班測驗，按測驗成

續把 213 名學生分到 60 個小班。剛上完一週的小班課，還沒有學生因中文水平不適而需要換班。

表 4、表 5 是四月份進行的測驗的基本統計表及相關系數表。

	詞彙聽力 (40)	完成說話 (40)	復讀 (20)	總分 (100)
平均	22.023	22.032	12.361	56.418
標準偏差	6.122	5.319	4.511	14.012
最低	7	10	1	26
最高	40	40	20	92

表 4 (基本統計表)

	詞彙聽力	完成說話	復讀	總分
詞彙聽力	1			
完成說話	0.682	1		
復讀	0.638	0.637	1	
總分	0.901	0.883	0.843	1

表 5 (相關系數表)

因為時間限制，我們還沒完全分析過這次測驗的詳細結果，也沒有進行過與 TECC 等其他測驗比較分析，但可以看得出這次測驗獲得了相當好的成果。

3 · 結語

我們自 2002 年開始推展 Tutorial Method 小人數漢語會話訓練以來，怎樣能進行最合適的分班，一直是重要的問題。這次我們開發的“Test of Ability for Chinese Communication(TACC)”是對這個問題的初步答案。只費 10 分鐘左右的測驗時間能做出相當好的分班結果。但 TACC 並不是完美的測驗，裡面有很多問題。比方說復讀題不是客觀分析考生的聲音評估的，目前採用自報選擇方式評估。這是一個暫時的措施。等到由於聲音分析技術的迅速發展，再等幾年一定能做出客觀評估。那時候我們也可以用新的技術評估復讀題。

附記：本稿發表得到自 2002 到 2004 年度日本文化科學省學術研究費(基盤研究 B-2：研究題目“漢語 Tutorial System 的國際開發” 課題號碼 143101136：研究代表 砂岡和子)的撥款，是該研究成果的一部分。

參考書目

1. 芝祐順(1991) “項目反應理論-基礎と応用-” (東京大学出版会)
2. 村上公一、砂岡和子、劉松(2003) “中国語レベル分けのための Web 上での中国語能力診断テスト” (日本中国語学会第 53 回全国大会予稿集)
3. 砂岡和子、村上公一 “以提升交際能力為目的之漢語教學以及水平測驗芻議-早稻田大學國際漢語遠距教學四年實踐見證”
4. 村上公一、砂岡和子、劉松(2003) “発話能力を測定するための中国語 Web テストの開発” (日本中国語学会第 54 回全国大会予稿集)

原載「第四全球華文網路教育研討會」2005年6月台湾

與語料庫相結合的中文 CALL 練習系統的開發

劉 松 砂岡和子 浦野義賴

摘要：

隨著人類在電腦多媒體技術、網路技術、資料庫技術、人工智慧等方面的快速發展，能夠以電腦或智慧型終端作為客戶機，通過 Internet 獲得長期的、高效的、即時方便的終身學習機會，已經成為新一代教育的發展方向。早稻田大學 2000 年開始了名為“Tutorial Chinese”的小班口語強化中文教學模式，並在 2003 年開始利用網路平臺進行教學資訊的共用和語料庫的收集與分析。2004 年，我們又開發了基於語料庫的 CALL 練習系統，這套系統同時支援 PC 與手機兩種端末。

本文將介紹語料庫的構成以及我們開發 CALL 練習系統時的具體實踐。

關鍵字：

Tutorial Chinese, CALL, 語料庫, 手機, 多語言支援

The Development of Chinese CALL Drill System based on Corpus

Abstract:

With the development of Multimedia, Network, Database and AI technology, E-learning has being a popular topic and new tendency in education field. Waseda University has been carried out a new Chinese language education mode named Tutorial Chinese since the year 2000. It takes the small class lesson and whose objective is to enhance the students' ability of oral conversation. In order to support this language education mode, we have developed Tutorial Chinese Platform in 2003, to gather, share and analyze the educational information and linguistics corpus. In 2004 we have constructed PC and mobile telephone oriented CALL-Drill system as one major assistant part of this Platform.

In this paper, the development process of CALL system and the construction of corpus will be introduced.

Key words:

Tutorial Chinese, CALL, Corpus, Mobile telephone, Multi-language supported

一·序言

隨著人類在電腦多媒體技術、網路技術、資料庫技術、人工智慧等方面的快速發展，能夠以電腦或智慧型終端作為客戶機，通過 Internet 獲得長期的、高效的、即時方便的終身學習機會，已經成為新一代教育的發展方向。遠端教學利用現代技術手段，打破地域和時間的界限，不僅使講授更為生動，而且使遠距離的雙向交流成為可能。語言教育也勢必受這一趨勢的影響。

語言教育的目標是培養學生的意思表達能力，要達到這種能力除了掌握語法以外，還需要通過實際交流來鍛煉。目前，海外漢語學習者的數量急劇增加，由於缺少與母語話者的交流機會，造成了學生經過長時間的學習卻不敢開口說話的現象，在這種情況下，早稻田大學於 2000 年開始進行了 Tutorial Chinese 教學，採用了一名 Tutor(輔導老師)指導四名學生的少人數口語會話強化班的模式，分為遠端型(圖 1)和面授型(圖 2)兩種。遠端型教學在上課時利用網路電視設備，由在臺灣或大陸的 Tutor 同期即時地指導日本學生；面授型教學則在日本早稻田大學校內進行，由早大的 Tutor 對學生進行面對面授課。目前 Tutor 分別由北京大學，臺灣師範大學及早稻田大學的母語話者組成，一年共進行 220 節即時遠端漢語口語課，參加人數約 200 名。



圖 1 遠端型教學



圖 2 面授型教學

爲了更加有效地支援這種新型的教學模式，我們于 2003 年開發了 Tutorial Chinese 教學管理平臺。由於 Tutorial Chinese 重視面對面的現場交際口語教學，所以該平臺也與一般的以靜態課程爲中心的不同，而是以動態口語教學爲中心，發揮網路教學的特點，把動態課堂上的資訊及時地回饋給學生，形成連接 3 個地區師生的強大用戶聯盟網路。第一階段的開發實現了網上提交教學報告、批改作業、記錄學習履歷、BBS 以及網上考試和問卷調查等功能。2004 年，我們又結合 Tutorial Chinese 的教材，建立起名爲 WTCC (Chinese Corpus named WTCC) 的語料庫，並基於該語料庫開發了同時支援 PC 和手機的 CALL 練習系統。本文將介紹 WTCC 語料庫的構成以及 CALL 練習系統的開發過程，並對將來的功能擴展做簡單的闡述。

二 · WTCC 語料庫的構築

WTCC 語料庫是結合 4 部分的語料庫建立起來的，它們分別是北京大學現代漢語語法資訊詞典語料庫(約 10000 詞)，HSK (漢語水準考試) 甲乙丙級辭彙庫(約 8000 詞)，日本第二外語教育用基礎辭彙庫(約 3000 詞)，以及 Tutorial Chinese 教材的辭彙庫。

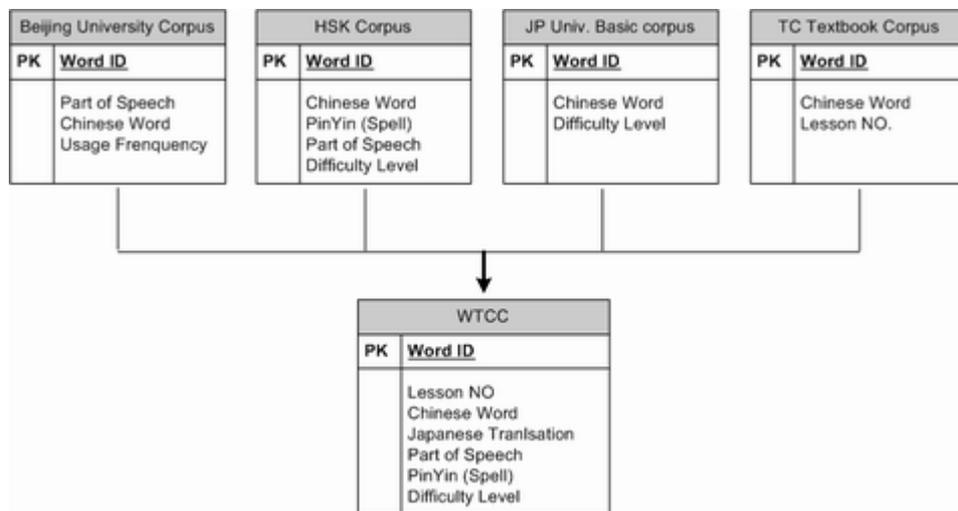


圖 3 WTCC 語料庫的構成

1	A F1	B 中文词汇	C 日文意思	D 词性	E 拼音	F 课别	G 词汇的初始难度级别
2970	你	你	あなた	pron.	ni3	14C	1
2971	喜欢	喜欢	好きだ	v.	xi3 huan0	14C	1
2972	羽毛球	羽毛球	バドミントン	n.	yu3 mao2 qiu2	14C	2
2973	早晨	早晨	朝	n.	zao3 chen0	14C	1
2974	人	人	人	n.	ren2	14C	1
2975	公园	公园	公園	n.	gong1 yuan2	14C	1
2977	足球	足球	サッカー	n.	zu2 qiu2	14C	2
2978	踢	踢	蹴る	v.	ti1	14C	2
2980	把	把	…を	prep.	ba3	14C	1
2981	球	球	ボール	n.	qiu2	14C	1
2983	运动员	运动员	スポーツ選手	n.	yun4 dong4 yuan2	14C	3
2985	方面	方面	方面	n.	fang1 mian4	14C	2
2988	有益	有益	役に立つ	adj.	you3 yi4	14C	3
2989	健康	健康	健康	n.	jian4 kang1	14C	2
2990	比赛	比赛	試合	n.	bi3 sai4	14C	2
2991	跳高	跳高	走り高跳び	n.	tiao4 gao1	14C	3
2992	冠军	冠军	優勝(者)	n.	guan4 jun1	14C	3
2994	跳远	跳远	走り幅跳び	n.	tiao4 yuan3	14C	3
2995	下午	下午	午後	n.	xia4 wu3	14C	1
2996	开始	开始	初め	v.	kai1 shi3	14C	1
2997	战胜	战胜	打ち勝つ	v.	zhan4 sheng4	14C	3
3000	真是	真是	本当にだ	int.	zhen1 shi0	14C	3
3002	虽然	虽然	…が	conj.	sui1 ran2	14C	2
3004	可是	可是	しかし	conj.	ke3 shi4	14C	2
3005	并	并	そして	conj.	bing4	14C	3
3006	灰心	灰心	がっかりしている	v.	hui1 xin1	14C	3
3007	教练	教练	監督	n.	jiao4 lian4	14C	3
3008	队员	队员	メンバー	n.	dui4 yuan2	14C	3
3009	不许	不许	不许	adv.	bu4 xu3	14C	3
3010	赢得	赢得	勝ち取る	v.	ying2 de2	14C	3

圖 4 WTCC 語料庫的內容

構建語料庫的主要步驟是：

將教材中的句子按照語素為單位進行分割，並按照常用詞句，課文，分組對話以及看圖聽故事的教材類別進行分類標號(C,K,G)；

通過資料庫查詢的方法從北大現代漢語語法資訊詞典語料庫、HSK 詞庫和日本第二外語教育用基礎辭彙庫中將單詞的詞性，詞義，拼音，難度查詢出來；

根據課文中的具體意思來刪除多義詞，多音詞的情況，僅保留教材原文出現的辭彙；

利用自動翻譯工具及人工檢查，確定單詞的日文解釋。

最終建成的語料庫中辭彙的屬性包括：課本中的 ID、中文解釋、日文解釋、詞性（例如：adj-形容詞，adv-副詞，aux-助動詞，clas-量詞，conj-連詞，n-名次，v-動詞，num-數詞，pron-代詞，等）拼音(音調分為 1,2,3,4,0)，音效檔案，難度(1-初級，2-中級，3-上級)。

目前我們主要利用 WTCC 語料庫來進行網上分級考試及 CALL 練習的題目編纂，因為該語料庫是根據 Tutorial Chinese 的教材建立起來的，所以基於它製作的練習和考試非常具有針對性，有利於學生的預習和復習。同時該語料庫中包含了從其他語料庫吸取的各種語法和發音資訊，利用這些資訊，結合出題理論的研究和程式的開發，可以進行半自動化乃至自動化出題。下一步，我們還將構築句子的語料庫以及在課堂上新出現的課本之外的語料庫，使其更加完善和具有即時性。

三·CALL 練習系統的實現

1. 系統結構

本系統基於 Browser/Server 結構，即瀏覽器/伺服器模式(圖 3)。用戶端，只要有一台安裝有操作系統 WINDOWS 9x(2000) 或 XP，可以上網，用 IE 等瀏覽器來瀏覽 Internet 資源就可以了，原則上不需要另外安裝任何軟件。與程式升級以及資料更新相關的工作放在伺服器端，Web 伺服器連接用戶瀏覽器與後臺資料庫，通過表單等形式在用戶與資料庫之間進行資料交換。這種基於 Web 的資料庫方式較好地結合了網路傳輸速度快的優勢和資料庫存取資料

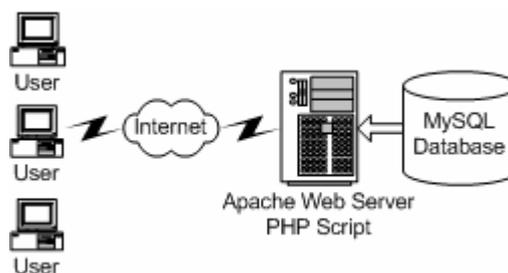


图 5 系统结构

2. 系統開發環境

教學平臺的開發採用了目前在教學研究單位非常通用的 LAMP 模式 (Linux + Apache + MySQL + PHP)，全部是免費的開源軟體。Linux 是全球最好的伺服器應用作業系統之一，也是 UNIX 的免費替換器。Linux 以開放源碼、安全、高效、應用廣泛、技術支援完備而著稱。全球有超過 55% 的伺服器採用 Linux 作業系統。Apache 是世界排名第一的 Web 伺服器軟體，世界上百分之七十以上的 Web 伺服器都在使用它。PHP 是 Hypertext Pre-Processor (超文本預處理器) 的縮寫，是目前最為流行的 Web 腳本語言之一，它運行在 Web 伺服器端，根據用戶請求或伺服器端的資料產生動態網頁。MySQL 是一種多用戶、多線程的關聯資料庫系統，可用於各種流行作業系統平臺，它擁有客戶機/伺服器體系結構的分散式資料庫管理系統，支援 SQL 語法，沒有 Y2K 問題，具有功能強、使用簡單、管理方便、運行速度快、可靠性高、安全保密性強等優點。

使用 Linux + Apache + MySQL + PHP 的開發環境，花費的代價很低，卻擁有較好的性能價格比；並且由於公開源代碼，能在使用中及時發現錯誤和漏洞，進行修補；還可以依據自己的實際情況對程式進行修改，使其更符合自己的開發環境。對於學校或科研單位來說是一個非常好的選擇。

3. 系統設計

CALL 練習系統主要有 4 個模組構成：出題模組、出卷子模組、語料庫管理模組以及學生答題模組。教師主要利用 PC 登錄系統，經過認證後進行題目和試卷的製作以及語料庫的修改編輯等工作，學生則可以利用 PC 或者手機登錄系統並做練習，做練習的履歷和成績都記錄在資料庫中，便於將來教師的分析和管理的。目前 PC 上的題目類型有單項選擇題、填空題和排序題三種，手機上實現了選擇題一種題型的功能。

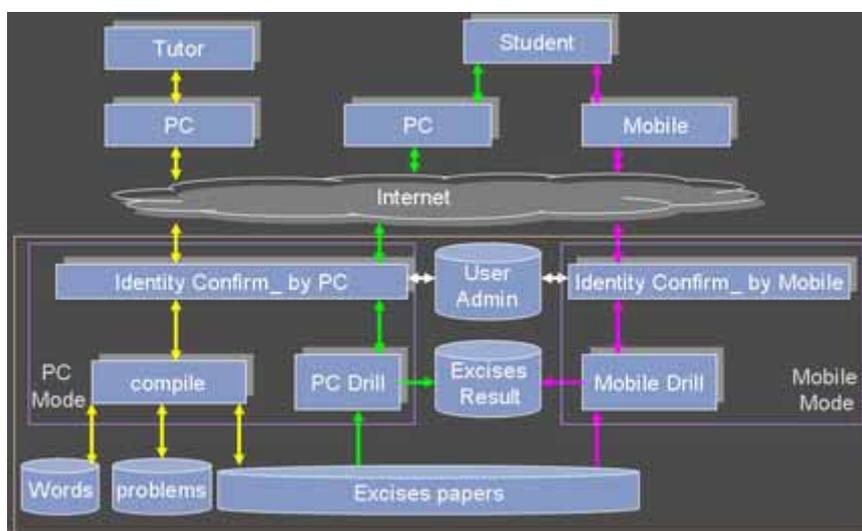


圖 6 CALL 系統的工作流程

我們在設計 CALL 系統時，有兩個比較明顯的特點，第一是分層的結構設計，第二是半自動化的出題模式。

我們把練習的介面，試卷的結構，題目的結構以及語料庫結構分別按照層次進行設計和存儲，這樣同一個單詞可以被多次應用在不同的練習題中，同一道練習題可以被多次重複應用在不同的練習試卷裏，同一份試卷可以有 PC、手機兩種不同的介面。這樣，出題的教師在出完每一課的練習題之後，可以方便的組合現有題目製作出中期考試或總復習的試卷，非常靈活得重複利用了語料庫及練習題的資源。

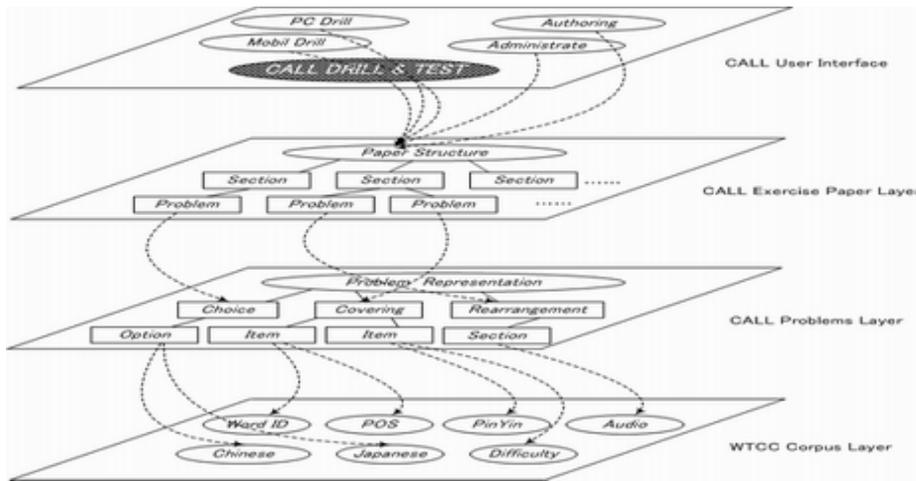


圖 7 CALL 的分層結構設計

另外一個特點是半自動化的出題方式，因為應用了語料庫中的拼音、日語解釋等資訊，在製作選擇題時，選擇一個單詞，並從“看中文選擇日文”、“看中文選擇拼音”、“看拼音選擇日文”三種類型中選擇一個之後，選擇題的題幹和正確答案就自動從語料庫中抽取出來，然後再手工填寫其餘的不正確選項之後就可以方便得出好一道題目，這種半自動化的出題方式可以大大節省教師的時間和精力。



圖 8 半自動出題的設計

4 · 系統介面



圖 9 登錄介面

編號	詞類	中文	日文	詞性	詞性	詞性	詞性	詞性
4.1	代	第一	第一	adj	adj			普通 動詞
4.2	動	叫	呼ぶ	v	call	call		普通 動詞
4.3	動	叫	呼ぶ	v	call	call		普通 動詞
4.4	動	叫	呼ぶ	v	call	call		普通 動詞
4.5	動	叫	呼ぶ	v	call	call		普通 動詞
4.6	動	叫	呼ぶ	v	call	call		普通 動詞
4.7	動	叫	呼ぶ	v	call	call		普通 動詞
4.8	動	叫	呼ぶ	v	call	call		普通 動詞
4.9	動	叫	呼ぶ	v	call	call		普通 動詞
4.10	動	叫	呼ぶ	v	call	call		普通 動詞

圖 10 語料庫管理介面



圖 11 出題模組介面



圖 12 出卷子模組介面



圖 13 答題模組(PC版)介面



圖 14 答題模組(手機版)介面

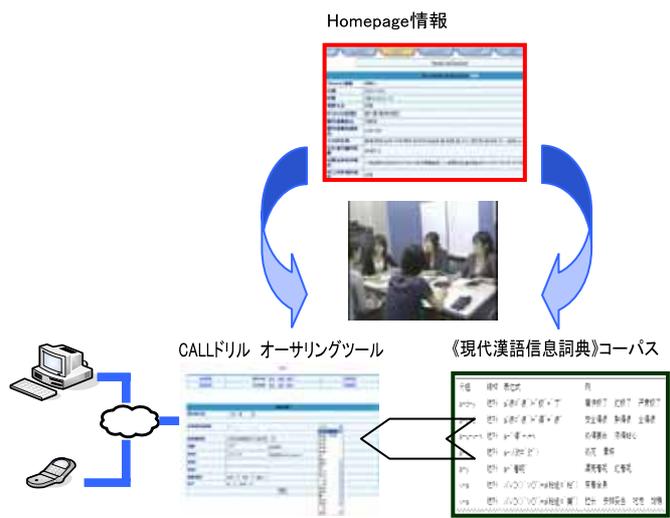
四·總結和展望

目前，CALL 練習系統經過測試已經開始在 Tutorial Chinese 系統中試運行，2005 年 4 月開始將正式投入使用。本系統具有以下幾個特點：

- (1) 利用免費、開源的開發環境建立起基於 Browser/Server 結構的網上應用程式模式，這樣不需要用戶個人電腦安裝特殊的設備和軟體，只要能夠上網就可以使用本系統的各项功能。對於管理者來說，所有的系統維護工作也都可以在網上完成。
- (2) 緊密結合基於教材的 WTCC 語料庫，可以出有針對性的練習題目，有利於提高學生的預習或復習效果。
- (3) 試卷、題目和語料庫的分層存儲結構實現了練習題的資源重複利用。
- (4) 半自動化的出題方式節省教師的出題時間和精力。
- (5) 建立了 PC 和手機兩種練習模式，使學生可以在有限的時間裏作預習或復習，並且不拘泥於空間的限制。

下一步，我們將進一步完善自動化出題的功能，不僅使正確選項自動出現，而且完成與正確選項接近的非正確選項的自動填充，進一步減輕教師的負擔。

另外，我們將把 CALL 練習系統的語料庫與 Tutorial Chinese 的 Tutor 報告相結合，使 Tutor 報告中“課堂上出現的新單詞”專



案中的內容自動增加進語料庫中，這樣我們的 CALL 練習不僅僅局限於課本單詞，而且會隨時增加反映學生興趣新單詞，實現“與時俱進”的練習效果。

圖 15 與 Tutor 報告結合的 CALL 系統

參考文獻

- [1] 砂岡和子，村上公一，“以提升交際能力為目的之漢語教學以及水準測驗芻議-早稻田大學國際漢語遠端教學四年實踐見證”，第三屆全球華文網路教育研討會論文集，p. 1-9，2003 年 10 月 24 日—10 月 26 日，臺北。
- [2] 砂岡和子等「基於 *Active Database* 與師生課本互動的國際漢語口語教學」第四屆中文電化教學國際研討會（北京）p. 328-334, 2004，7，12 數位化對外漢語教學理論與方法研究（清華大學出版社）
- [3] 劉松·砂岡和子·浦野義賴「日中台遠隔中國語口語レッスン管理プラットフォームの開発-早稻田大學 *Tutorial Chinese E-Learning Platform*-」2004，8，4 PC Conference 論文集 P142-145
- [4] LanZhang · SongLiu · JianmingWu · Yoshiyori URANO「*Chinese Language Learning Systems with 3G Mobile Phones*」Asian Info-communications Council WG3, 2004, 11 Shen z hen, China. Document No. 91
- [5] 劉松等「遠程網絡漢語口語教學管理平臺的開發」第四屆中文電化教學國際研討會（北京）2004，7，12 數位化對外漢語教學理論與方法研究（清華大學出版社）P293-299

原載「第四全球華文網路教育研討會」2005年6月 台灣

與評獎和 Self-Editing 相結合進行的網絡華文寫作教學

砂岡 和子，村上 公一，劉 松

概要:

本文將介紹早稻田大學“Tutorial 漢語”教學實行的非同期網絡寫作實踐。本系統充分應用多媒體與網路互動的特點，跟面授課同步進行，透過把口語文字化的過程，能使學生發現平時講話時所沒發現的問題，對提高口語教學有一定的效果。但是外文寫作反映個人的語言綜合能力，我們用批改標號的方式分層次分概念地進行批改。並面對日本學生不太注意評語，導致寫作練習中一再犯錯的缺點，我們採用了發榜優秀作品，以評比和 Self-Editing 方式，讓學生養成有定時上網去欣賞他人作品和老師批語的習慣。我們要介紹批改評獎的具體工作內容，希望共同探討建構自動批改華語寫作網站的未來。

關鍵詞：線上寫作， Self-Editing， 線上語法修正， 電腦評分

I 早稻田大學“Tutorial 漢語”教學的特點

早稻田大學Tutorial漢語教學是在只能使用漢語的環境下，〔4名學習者+1名教學者〕的小班制會話課程。分為遠距型（圖1）和面授型（圖2）兩種授課方式，皆是以提升華語口語能力為目標的強化課程。遠距教學在上課時利用網路電視設備，由在臺灣師範大學華語文教學研究所或北京大學對外漢語教育學院的Tutor直接為身處日本的學生上漢語口語課，把國際間的交流實踐帶進課堂裡。隨著網絡技術日益進步，遠距教學已經幾乎不受物理環境的負面因素影響，創造了必須使用漢語的環境，使學生掌握了面對不同文化時的應對能力，提高了他們學習外語的動力¹³。面授型教學則在日本早稻田大學校內進行。由早大的Tutor對學生進行面對面授課，強調教師與學生之間的互動，用啟發、推想和引導等多種教學技巧，激發學習者自我發展中文口語表達之潛力，完成語言交際任務。



圖1 遠距型教學（臺灣師範大學） 圖2 面授型教學（早稻田大學）

早大Tutorial漢語教學的首要目的是提高學生的口頭表達能力。我們經過3年多的實踐，連接在3個地區的師生進行互動教學，在真實和自然的交際情境中，學生自由地表達個人意見。為了有效地支援動態口語教學，開發了網絡教學管理平臺。在此實現了網上提交教學報告、批改寫作教學、記錄學習履歷、BBS以及網上考試和問卷調查等功能。最近又開發了基於WTCC(Waseda Tutorial Chinese Corpus)語料庫的同時能支援PC和手機的CALL練習系統¹⁴。提

¹³ 郑艳群 2004

¹⁴ 劉松 2005年

升華語口語能力的教學，已經取得了初步的成果。2004 年共進行 480 節即時面授型教學，360 節即時遠距漢語口語課，參加人數約 200 名。2005 年進而擴大師生陣容，教學上向“聽說讀寫”全面發展。本文討論的重點在於我們寫作教學的開發。

“寫”和“說”都屬於復用性技能 (reproductive skill)，比被動的“聽”和“讀”這種接受性技能 (receptive skill) 要困難得多。而且外文寫作反映個人的語言綜合能力，指導的標準包括教學理論、教育心理、教學設計、教育測量和教學評價等諸多領域。我們通過 3 個地域老師們的反覆討論，根據如下原則來進行批改寫作工作。

II 寫作教學模式

①與口語教學同步進行掌握華文規範

要學好一種外語，應該約定俗成，如果不尊重該語言的基本語法規範和概念，就無法被接受¹⁵。我們在口語教學方面，用啟發、推想和引導等教學方法來協助學生完成語言交際任務，已經取得了一定的成果。本系統充分應用多媒體與網路互動的特點，與口語教學同步進行寫作教學，給師生提供編輯工具 (authoring tool)，設計符合自己課程需要的各種寫作和管理平臺功能 (圖 5)。寫作教學跟上課的節奏一致開展。學生下課後點上 [作業] 菜單，就出現 (圖 3) 的畫面。[作業題目] 裡，已經貼出幾種寫作作業題目。按照語言水平由台灣師範大學和北京大學的老師分別提出兩種課題。題目跟剛剛上課的討論內容有關聯，促進學生完成作業的興頭，同時通過寫作功課掌握語法點，以改善口語石化的現象。

优秀作业				
Text	Level	Title	Point	Edit
第1课	中高級	自我介紹	8	view
第1课	中高級	我覺得我是一個.....的人	8	view
第1课	中高級	我覺得我是一個.....的人	9	view

作业题目				
Text	Teacher	Level	Title	Description
1	林容年	初級	我的中文學習經驗	請談談你的中文學習經驗，為什麼學習中文？在學習中文時，覺得最困難的是什麼？遇到學習上的困難時，你怎麼解決？(有趣的、難過的、特殊的事件都可以寫)100字
		中高級	我覺得我是一個.....的人	請說說你覺得自己是一個怎麼樣的人呢？你具備什麼樣的特質或能力？你的性格又是如何？(可以描述事件來輔助說明)200字
1	夏維勇	初級	我的學校	請向大家介紹一下你的學校(小学, 中学或者大学都可以)。比如说学校的名字, 大小, 你对学校以及老师和同学的看法等等。字数: 100字。
		中高級	自我介紹	假设你正在向应聘单位介绍自己的情况。你希望用人单位了解你哪方面的信息？你觉得自己的卖点(和别人相比自己的优势)是什么？如果你被聘用的话，你的目标是什么？字数: 200字。请使用惯用说法：“动词+一下”，“动词+过”和“如果……的话”。

等待评分作业						
Student	Text	Level	Title	Teacher	Submit Time	Edit

圖 3 2005 年寫作作業佈置表

學生按照自己的語言水平以及對課題的興趣程度，選擇一個 [作業題目]，在網頁上進行寫作後，投稿上去。最上面是 [優秀作業]，老師給學生的作業打分數，獲得 8 分以上的優秀作品，自動地發榜到這裡。寫作任務屬與半控制的作業，不過我們同時注意把學習者看作是信息加工的主體，盡可能給學生更大的選擇自由和寫作空間，評分的時候也注重作品的邏輯和主張。

②分層次分錯誤概念進行批改

外文寫作反映個人的語言綜合能力，可是大多批改作業往往無視分層次批改的寫作需要。我們根據不同學生的水平，給他們提出初級和中高級兩種不同要求的課題和字數的限制。

¹⁵高垣由美 2004 年

對於初級水平學生，我們強調語法規則的正確性，注重基本句子結構的合理性。要求老師出題中應包含兩句當天課文裏的關鍵語法。寫作字數限制在 100 字以內¹⁶。主要採用問答方式進行訓練。中高級水平學生，我們注重句與句之間的連貫性和所表達主題的豐富性。要求寫一段包含當天課文裏的關鍵說法。寫作字數限制在 200 字以內 (圖 4)¹⁷。

爲了指出寫作上的不同類型的毛病,我們以短句做爲批改的單位,用如下 9 種批改標號,用概念和類型分類進行批改工作,啓發學生不要再犯同樣的造句錯誤。

Text	Teacher	Level	Title	Description
第 12 課	林金錫	初級	我的暑假	請用「雖然...可是...所以」、「是...的」說說你的暑假。50 字
		中 高級	最有意義的假期	請說說你覺得最有意義的假期。100 字。
	夏維勇	初級	暑假回想	回想一件你在暑假期間的一次活動。其中一句話要使用“.....以後,就.....”這一句型。字數:約 50 字
		中 高級	我的悠長假期	暑假裏你都做什麼了?寫一段話,請用上“雖然.....可是.....”以及“.....以後,就.....”等句型。字數:約 100 字。

圖 4 2004 年後期部份作業題目

圖 5 寫作平台上的批改標號菜單

③評獎和 Self-Editing 相結合

傳統作文教學中，老師費時費力的批改完作文後，學升看完評語就隨手丟棄了，很難從寫作練習中了解自己的缺點，因此同樣的問題常一再出現。本系統自動記錄學生所犯的每一個錯誤標號。教員可整理記錄,從中可以發現學生常犯的缺點，並據以改進教學方法。

¹⁶ 2004 年定爲初級字數限制到 50 字,2005 年改爲 100 字.

¹⁷ 上同, 2004 年 100 字,2005 年增加爲 200 字.

據日本學生對外文寫作學習特點的調查分析,日本學生與美國學生相比,不太注重老師的評語,批改作業的效果更差。因為日本大學的外文教學缺乏實踐性,對大多學生來說,寫作功課沒有迫切的需要¹⁸。所以別人如何評價自己的文章,沒有太大的關心,對自己寫作的毛病也不在意。面對日本學生沒有養成反饋批改結果的學習態度,我們採用了以評比和Self-Editing方式,設計評分欄,讓寫作者互相評分的方式完成作業任務的方法。老師給學生的作業打分數的同時,發榜優秀作品(圖 3)。促使學生之間對評語進行反饋,有效地形成自力提升寫作能力的教學效果。

④中台用受限華語批改作文

我們教學的對象大多是華語初中級水平的日本學生。因此不能要求全面的作文教學,只能要求寫片段文字或短文。日本學生常寫隨筆式外文,不太注重文章結構的邏輯性。我們批改教學的重點放在華文表達概念的掌握,以及造句用詞、布局謀篇的文字表達能力。儘管我們教學的目標是華文的全面掌握,然而在具體教學階段,只能用規範化的受限漢語來教學。

所謂“受限漢語”是指全世界華人共同可接受的規範化的華文。本來是自然語言處理系統的概念,可以用于外語教學。我們網絡寫作教學,由中台兩岸的老師分級出題並進行修改(圖 4)。學生可以選擇水平和話題,投稿作業并受到中台兩岸華文、由簡繁兩種字體的批改指導¹⁹。同是華語普通話,各地有不少差異。中台兩岸的老師對學生作品的評價批語又不同。可是語言的發展總是伴隨著異化與同化的交互作用。受限華語下面的寫作教學將促進漢語的規範化²⁰。

⑤分析教學數據庫以備建構自動批改寫作網站

本系統應用多媒體與網路互動的特點,實現了不受時、空間限制的華文寫作教室。同時可以根據幾千篇批改作文檔案,建構網絡數據語料庫(圖 6)。可以在這裏獲得的數據按批改標號分類,進行錯誤類型的統計與分析。我們將對華文中的單詞、詞組和句子結構和寫作內容與系統內的模式進行比較。這些對寫作數據庫的分析有助於掌握學習華文的盲點(圖 7)。

	前期	後期	計
寫作作業	6 2 8 件	6 6 3 件	1 2 9 1 件
分類標號總數	2 万 6 0 6 4 条	未統計	

圖 6 2004 年寫作作業總數

¹⁸ 西村由起子 2000 年

¹⁹ 參考文獻 [1]

²⁰ 俞士汶、朱學鋒 1995 年

8	那里有一
9	不好个坏点的地方。无论什么时候
0	有很多人太多子，特别是星期六和星期天。因为我怕人多的地方，所以
1	好一般在节假日以外
2	时候去那里。
3	在那里我常买衣服，唱片
4	和杂货等。有一个小商店卖中国，台湾和韩国明星的东西。
5	定大家一定都不知道
6	个商店，因为这个商店不
7	有名。在日本找这些东西很难，
8	个商店大概是找这些东西最好这里很好找的地方。
9	在中国我们能随便讲价，很有趣，可是在日本，除了菜市场，鱼市场等以外，我们不
20	能讲价，
21	的东西就是很贵价格标的是多少就是多少。
22	我去原宿和she4涉谷购物过，因为这些地方跟新宿差不多
23	近。可是我觉得这些地方并不适合经常去是平常去的地方。人多极了，人很容易累子。
24	我想请朋友推荐好
25	购物的地方。

圖 7 2004 年 批改標號分類統計庫

III 批改評獎工作舉例

下面舉出具體例子來介紹批改作文和評獎工作。(圖 8) 是一件得到 9 分的優秀初級作品。我們準備修改用的標號共有 9 類(圖 9)。擔任批改工作的老師在把錯的地方用這些標號標示出語法等問題所在的同時，再修改錯處寫出評語和建議,向全體學員提示一份模範答案。關於批改用的標號，在歸類標準以及分類概念上，還有修改的餘地。不過廣義上這些應該已經能包括各種錯誤。上學期大致看來，有一定的成效。

Text	第19課
Remark by	
Level	初級
Title	假如我是政治家
Description	你曾經想過要當一個政治家嗎？如果你是一個政治家,你想為你服務的人民做些什麼？你覺得你應該培養什麼樣的能力？50字
Point	9
Content	<p>我沒有想過 D 當一個 E 政治家。但是如果我 D 當一個 E 政治家, D 加強外國語教育。我從中學到大學學習 D 八年英語。但是不會用英語跟外人說話。在學校我學習文法,但是沒有實際使用得 E 的知識 C。大部分 C 看起來 C 覺得日本的義務教育不夠 D 練習實用的會話 P。</p> <p>在國際化 D, 為了聽別人的意見又 C 陳述自己的意見, 我們先要 P 打破言語的大關。正因此 C 加強外語教育, 特別 D 加強會話能力是非常有意義的政策 21。</p> <p>評語: 內容十分豐富, 完整. 再多學習一些中文的說法就更好了。</p> <p>建議: 我從來沒想過要當政治家. 如果我要當政治家的話, 我會加強外國語教育. 我從中學到大學學習了八年的英語, 但是還不會用英語跟外人說話. 在學校我學習文法, 卻沒有實際使用的經驗. 大致上看來, 日本的義務教育不夠注重練習實用的會話。</p> <p>在國際化的時代, 為了聽別人的意見并陳述自己的意見, 我們要先打破語言的大關. 正因此 C 加強外語教育, 特別 D</p>

圖 8 批改後的優秀初級作品

批改文本: 我沒有想過 **D** 當一個 **E** 政治家。但是如果我 **D** 當一個 **E** 政治家, **D** 加強外國語教育。我從中學到大學學習 **D** 八年英語。但是不會用英語跟外人說話。在學校我學習文法, 但是 **C** 沒有實際使用得 **E** 的知識 **C**。大部分 **C** 看起來 **C** 覺得日本的義務教育不夠 **D** 練習實用的會話 **P**。

在國際化 **D**, 為了聽別人的意見又 **C** 陳述自己的意見, 我們先要 **P** 打破言語的大關。正因此 **C** 加強外語教育, 特別 **D** 加強會話能力是非常有意義的政策 **21**。

評語: 內容十分豐富, 完整. 再多學習一些中文的說法就更好了。

建議: 我從來沒想過要當政治家. 如果我要當政治家的話, 我會加強外國語教育. 我從中學到

²¹ 沒帶括號的ED等標號, 批改老師的批號。帶括號的標號, 根據建議里的模範答案, 由筆者來補充。

大學學習了八年的英語，但是還不會用英語跟外國人說話。在學校我學習文法，卻沒有實際使用的經驗。大致上看來，日本的義務教育不夠注重練習實用的會話。

這位學生雖然是初級水平，但是他寫作上的主要錯誤在於[D;殘缺]以及[C;搭配不當]上，比其他日本同學作品所常看到的[L; 缺乏邏輯]以及[P;詞語的位置不當]毛病要好的多。寫作也是把口語文字化，寫作作業能使學生發現平時講話時所沒發現的問題。所以對於提高口語教學有一定的效果。我們向批改老師進行調查時，有位老師就寫到；“這學期學生程度比較好，無論篇幅或是錯誤，都比以往的要少很多，可以看得出進步很多。”

批改標號提示： 批改標號以句為單位，可以選擇多項標號。 因為對學生的寫作只作初中水平的要求，所以並不對修辭做過多準確的要求。	
L : logic	缺乏邏輯（缺乏邏輯，語義不明白）
P : position	詞語的位置不當（介詞，動詞，賓語，助動詞，時間詞，數量詞，否定詞，主語，虛詞等的位置不對）
D : deformity	殘缺（謂語，動態，賓語，主語，語氣詞，疑問詞，虛詞等殘缺）
E : extra	多餘（連詞，助詞或主語等句法成分贅餘）
C : correspondence	搭配不當（主謂，動賓，述補，介動，複句連接等搭配不當）
R : refer	指代不明（所指人物，事務，動態，數量等前後不對應）
M : mix up	混淆（生造詞語，口語和書面語混淆，不同文體混淆）
W : wrong	字碼不對（漢字錯誤，標點符號錯誤）
O : others	其它

圖 9 批改標號表

打分數後發榜優秀學生作品，對作文教學有沒有效果？有幾位批改老師就寫到；“優秀學生作品對於其它的學生寫作應該起到了促進作用”，“或許是評分制度，感覺上學生比以前積極，這可從篇幅上看出來。”等支持意見。有一位老師指出問題說；“有無效果得視學生而定，如果學生有定時上網去欣賞他人作品的習慣，那應該對於寫作學習有所幫助。可以在期末時，要學生選一篇他們覺得寫得最好的，再修改重寫發表，以確認學生吸收了正確的寫法或改正了偏誤的地方。不然錯誤總是一犯再犯。²²”可見如果要提高寫作平台的教學效果，除了觀察學生在網路作文的表現以外，還需要給學生們討論他們作文的機會，實際觀察他們真正口語的表現如何。目前我們的教學規模較大，在時間和財政上，只好把從事寫作教學的教師隊伍跟從事口語教學的分開來組織運行。今後我們將用測驗統計的辦法在次分析這種啓發式寫作教學對口語教學的效果。然後從人機兩個方面採取措施，想辦法解決網絡學習與面授教分離的問題。

另外，本來是在寫作平台上要求學生用句型來造句，不過學生表現得不太好，常常忽略這點，教學工作開始不久，以後的出題就很少出句型造句的作業。我們將寫作平台以外，另外設計了電腦輔助練習庫平台，把基本句型的練習和寫作教學分開進行。以便保證寫作作業的自發性和語法練習的規範性。

我們將對學生寫作作品上的常犯錯誤資料進行整理，準備製作一個錯誤語料庫。這件語料庫除了作為編教材以及製作電腦輔助練習平台的基本資料以外，對未來網絡自動批改寫作網站的開發，提供很有價值的數據。隨著使用者的增多而擴大，語料庫本身會因使用頻率的增

²² 參考文獻 [1] 林金錫老師的建議

多而隨著改進、充實，以備將來開發網絡自動批改寫作站提供可貴資料。

附記:本稿發表得到自 2002 到 2004 年度日本文部科學省學術研究費(基盤研究 B-2:研究題目“漢語 TutorialSystem 的國際開發” 課題號碼 14310136:研究代表 砂岡和子)以及 2001 年早稻田大學特定課題研究(國際共同研究)“中国語チュートリアル遠隔学習システムの国際共同開発—学習プログラムの企画とデジタル教材作成—” 課題號碼 20003C-002:研究代表 砂岡和子)的撥款，為該項研究成果的一部分。

参考文献

- [1]本稿得到林金錫老師(臺灣師範大學國際華語研習所)、林容年老師(臺灣師範大學華語文教學研究所)、李利津老師((同上)、王玉老師(北京大學對外漢語教育學院)、夏維勇老師(北京大學國際關係學院研究生)、齊慧老師(早稻田大學研究生)等各位老師的寶貴意見.特此表示感謝.
- [2] 西村由起子 2000 年 英作文エラー添削(error -crrection)における同輩・自己訂正演習(peer-self-editing)の可能性-』『東洋学園大学紀要』No8
- [3] 高垣由美 2004 エクリに関する知識と実践—対照言語学的観点から—Rencontres Pédagogiques du Kansai Thème1
- [4] 俞士汶、朱學鋒 1995 年 受限漢語研究的必要性 <語文現代化論叢>第三輯
- [5] 劉松、吳劍明、浦野義賴 2003 年基於 PHP 與 MySQL 的互動型漢語教學支持平臺的研究與實踐 第三屆全球華文網路教育研討會
- [6] 劉松、砂岡和子、浦野義賴 2005 年 日本的手機在中文教育上應用的問題與實踐台灣 ITCSL 予稿集印刷中
- [7] 劉松、砂岡和子、浦野義賴 2005 年 與語料庫相結合的中文 CALL 練習系統的開發 4th International Conference on Internet Chinese Education (ICICE 2005) 台灣予稿集印刷中
- [8] 村上公一、砂岡和子、劉松 2005 年 Computerized Adaptive Testing (CAT) 方式的網路中文口語測驗的開發 4th International Conference on Internet Chinese Education (ICICE 2005) 台灣予稿集

原載「4th International Conference on Internet Chinese Education (ICICE2005)」

2005 年 6 月台灣

日本的手機在中文教育上應用的問題與實踐

劉松 砂岡和子 浦野義賴

摘要

Internet 的急速發展使遠端教育成爲語言教育領域中不可忽視的一部分。近年來，日本利用手機連接網路的用戶迅猛增加。使用輕巧便利的手機，在任何時間、任何地點都可以方便地連接網路，這樣的優點使它同樣在中文教育領域中可以大顯身手。然而，日本的手機在多語言對應上的局限性，以及多家手機運營商之間的標準不統一，使開發手機教材及網站時，需要解決諸多 PC 上沒有遇到的問題。

早稻田大學 2000 年開始了名爲“Tutorial Chinese”的小班口語強化中文教學模式，並在 2003 年開始利用網路平臺進行教學資訊的共用和語料庫的收集與分析。2004 年，我們又開發了基於語料庫的 CALL 練習系統，這套系統同時支援 PC 與手機兩種端末。本文將介紹日本的手機在中文教育上應用時會遇到的問題，以及我們開發 CALL 練習系統時的具體實踐。

關鍵字

Tutorial Chinese, CALL, 語料庫, 手機, 多語言支援

一·序言

隨著人類在電腦多媒體技術、網路技術、資料庫技術、人工智慧等方面的快速發展，能夠以電腦或智慧型終端作爲客戶機，通過 Internet 獲得長期的、高效的、即時方便的終身學習機會，已經成爲新一代教育的發展方向。遠端教學利用現代技術手段，打破地域和時間的界限，不僅使講授更爲生動，而且使遠距離的雙向交流成爲可能。語言教育也勢必受這一趨勢的影響。

語言教育的目標是培養學生的意思表達能力，要達到這種能力除了掌握語法以外，還需要通過實際交流來鍛煉。目前，海外漢語學習者的數量急劇增加，由於缺少與母語話者的交流機會，造成了學生經過長時間的學習卻不敢開口說話的現象，在這種情況下，早稻田大學於 2000 年開始進行了 Tutorial Chinese 教學，採用了一名 Tutor(輔導老師)指導四名學生的少人數口語會話強化班的模式，分爲遠端型(圖 1)和面授型(圖 2)兩種。遠端型教學在上課時利用網路電視設備，由在臺灣或大陸的 Tutor 同期即時地指導日本學生；面授型教學則在日本早稻田大學校內進行，由早大的 Tutor 對學生進行面對面授課。目前 Tutor 分別由北京大學，臺灣師範大學及早稻田大學的母語話者組成，一年共進行 220 節即時遠端漢語口語課，參加人數約 200 名。



圖 1 遠端型教學



圖 2 面授型教學

為了更加有效地支援這種新型的教學模式，我們于 2003 年開發了 Tutorial Chinese 教學管理平臺。由於 Tutorial Chinese 重視面對面的現場交際口語教學，所以該平臺也與一般的以靜態課程為中心的不同，而是以動態口語教學為中心，發揮網路教學的特點，把動態課堂上的資訊及時地回饋給學生，形成連接 3 個地區師生的強大用戶聯盟網路。第一階段的開發實現了網上提交教學報告、批改作業、記錄學習履歷、BBS 以及網上考試和問卷調查等功能。2004 年，我們又開發了基於課文辭彙語料庫的同時支援 PC 和手機兩種端末的 CALL 練習系統。由於日本的手機在多語言對應上的局限性，使開發手機教材及網站時需要解決諸多 PC 上沒有遇到的問題。本文將介紹日本的手機在中文教育上應用時會遇到的問題，以及我們開發 CALL 練習系統時的具體實踐。

二．日本手機的現狀及在中文教育上應用的問題

根據日本總務省《2004 年版情報通信白書》，截止到 2003 年末，6 歲以上利用 Internet 的用戶有 7,730 萬人，其中利用個人電腦（PC）上網的用戶有 6,164 萬人，利用手機、PHS 等移動端末上網的用戶有 4,484 萬人。手機在日本已經成為上網的一種不可替代的工具。然而，在語言教育的 E-learning 系統中聲音、圖像之類的多媒體內容占非常重要的地位，過去的手機由於技術的限制，只能提供通話和收發 e-mail 等有限功能，通信的速度也非常慢。2001 年，日本開始提供 3-G 手機，實現了高速通信並支援音樂、圖像、電視等多媒體檔的播放，目前的主流方式是 DoCoMo 的 W-CDMA 和 AU 的 CDMA2000。2003 年 11 月，AU 又開始了 Packet 定額制的服務，只要每月交固定的使用費，通過手機上網流覽網頁，收發 email，下載音樂，電視節目，遊戲等都可以不再交額外的通信費。有了這樣的服務，通過手機上網流覽教學網站，下載多媒體教材都可以免除費用過高的顧慮，使開發手機版的教學網站成為可能。

日本手機在訪問中文教育網站時還會遇到字體顯示的問題，過去第二代日本手機的作業系統中多採用 Shift-JIS、EUC 等日語字體，即使是 3-G 手機，雖然支持 Unicode，但是由於機體本身存儲量的限制，大多數只是載入了日語部分的 Unicode，在這樣的情況下流覽中文網站時會出現亂碼。要解決字體顯示的問題，目前一般有三種方法：

- 與手機生產商合作，使手機載入中文字體。這種方法現在還很難實現，因為手機的存儲量比較有限，而中文字體至少要佔據幾兆到十幾兆的存儲量，在沒有看到非常大的使用需求的情況下（畢竟在日本，流覽中文內容的人還是少數）生產商是不會這樣做的。
- 製作可下載的應用程式，並在其中內嵌中文字體。這種方法可以使用內嵌的字體來顯示中文，但是需要的技術和成本比較高，在日本主要有 JAVA（DoCoMo 公司和 Vodafone 公司）和 BREW（AU 公司）兩種方式來開發，JAVA 的環境速度上有些欠缺，BREW 則需要與

AU 簽訂合約，經過授權才可以使用。

- 不使用字體，而是把中文字變成圖像並在手機裏顯示。這種方法比較容易實現，但是需要預先把字變成圖片，而且由於要傳送圖片，通比較大容易造成延遲。隨著 3-G 手機的普及，通和速度的問題已經逐步得到解決。

我們的教學系統採用了第三種顯示中文的方公開源碼的函數 GD 將 Unicode 的中國文字變為 Gif 用 Hash 函數為圖片命名，這樣就可以自動將中文字片以及自動讀取並顯示圖片。這種方法還可以把中的其他字體轉換成圖片，所以利用該方法還可以在機上顯示諸如韓文、法文、阿拉伯文……等等中包含的文字。



法，利用圖，並利轉換成圖 Unicode 日本的手 Unicode

三·CALL 練習系統的實踐

1·系統結構

本系統基于 Browser/Server 結構，即瀏覽器/ 服务器模式 (图 3)。用戶端，只要有一台安裝有操作系統 WINDOWS 9x (2000) 或 XP，可以上網，用 IE 等瀏覽器來瀏覽 Internet 資源就可以了，原則上不需要另外安裝任何軟件。與程式升級以及資料更新相關的工作放在伺服器端，Web 伺服器連接用戶流覽器與後臺資料庫，通過表單等形式在用戶與資料庫之間進行資料交換。這種基於 Web 的資料庫方式較好地結合了網路傳輸速度快的優勢和資料庫存取資料及安全機制上的特長。

图 3 手機上顯示多種語言

2·系統開發環境

教學平臺的開發採用了目前在教學研究單位非常通用的 LAMP 模式 (Linux + Apache + MySQL + PHP)，全部是免費的開源軟體。Linux 是全球最好的伺服器應用作業系統之一，也是 UNIX 的免費替換器。Linux 以開放源碼、安全、高效、應用廣泛、技術支援完備而著稱。全球有超過 55% 的伺服器採用 Linux 作業系統。Apache 是世界排名第一的 Web 伺服器軟體，世界上百分之七十以上的 Web 伺服器都在使用它。PHP 是 Hypertext Pre-Processor (超文本預處理器) 的縮寫，是目前最為流行本語言之一，它運行在 Web 伺服器端，請求或伺服器端的資料產生動態網頁。一種多用戶、多線程的關聯資料庫系於各種流行作業系統平臺，它擁有客戶器體系結構的分散式資料庫管理系統，語法，沒有 Y2K 問題，具有功能強、使管理方便、運行速度快、可靠性高、安強等優點。

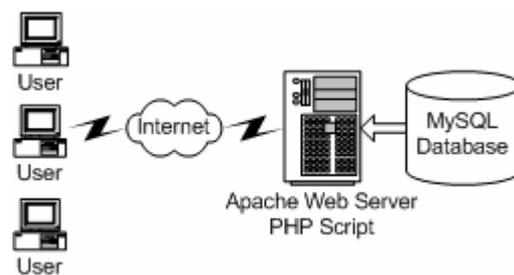


图 4 系統結構及開發環境

使用 Linux + Apache + MySQL + PHP 環境，花費的代價很低，卻擁有較好的性比；並且由於公開源代碼，能在使用中及時發現錯誤和漏洞，進行修補；還可以依據自己的實際情況對程式進行修改，使其更符合自己的開發環境。對於學校或科研單位來說是一個非常好的選擇。

3·系統設計

CALL 練習系統主要有 4 個模組構成：出題模組、出卷子模組、語料庫管理模組以及學生答題模組。教師主要利用 PC 登錄系統,經過認證後進行題目和試卷的製作以及語料庫的修改編輯等工作，學生則可以利用 PC 或者手機登錄系統並做練習，做練習的履歷和成績都記錄在資料庫中，便於將來教師的分析和管理的。目前 PC 上的題目類型有單項選擇題、填空題和排序題三種，手機上實現了選擇題一種題型的功能。

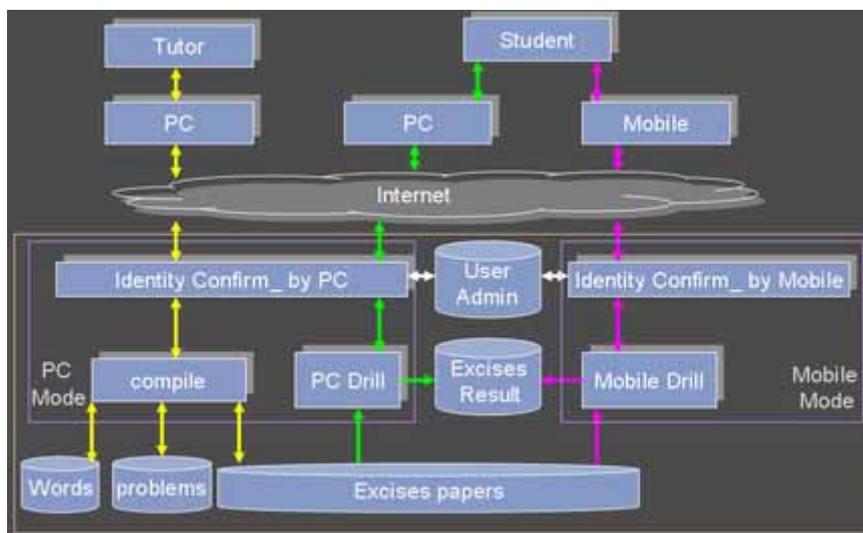


圖 5 CALL 系統的工作流程

我們在設計 CALL 系統時，有兩個比較明顯的特點，第一是分層的結構設計，第二是半自動化的出題模式。

我們把練習的介面，試卷的結構，題目的結構以及語料庫結構分別按照層次進行設計和存儲，這樣同一個單詞可以被多次應用在不同的練習題中，同一道練習題可以被多次重複應用在不同的練習裏，同一份試卷有 PC、手機兩種不同這樣，出題的教師一課的練習題之方便的組合現有題中期考試或總復習非常靈活得重複利用及練習題的資

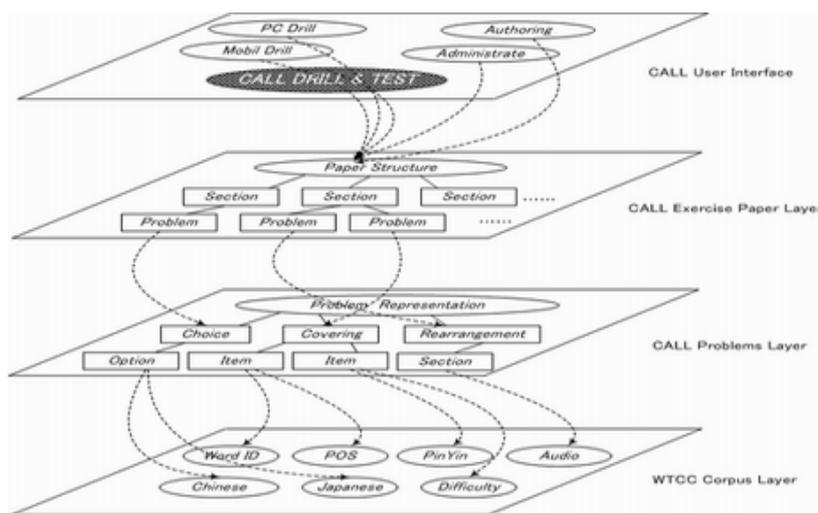


圖 6 CALL 的分層結構設計

另外一個特點化的出題方式，因語料庫中的拼音、等資訊，在製作選擇一個單詞，並文選擇日文”、選擇拼音”、

練習試卷可以有的介面。在出完每後，可以目製作出的試卷，用了語料源。

是半自動為應用了日語解釋選擇題時，從“看中“看中文

選擇一個之後，選擇題的題幹和正確答案就

自動從語料庫中抽取出來，然後再手工填寫其餘的不正確選項之後就可以方便得出好一道題目，這種半自動化的出題方式可以大大節省教師的時間和精力。



圖 7 半自動出題的設計

4· 系統介面



圖 8 登錄介面

ID	級別	年次	詞類	詞	次數	詞性	備註
401	4	第一	一	ping	1	動詞	
402	4	第一	動詞	v	1	動詞	
403	4	第一	動詞	v	1	動詞	
404	4	第一	動詞	v	1	動詞	
405	4	第一	動詞	v	1	動詞	
406	4	第一	動詞	v	1	動詞	
407	4	第一	動詞	v	1	動詞	
408	4	第一	動詞	v	1	動詞	
409	4	第一	動詞	v	1	動詞	
410	4	第一	動詞	v	1	動詞	

圖 9 語料庫管理介面



圖 10 出題模組介面



圖 11 出卷子模組介面



圖 12 答題模組(PC 版)介面

圖 13 答題模組(手機版)介面

四·總結和展望

目前，CALL 練習系統經過測試已經開始在 Tutorial Chinese 系統中試運行，2005 年 4 月開始將正式投入使用。本系統具有以下幾個特點：

- (6) 利用免費、開源的開發環境建立起基於 Browser/Server 結構的網上應用程式模式，這樣不需要用戶個人電腦安裝特殊的設備和軟體，只要能夠上網就可以使用本系統的各项功能。對於管理者來說，所有的系統維護工作也都可以在網上完成。
- (7) 緊密結合基於教材的 WTCC 語料庫，可以出有針對性的練習題目，有利於提高學生的預習或復習效果。
- (8) 試卷、題目和語料庫的分層存儲結構實現了練習題的資源重複利用。
- (9) 半自動化的出題方式節省教師的出題時間和精力。
- (10) 建立了 PC 和手機兩種練習模式，使學生可以在有限的時間裏作預習或復習，並且不拘泥於空間的限制。

下一步，我們將進一步完善自動化出題的功能，不僅使正確選項自動出現，而且完成與正確選項接近的非正確選項的自動填充，進一步減輕教師的負擔。同時，爲了對應日本各個手機生產商的機型，我們將在現有系統的通用化課題上作進一步研究。

附記:本稿發表得到自 2002 到 2004 年度日本文部科學省學術研究費(基盤研究 B-2:研究題目“漢語 TutorialSystem 的國際開發” 課題號碼 14310136:研究代表 砂岡和子)以及 2001 年早稻田大學特定課題研究(國際共同研究)“中国語チュートリアル遠隔学習システムの国際共同開発—学習プログラムの企画とデジタル教材作成—” 課題號碼 20003C-002:研究代表 砂岡和子)的撥款，爲該項研究成果的一部分。

參考文獻

- [6] 砂岡和子，村上公一，“以提升交際能力爲目的之漢語教學以及水準測驗芻議—早稻田大學國際漢語遠端教學四年實踐見證”，第三屆全球華文網路教育研討會論文集，p.1-9，2003 年 10 月 24 日—10 月 26 日，臺北。
- [7] 平成 16 年版 情報 通信 白 書
<http://www.johotsusintokei.soumu.go.jp/whitepaper/ja/h16/>
- [8] 砂岡和子等「基於 Active Database 與師生課本互動的國際漢語口語教學」 第四屆中文電化教學國際研討會(北京) p.328-334,2004，7，12 數位化對外漢語教學理論與方法研究(清華大學出版社)
- [9] 劉松·砂岡和子·浦野義賴「日中台遠隔中國語口語レッスン管理プラットフォームの開発-早稻田大學 Tutorial Chinese E-Learning Platform-」 2004，8，4 PC Conference 論文集 P142-145
- [10] LanZhang·SongLiu·JianmingWu·Yoshiyori URANO「Chinese Language Learning Systems with 3G Mobile Phones」 Asian Info-communications Council WG3,

2004,11 Shen z hen, China. Document No. 91

- [11] 劉松等「遠程網絡漢語口語教學管理平臺的開發」 第四屆中文電化教學國際研討會
(北京) 2004,7,12 數位化對外漢語教學理論與方法研究(清華大學出版社)P293-299

原載「二十一世紀中文項目營運策略國際教學研討會」2005年6月台灣

Mobile CALL-drill and Web-test Tools for the Communicative Chinese Conversation Lessons

Song LIU

Doctor Course, Global Information and Telecommunication Studies, Waseda Univ.
Tokyo, Japan
Liusong@suou.waseda.jp

Kazuko SUNAOKA

Professor, School of Political Science & Economics, Waseda Univ.
Tokyo, Japan
ksunaoka@waseda.jp

Yoshiyori URANO

Professor, Global Information and Telecommunication Institute, Waseda Univ.
Tokyo, Japan
urano@waseda.jp

Abstract

The design and implementation of mobile CALL-drill and Web-test authoring tools is described. This system integrates modern networking, computer and mobile telephone technologies with linguistics and language pedagogy. This paper also aims to introduce the construction of the Chinese Corpus WTCC (Waseda Tutorial Chinese Corpus) which supports our Chinese conversation classes and the development of the mobile CALL-drill and the Web-test that are based on these Corpus resources.

Keywords

Tutorial Chinese, Computer Assisted Language Learning (CALL), Mobile telephone, Corpus

1. Introduction

As of 2004 October, the Chinese language CCDL (Cross-Cultural Distance Learning) project at Waseda University affiliates with Peking University, National Taiwan Normal University and administrates thirty tutorial classes of four students each. We place native speakers in Japan, Beijing and Taipei as tutors to conduct Chinese conversation lessons by communicative approach. This education mode is called Tutorial Chinese, which has two education types: Distance Mode and Face to Face Mode.



Fig.1 Distance



Fig. 2 Face to Face

In order to support this new language education mode, we have developed Tutorial Chinese Platform (TCP) in 2003. We have established an education management server on campus to administrate all the classes that stretch over multiple countries. Being different from other distance education platform based on normal static class, Tutorial Chinese Platform is based on interactive dynamic conversation lesson. Through this platform, the education information in class will be fed back to teacher and students by tutors soon after the lesson. Up to now, we have already developed the functions such as report submission, homework correction, gathering studying log, BBS, web test and web questionnaire, etc.

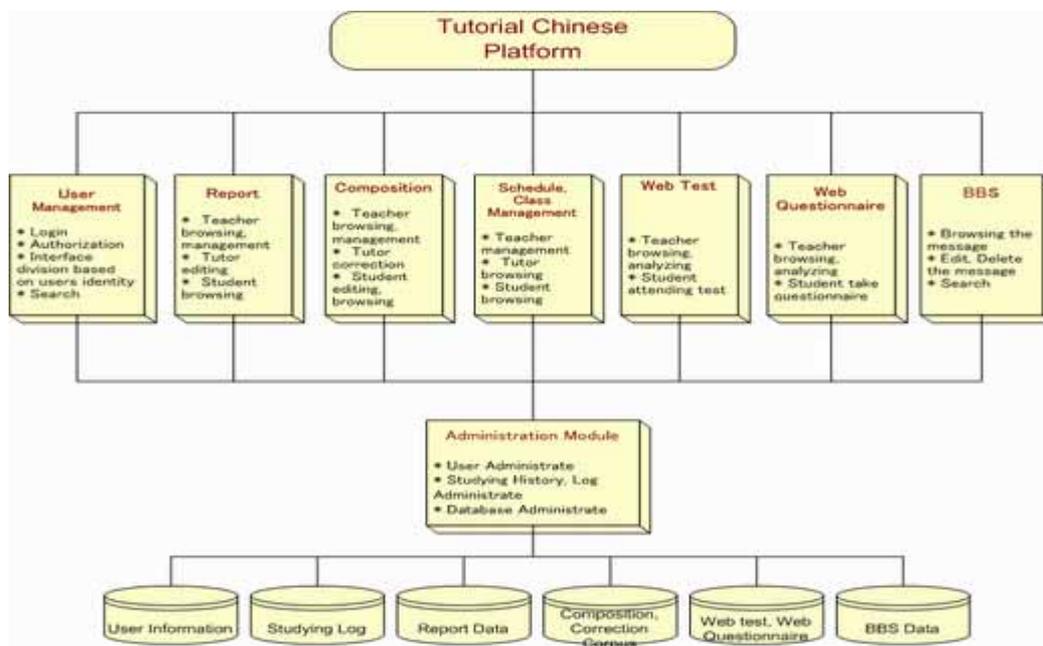


Fig. 3 TCP Module Structure

As one major assistant part of Tutorial Chinese Platform, we have constructed CALL-Drill and Web-Test tool from May, 2004. This tool has the following advance features:

- (11) Web based interface and transmission mechanisms.
- (12) Combined with Waseda Tutorial Chinese Corpus.
- (13) Constructed the mechanism that saved words corpus, problems and exercise papers as different layers, to make the authoring tool more flexible and reusable.
- (14) Constructed ubiquitous drilling environment with the mobile telephone technology.

Integrating these various features, we have developed corpus-based, flexible CALL-Drill and Web-test authoring tools for teachers and ubiquitous drilling environment for students.

In this paper, we will introduce the content and construction process of Waseda Tutorial Chinese Corpus; the implementation of CALL-Drill Authoring tool and Web-test interface for PC and Mobile telephone; and the directions for future research plan.

2. WTCC (Waseda Tutorial Chinese Corpus)

Chinese Corpus named WTCC (Waseda Tutorial Chinese Corpus) is based on three sources: The Grammatical Knowledge-base of Contemporary Chinese corpus by Beijing University (approx. 10000 words), HSK (Hanyu Shuiping Kaoshi, approx. 8000 words) and Japanese Basic University Educational Vocabulary corpus (approx. 3000 words). We have added translations in Japanese, pronunciation signs in Pinyin and the difficulty level tags to these basic corpus sources and listed it on the class website as

Vocabulary Corpus. Students take the computerized achievement tests based on this Vocabulary Corpus. Furthermore, we have added 8000 vocabularies from the class textbook to this Vocabulary Corpus information to make CALL drills to check the daily improvement of the students.

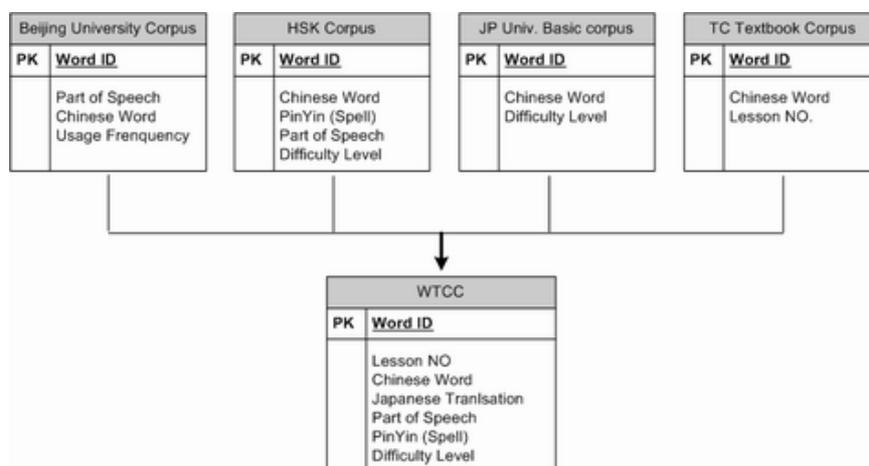


Fig. 4 WTCC Construction

The construction process consists of 5 steps:

- (1) Split sentences in TC textbook into words and morpheme.
- (2) By the way of database searching of Beijing University Corpus, HSK corpus and Japanese Basic University Educational Vocabulary corpus, to get the words' POS, Pinyin spell and difficulty level information.
- (3) Select the target word from polyphonic or multi-meaning words based on TC textbook.
- (4) Get the words' Japanese meaning using auto translation tools and checking by hand.
- (5) Construct the corpus into Internet database.

ID	Lesson	Chinese	Japanese	POS	Spell	Diff. Level
618	1	大	大きい	adj	da4	1
729	1	...过	...したことがある	pre.	guo4	1
730	1	参加	参加する	v	can1 jia1	1
731	1	多大	どのくらい、いくつ	num	duo1 da4	1
732	1	房间	部屋	n	fang2 jian1	1
733	1	高兴	喜ぶ	v	gao1 xing4	1
734	1	高兴	うれしい	adj	gao1 xing4	1
735	1	功课	授業	n	gong1 ke4	3
736	1	号码	番号	n	hao4 ma3	2
737	1	机会	チャンス	n	ji1 hui4	1

Fig. 5 WTCC Structure

3. Implementation of CALL-Drill and Web-Test Tools

3.1 Development Environment

We have taken B/S (Browser/Server) structure in this system. In the server side, web server has been used as

the bridge between user and database server. Through *Forms*, users' requests will be sent to database server and then feedback the results to users. In the client side, no special software installs are required and it ensures that fairly standard computers will be ready to use the system.

The LAMP (Linux, Apache, MYSQL, PHP) mode has been used as the development environment in this system, which has many advantages: (a) the entire technology stack is available through open-source; (b) it works fine for most applications; (c) it is easy to learn; (d) it allows one to build a web application quickly; and (e) there are many open source code examples available that make creating an entire web application even easier. The combination of Linux, Apache, PHP and MYSQL has formed an ideal network database environment. So this kind of technology stack has rapidly being adopted for building web applications in these days.

3.2 System Design

The CALL-Drill & Web-Test tools have four modules: Corpus Management, Problems Authoring, Exercise Paper Authoring, Drill and Administration Interface. Different from normal system, we have constructed the mechanism that words corpus, problems and exercise papers are saved as different layers. Therefore, when the teacher wants to construct exercise paper, she/he only needs to select problems that already constructed and saved. The same problem can be utilized multi-times in different papers. Based on this mechanism, the authoring tool is more flexible and reusable.

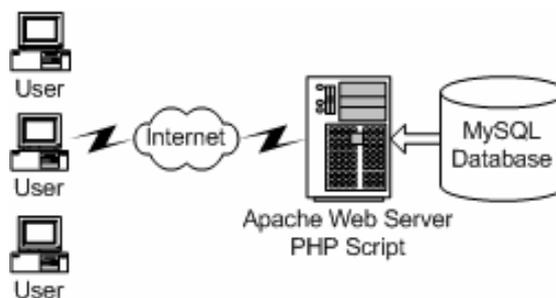


Fig.6 Development Environment

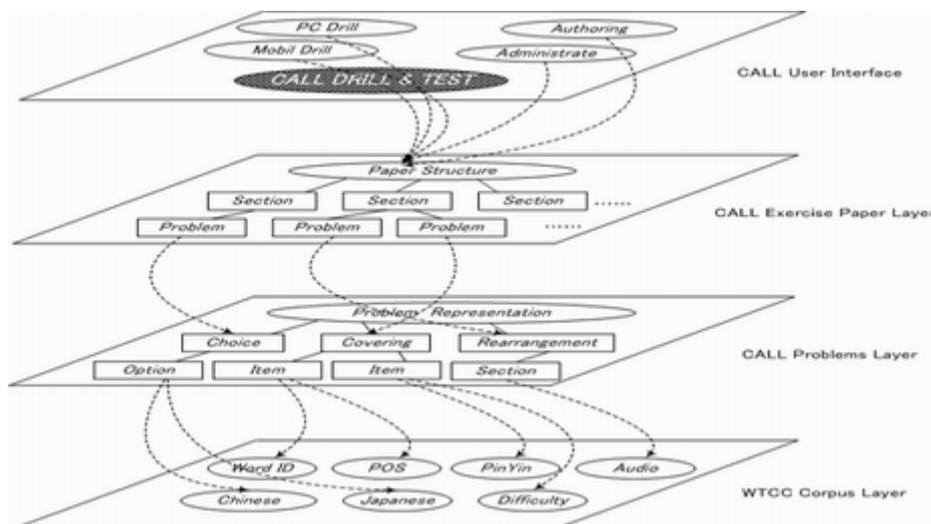


Fig.7 Layer Mechanism

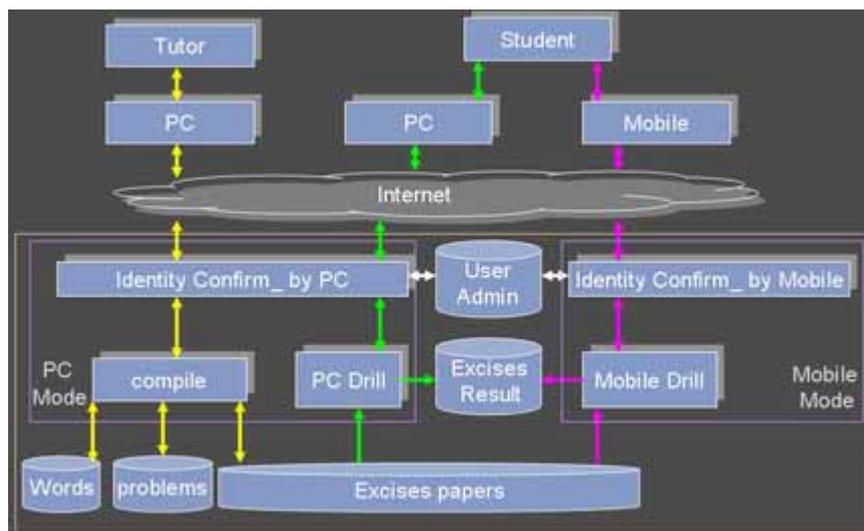


Fig.8 System Structure

3.3 Interface



Fig.9 Login

序号	词频	词性	词义	例句	例句	例句	例句	例句	例句
413	4	副	第一						
414	4	副	第二						
415	4	副	第三						
416	4	副	第四						
417	4	副	第五						
418	4	副	第六						
419	4	副	第七						
420	4	副	第八						
421	4	副	第九						
422	4	副	第十						

Fig.10 Corpus Management



Fig.11 Problem Authoring



Fig.12 Paper Authoring



Fig. 13 Web-test (PC)

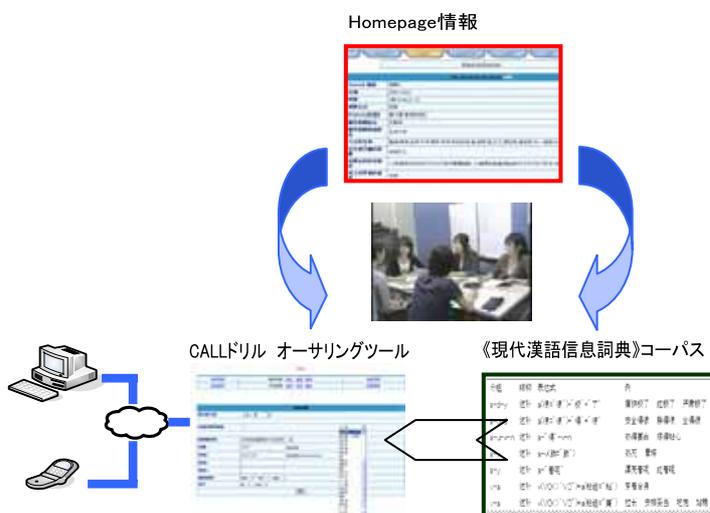


Fig.14 Web-test (Mobile)

4. Conclusions and Future Work

This mobile CALL-Drill and Web-Test tools has many advantages which distinguish it from conventional software. Among these advantages are: (a) a web-based user interface for students and teachers to facilitate system access, (b) layer architecture to make authoring flexible, (c) based on textbook corpus to make the drill effective, (d) ubiquitous environment to make the drill usable whenever, wherever users are.

In the next step, we will develop new functions as follows: We hope to build an automatic registry system that adds new vocabularies used in the live tutorial lessons into the Corpus automatically. New vocabularies will be reported by the tutors after each class and will be registered in the server. The server refers to the WTCC Corpus database and if it is already registered, it will provide the educational data for further study. If it is not yet registered, it will be acknowledged as a new entry and the server makes an additional entry. This additional information will be circulated among the students through their computers and mobile phones. This whole process will be computerized.



References

砂岡和子,村上公一,“以提升交际能力为目的之汉语教学以及水平测验刍议—早稻田大学国际汉语远程教学四年实践见证”,第三届全球华文网络教育研讨会论文集,p.1-9,2003年10月24日—10月26日,台北。

砂岡和子等「基于 Active Database 与師生課本互動的國際漢語口語教學」 第四屆中文電化教學國際研討會(北京) p.328-334,2004,7,12 數字化對外漢語教學理論與方法研究(清華大學出版社)

劉松,吳劍明,浦野義賴,“基于 PHP 与 MySQL 的互动型汉语教学支持平台的研究与实践”,第三届全球华文网络教育研讨会,CDROM,2003年10月24日—10月26日,台北。

劉松等「遠程網絡漢語口語教學管理平台的開發」 第四屆中文電化教學國際研討會(北京)p.293-299,2004,7,12 數字化對外漢語教學理論與方法研究(清華大學出版社)

劉松·砂岡和子·浦野義賴「日中台遠隔中國語口語レッスン管理プラットフォームの開發-早稻田大学 Tutorial Chinese E-Learning Platform-」 p.142-145,2004,8,4 PC Conference

LanZhang·SongLiu·JianmingWu·Yoshiyori URANO「Chinese Language Learning Systems with 3G Mobile Phones」 Asian Info-communications Council WG3,2004,11 Shenzhen,China

原載

「The 18th Pacific Asia Conference on Language, Information and Computation」
2004年12月

日中台遠隔中国語口語レッスン管理プラットフォームの開発 -早稲田大学 Tutorial Chinese E-Learning Platform-

劉松・砂岡和子・浦野義頼

はじめに

2000年から開始した早稲田大学の Tutorial Chinese 科目は、チューター1名が学生4名を指導する少人数クラスを採用し、口語強化の授業を行う教学モデルである。遠隔型(図1)と対面型(図2)の2タイプがあり、遠隔型はテレビ会議システムを利用し、中国大陸および台湾のチューターが現地から日本の早稲田大学の学生にリアルタイムで授業を配信する。一方、対面型は早稲田大学校内で、大学のチューターが学生と直接顔を合わせて対面式授業を行う。チューターは全員、北京大学、台湾師範大学および早稲田大学に在籍する言語の専門訓練を受けた母語話者から構成される。



図1 遠隔型



図2 対面型

この新しい教育モデルをより効率的にサポートするために、我々は2003年から Tutorial Chinese 授業の管理プラットフォームを開発した。Tutorial Chinese はリアルタイムの対話型口語教学がメインであるため、本プラットフォームは通常の静的素材を主にする E-Learning と異なり、ダイナミックな口語教育を中心に据え、ネットワーク教育の特徴を發揮して、教場での情報を直ちに学生へフィード・バックし、三地域の教師と学生が強大なユーザー連合を構成するプラットフォームを目指した。現在、すでにオンラインのレポート提出や、宿題の添削、学習履歴の記録、ならびにウェブテスト・ウェブアンケートなどの機能を実現している。このプラットフォームにより、多くのペーパーによる教務作業が、ネット上の電子管理へと転移し、データの提出、保存と統計など、資源と時間を節約できた。同時に、データベースに保存されたレポートや宿題の添削情報は、絶好なコーパス資源となり、今後の教育研究およびテキスト編纂に貴重な資料を提供することが可能になる。

本論文は Tutorial Chinese 授業プラットフォームの開発経緯、主要モジュールのロジカル構造とコア技術を紹介し、また将来的なシステム拡張についても触れる。

1. システムの設計

プラットフォームの開発は「LAMP」モード(Linux + Apache + MySQL + PHP)を採用している(図3)。「LAMP」はオープンソース・ソフトウェアが普及する中で、注目を集めるソフトウェア群である。従来は規模の小さい Web アプリケーションで利用されていたが、最近は大規模なもの(エンタープライズ用

途) まで適用範囲が広がりつつある。

Linux は周知のように、自由に再配布することのできる、独立した Unix 系オペレーティングシステム(OS)である。Apache は NCSA HTTPd 1.3 より派生した HTTPD サーバで、パフォーマンスの良さと、モジュールにより機能を追加・変更できる拡張性の高さ、費用がかからず使用制限がないこと、などから多くのサイトで使われるようになってきている。現在インターネットにおいて世界一のシェアを誇る。MySQLは大規模に展開されるソフトウェアへの組み込みにも対応するように設計されている。最大の特徴は、検索処理や更新処理を高速にできることにある。PHP (Hypertext Preprocessor) は広く使用されている汎用スクリプト言語で、HTML中にスクリプトを記述出来るという特性によりWeb開発に適している。多くの構文は C、Java、Perl 言語から転用したもので、簡単に習得することが出来る。

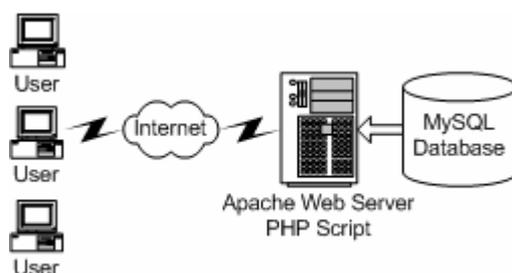


図3 システム環境

Linux オペレーティングシステム、Apache web サーバ、MySQL データベース、PHP プログラミング言語で動作することによって、強健性と安全性を備えた高速でダイナミックな Web サイトを構築することができる。

プラットフォームはユーザー認証、クラススケジュール管理、レポート、宿題・添削、ウェブテスト、ウェブアンケート、BBS など、幾つかのモジュールから構成されている。モジュール構成図を、図4に示す。

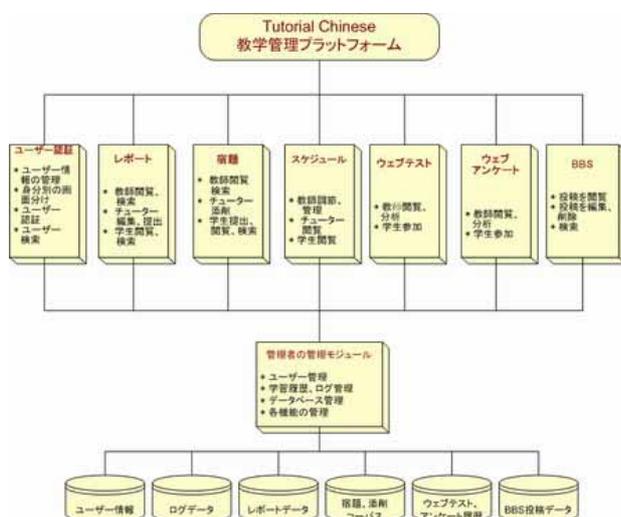


図4 システムモジュール構成

Tutorial Chinese プラットフォームは、システムがロジカルで三つのユーザー階層から構成され、教師、チューター、学生がそれぞれ権限により、特定のインタフェースを設計する(図5)。1 駒の授業が終了するごとに、チューターと学生間で宿題の提出と添削などの作業を行い、チューターと教師は授業レポートの提出と分析作業を行う。データベースにはユーザーのプロファイル、学生の学習履歴、チューターの作業履歴、宿題・添削コーパス、教学レポートなどの情報が全記録されており、教学研究に貴重な資料を提供する。

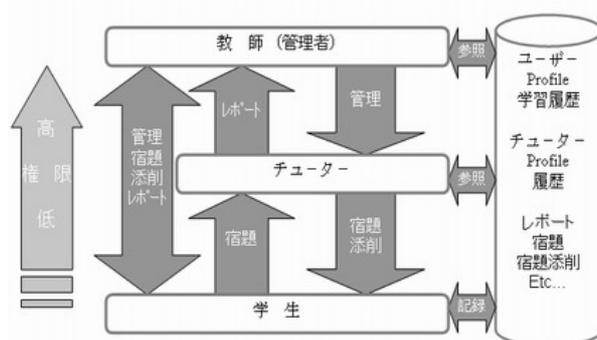


図5 システムロジカル階層

2. システムの実現と技術

2003年6月、システム提案から9月までに第一段階の開発を完成、2003年10月から2004年1月までの半年間で、約200名の学習者に対し、220時間のリアルタイム口語授業のサポートを実施した。



図6 システムのログイン画面



図7 システムのメイン画面

プラットフォームにダイナミックな口語教育を中心に据え、日中台3地域にわたる教場の情報を直ちに学生へフィード・バックしなければならないため、システム実現には独特な技術手段を採用した。以下はその中の事例である。

2.1 全面的に Unicode 技術を採用 (フル・ユニコード技術)

対外中国語の教育システムにとって、中国語の表示と入力は一解決困難な課題である。かつて日本のユーザーに対する中国語教育システムの解決案は、日本語 shift-JIS のエンコードへマッピング(mapping)して日本語 OS 上で中国語の表示問題を解決するものであった。しかしこの方法では入力時、特殊なツールを利用しなければならず、同時に非 shift-JIS コード利用ユーザーにとって、実現が難しい。以上の問題を考慮した上で、当システムは Unicode でこれらの問題を解決する。

Unicode は新しい概念ではなく、多言語表示の解決手段としてすでに利用されている。しかしソフト

ウェア/ウェブサイトシステム中で全面的に Unicode を使用するの珍しいケースである。当システムはサーバー——OS、Apache、MySQL で Unicode を支持するため、プログラム作成——各種言語コードと Unicode 間の転換、ライブラリ (library) の使用まで、Unicode に対する全面的な支持を実現した。

フル・ユニコード技術を採用することによって、全面的に日本語、簡体字中国語、繁体字中国語と英語など、多種類の文字コードをサポートし、多地域の教育と受講者管理に至便なサービスを提供している。

2.2 Richtext フォーマットの宿題編集 (図 8)

Tutorial Chinese 科目では学生に定期的に作文提出を課しており、教師が作文を添削し、学生にフィード・バックする。当システム開発以前は、学生が Word などテキストのエディターを利用して宿題を提出していたため、教師の添削作業も自身のコンピュータ上でテキストエディターを用い、教師と学生間のファイル交換は電

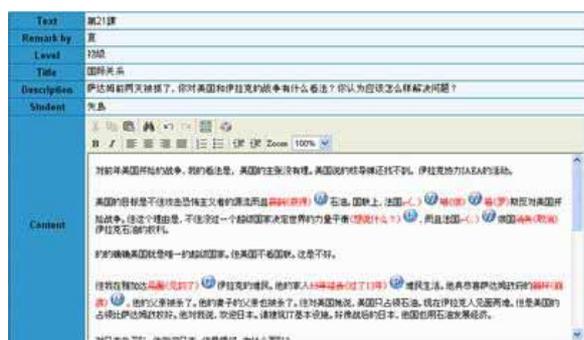


図 8 宿題添削画面

子メールで行っていた。この方法は効率が低く、データ管理が難しいなどの欠点がある。現行システムでは、JAVASCRIPT 言語で DHTML に基づきオンラインのエディターを設計した。Richtext ベースの、Word などテキストエディターに近い編集機能を提供することによって、オンラインの宿題編集および添削機能を実現した。統一のオンラインエディターを採用して、宿題の提出、添削など作業効率を高めるとともに、フォーマットも標準化した。すべての記録を XML/XHTML のフォーマットでデータベースに保存する。各種文法の誤りを特定の XML/XHTML 標示で記録してあるため、大量のコーパスデータに対し、数量化分析を行うことができた。

3. むすび

Tutorial Chinese プラットフォームの第 1 段階は 2003 年 10 月に完成し、半年間の運行結果、すでに 1282 件のレポートと、286 件の宿題、312 件の BBS 投稿、および 6000 数条の活動履歴を蓄積している。これらのコーパス資源は今後の教育研究およびテキスト編纂に貴重な資料を提供することができるであろう。

2004 年 5 月、開発第 2 段階に入ったが、今後は重点的に以下の課題を研究してゆく。

- ウェブレベル分けテストの成績による、自動クラス編成と授業スケジュールの調節。
- 携帯電話、PDA などモバイル端末と連携し、未習者への警告メッセージ自動送信や、モバイル端末上での直前単語練習の実現。
- SCORM 標準を参考にし、Tutorial Chinese プラットフォームのコーパスと連動可能な新型 CALL 教材の開発。

本稿は、科学研究費補助金研究(B)(2)「国語チュートリアルシステムの国際共同開発」(課題番号: 14310136・研究代表者: 砂岡 和子)による成果の一部である。

参考文献

- [1] 砂岡和子, 村上公一, “以提升交际能力为目的之汉语教学以及水平测验刍议-早稻田大学国际汉语远程教学四年实践见证”, 第三届全球华文网络教育研讨会论文集, 2003年10月24日-10月26日, 台北。
- [2] 刘松, 吴剑明, 浦野义赖, “基于 PHP 与 MySQL 的互动型汉语教学支持平台的研究与实践”, 第三届全球华文网络教育研讨会, 2003年10月24日-10月26日, 台北。
- [3] Apache HP: <http://www.apache.org/>
- [4] PHP HP: <http://www.php.net/>
- [5] MySQL HP: <http://www.mysql.com/>

原載「2004 PC Conference 研究大会予稿集」2004年8月

PC、携帯端末に対応するコールドリルシステムの開発

劉 松,砂岡和子,浦野 義頼

Abstract: CALLドリルオーサリングシステム及びPC、携帯端末に対応する中国語テストシステムの開発について述べる。多国間に分散する受講者の至便性と、ライブ授業を支援するため、多種文字コードの採用、コーパス準拠のオンラインドリル問題作成を実現した開発経緯を紹介する。

Keywords: Computer Assisted Language Learning (CALL), Corpus, Mobile telephone, Multi-language

キーワード : CALL、コーパス、携帯電話、多言語対応

1. はじめに

近年、中国の急速な経済発展を背景に、日中間の往来がより頻繁となり、中国語や日本語の学習需要も増大している。インターネットを使った語学 E-Learning が普及する一方、移動通信技術の急速な普及につれ、携帯電話を用いた語学学習プログラムが新たな注目を集めている。特に第3世代移動通信方式である IMT-2000 は、世界統一標準を目指しており、一台の携帯電話で、何時でも何処でも、世界中の学友と交流できるユビキタス・ラーニングが可能となる。

早稲田大学では2000年から、学部の1・2年生を対象とする選択演習科目として、新たに「チュートリアル中国語」コースを開設し、ライブ型レッスンを中核とする、総合的中国語学習プログラムを導入した。学生4人とネイティブのチューター1人からなる、フェイス・トゥ・フェイスの中国語会話レッスンと、台湾師範大学や中国北京大学の教師による、テレビ会議やビデオ・チャットを組み合わせたハイブリッド遠隔講義を行っている。

この新しい教育モデルをより有効的にサポートするために、我々は2003年から Tutorial Chinese 授業の管理プラットフォームを開発し、さらに2004年から予習、復習ためのCALLドリルシステムを開発した。このドリルシステムは、テキスト語彙コーパスに準拠し、言語事実に基づいた語学訓練効果が期待できる。また、PCと携帯電話端末双方に対応するため、学習者が、時と場所を選ばずアクセスしやすいユビキタスラーニング環境を築いた。

本論文はCALLドリルシステム開発経緯、主要モジュールのロジカル構造とコア技術を紹介し、また将来的なシステム拡張についても触れる。

2. システムの設計

2.1 システム開発環境

CALLドリルシステムの開発は「LAMP」モード (Linux + Apache + MySQL + PHP) を採用している (図1)。「LAMP」はオープンソース・ソフトウェアが普及する中で、注目を集めるソフトウェア群である。従来は小規模の Web アプリケーションで利用されていたが、最近は大規模システム? (エンタープライズ用途) まで適用範囲が広がりつつある。

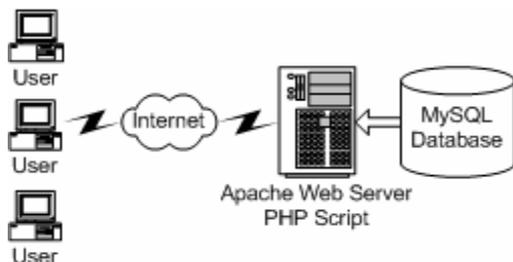


図1 CALLシステム開発環境

2.2 コーパス設計

本システムが利用した語学コーパスの資源は、北京大学 The Grammatical Knowledge-base of Contemporary Chinese corpus 約 10000 語と HSK (中国漢語水平考試) 甲乙丙級約 8000 語、日本の大学教育用基礎語彙コーパス約 3000 語、およびチュートリアル中国語会話レッスン用テキスト中の約 8000 語句からなる計 4 種類のデータを基盤とした。これらの語句にそれぞれ、品詞情報、日本語訳、発音情報

(ローマ字音注)、語句の難度情報を付加してコーパスとし、「チュートリアル中国語」専用のウェブデータベースに搭載した。授業進行に合わせ、上記コーパスに典拠し、CALL 用ドリルを作成する。別途、受講者のレベル判定用の、コンピュータテストの作問と出題のデータベースにもなる。

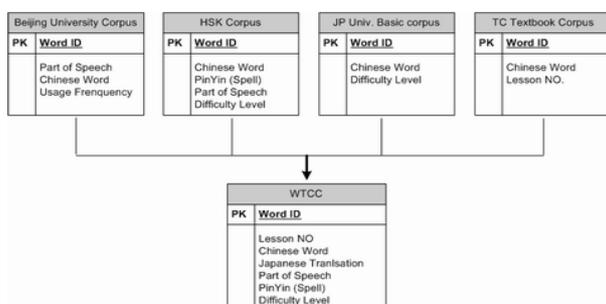


図2 CALL コーパス構成

2.3 システムの構成

CALL ドリルシステムは、「練習問題の作成」、「練習ペーパーの作成」「練習画面」、及び「コーパス管理」の、四つの部分から構成される。それぞれの役割は：

- 「練習問題の作成」：テキスト単語コーパスと連携し、半自動的に作問できる。
- 「練習ペーパーの作成」：過去問を自由に組合せてペーパーを作る。
- 「練習画面」：PCあるいは携帯電話端末で、予習、復習を実施する。
- 「コーパス管理」：既存のコーパスの修正、検索及び新しい単語コーパスの追加などの管理。

現在、PC版の練習問題には、「穴埋め」、「選択」、「並べ替え」の三タイプがあり、携帯電話版練習問題は「選択問題」一種である。教師と学生がPCあるいは携帯電話でシステムにアクセスし、認証を受けてから、それぞれの操作が始まる。

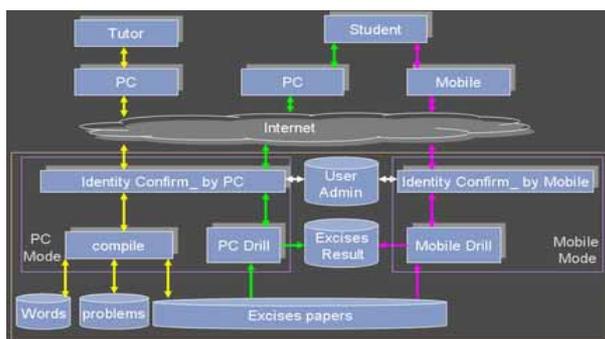


図3 CATIシステム構成

3. システムの実現

本システムは2004年5月に提案され、9月まで第一段階の開発を完成、2004年12月から Tutorial Chinese 授業の管理プラットフォームと連携し、約200名の学習者に対し、テスト運用を実施した。

PC版システムは既にFull・Unicode技術を採用し、全面的に日本語、簡体字中国語、繁体字中国語、英語など、多種類の文字コードをサポートしていた。携帯電話版のシステムも多言語対応できるため、プログラムで自動的Unicodeの文字を画像データに変換する。さらに画像圧縮技術を利用し、16X16画素の文字を2色(白黒)のGIF方式で圧縮すると、通常ビットマップで300~500バイトの文字が、およそ40~60バイトに圧縮し、約1/8~1/10の圧縮率が達成できる。現在の携帯の伝送速度は一秒当たり28.8から2100キロバイトの能力があるため、上記の画像データ閲覧なら問題はない。中国語以外の言語コードも表示できるので、外国語学習に広く応用できる。



図4 携帯電話練習画面

図 5 携帯電話で
多言語表示



4. むすび

チュートリアル式中国語レッスンの教学支援用にCALLドリル及びPC、携帯端末に対応する中国語練習システムの開発を紹介した。今後の課題は、①レッスンで使った新出語彙のコーパス自動追加登録の実現、②確認テストの能力別・個別カスタマイズ化開発、③AUのW21SとA5503SAでの実験に加え、今後各携帯機種に対応可能なシステム環境開発、を目指す。

参考文献

- [1] 砂岡和子等:「基于 Active Database 与師生課本互動的國際漢語口語教学」, 数字化對外漢語教學理論与方法研究(清華大学出版社), p. 328-334(2004)
- [2] 劉松・砂岡和子・浦野義頼:「日中台遠隔中国語口語レッスン管理プラットフォームの開発-早稲田大学 Tutorial Chinese E-Learning Platform-」, PC Conference 論文集, P142-145(2004)
- [3] LanZhang・SongLiu・JianmingWu・Yoshiyori URANO:「Chinese Language Learning Systems with 3G Mobile Phones」, Asian Info-communications Council WG3, Document No. 91(2004)

原載「ケータイ・カーナビの利用性と人間工学」2005年3月

汉语语料库的建设及其在汉语教学中的应用*

俞士汶 段慧明 朱学锋
北京大学信息科学技术学院
北京大学计算语言学研究所

1. 引言

语言学家历来重视第一手语言资料的获取和运用。在个人计算机普及以前，语言学家只能靠手工用卡片收集少量的典型的资料。所以，中国前辈语言学家曾感叹他们是在用小本钱做大买卖。计算语言学把传统的语言学研究同自然语言信息处理技术结合起来，为语言学的发展、为语言教学研究注入了新的活力。

计算语言学 (Computational Linguistics) 这个专业术语自 1960 年代在文献中出现以来[1]，这门学科已经有了 40 多年的发展历史。1990 年代，语料库语言学异军突起，这是计算语言学发展历史上的一个重要的里程碑。由于计算机以及应用技术（电子排版、因特网等等）的发展，大规模的电子文本唾手可得，自然语言处理技术专家有了从大规模真实文本中获取语言知识（词频、词义、词语搭配、句子结构规则等等）的物质基础，基于大规模语料库的统计语言模型(如：n 元语法、隐马尔可夫模型、向量空间模型等)在诸如词性标注、文本分类等语言工程中的成功应用[2]更加有力地推动了语料库语言学的发展，至今仍是方兴未艾。

推动语料库语言学发展的基础是由大规模真实文本资料构成的语料库。可以引用以下事实作为语料库价值的佐证。2002 年 3 月，中国教育部发布《第一批异形词整理表》时，李行健先生解释了整理异形词所遵循的三个原则：通用性、理据性、系统性。并进一步说明：“通用性怎么确定呢？就是用计算机做词频统计。”这项词频统计工作是应用计算机技术、选用 1995 年至 2000 年 6 年共计 1.5 亿字的《人民日报》语料作为对象进行的。

如果没有理解错，笔者以为李行健先生所做的“词频统计”是仅仅针对预先收集到的几百组（最多不过 1500 组）异形词进行的。即使针对性如此明确，在中文文本中统计这些异形词时，仍会遇到困难。原因同日语一样，书面汉语的词与词之间不留间隔。“过分”和“过份”是一组异形词。如果在原始文本中简单地统计汉字串“过分”的出现次数，就会出现误差。例如，“有人做过分析”、“他们有过分歧”、“已经检查过分会场了”、“北大在汉中办过分校”等等汉字串中的“过分”都不是一个词。汉字串“过分”的频度并不等于“过分”这个词的频度。笔者以为，进行异形词统计的研究者对机器检索结果是进行了人工校对的。但这样的工作并不能得到关于数以万计的汉语通用词语使用频度的整体认识。对于中文来说，实现大规模文本的正确的词语切分是词频统计的基础。相对于原始的未经过任何加工的（生）语料，将汉语句子正确地切分为词语序列（即在词语同词语之间加上区分标记，如：空格）的语料就是经过加工的（熟）语料。面向不同的应用，语料库需要有不同深度的加工。只作切分，只能得到不区分词性的词频统计。以“把”为例，至少有 3 种词性：介词，量词，动词，意

义和用法截然不同。只有在大规模文本中，在切分的基础上，进一步标注“把”的词性，才能统计出带词性的词频。显然，带词性的词频信息无论对信息处理、词典编纂还是汉语教学都有重要的意义。通常，将词语切分和词性标注这两件密切相关的加工任务叫做汉语语料库的基本加工。

语料库语言学为汉语语料库的基本加工提供了理论模型和实现技术。面向语言信息处理的理论、技术和应用，北京大学计算语言学研究所（以下简称“北大计算语言所”）研制了系列化的经过加工的语料库和其他语言知识库。

*本文相关研究得到中国国家自然科学基金项目 69483003、973 项目 G1998030507-4、863 项目 2001AA114040 的支持。

2. 北京大学计算语言学研究所语料库建设概况

北大计算语言所从 1992 年起开始研究现代汉语语料库的多级加工技术，历时已有 10 载，对从词语切分和词性标注的基本加工到树库的构建都进行过探索[3,4,5]，并取得了一些有价值的成果。

《人民日报》标注语料库

目前，北大计算语言所拥有的加工语料库首推自 1999 年 4 月至 2002 年 4 月历时 3 年完成的《人民日报》标注语料库。该语料库包含 1998 年全年 2600 多万汉字文本，对全部文本已完成词语切分和词性标注等基本加工。该项成果通过了合作单位 Fujitsu 的验收。其中 1 月份的 200 多万字的标注语料已在计算语言所的主页（www.icl.pku.edu.cn 或 icl.pku.edu.cn）上公布，可以免费下载。对于上半年的 1300 多万字的标注语料，业界可以通过人民日报社新闻信息中心获得许可使用权。

语料库的加工离不开详细的、严谨的规范的指导。《人民日报》语料库的加工规范是 1999 年 3 月制订、2001 年 7 月修订的《现代汉语语料库加工规范—词语切分与词性标注》。这个规范以《北京大学现代汉语语料库基本加工规范》[6]为题，已在《中文信息学报》2002 年第 5 期和第 6 期连载。

从遵照这个加工规范完成的《人民日报》标注语料库中，摘取一段样例如下：

原文是《人民日报》1998 年 12 月 1 日第一版第二篇文章的第 1,2,3,4 段。经加工后，词语之间有了空格，斜杠之后的字母是该词语的标记。《规范》的标记集包括约 40 个标记。其中包括 26 个词性标记（依据北大计算语言所开发的《现代汉语语法信息词典》的词类体系[7]，基本词类有 18 类：名词 n、时间词 t、处所词 s、方位词 f、数词 m、量词 q、区别词 b、代词 r、动词 v、形容词 a、状态词 z、副词 d、介词 p、连词 c、助词 u、语气词 y、叹词 e、拟声词 o。另外增加了成语 i、习用语 l、简称 j、前接成分 h、后接成分 k、语素 g、非语素字 x、标点符号 w 等 8 个附加类别）、专有名词标记（人名 nr，地名 ns，团体机构名称 nt，其他专名 nz 等）、语素子类标记（如名语素 Ng，动语素 Vg，形容语素 Ag，副语素 Dg 等）以及动词和形容词的特殊用法标记（如名动词 vn，副动词 vd，名形词 an，副形词 ad）。不妨将

这些标记笼统地称为“词性标记”。短语型专名“”、“”先切分标注，再用方括号标记出来，结果为。

读者也可以在本所的主页上测试中文文本词语切分和词性标注软件。

注音语料库

中文汉字的一字多音现象虽然没有日文那么严重，对于初学者和外族人，仍然是令人头痛的。如果能实现中文文本的自动注音，对汉语教学是很有帮助的。这也是实现语音识别、语音合成等汉语信息处理系统的基础技术之一。为了这些目的，北大计算语言所也开发了一定规模的注音语料库。摘录一段样例如下：

咱们/r 中国/ns 这么/r 大{da4}/a 的{de5}/u 一个/mq 多/a 民族/n 的{de5}/u
国家/n 如果/c 不/d 团结/a , /w 就/d 不/d 可能/v 发展/v 经济/n , /w 人民/n
生活/n 水平/n 也/d 就/d 不/d 可能/v 得到/v 改善/vn 和{he2}/c 提高/vn 。 /w

置于花括号内的字母数字串是多音词的汉语拼音（分别用 1,2,3,4,5 代表第一声、第二声、第三声、第四声和轻声）。它是在切分标注的基础上加上去的。对于单音词，直接查词典就可以知道他们的读音了。读者也可以在本所的主页上测试中文文本自动注音软件。附带说明一下，mq 表示前面的词是数量词（数词的一个子类）。

英汉对照双语语料库

北大计算语言所于 1980 年代和 1990 年代曾进行了机器翻译译文质量自动评价研究。为此，建立了“英汉机器翻译译文质量测试大纲”（从本所主页可以下载）和根据这个大纲精选的近万句英语句子，并给出了对应的译文。这是北大计算语言所建立英汉对照双语语料库的基础。为了实现基于实例的机器翻译（Examples Based Machine Translation），更需要一个规模更大的不同级别对齐的双语语料库（aligned bilingual corpus）。现在，双语语料库的规模如下：

篇章级对齐的，汉语 2000 万字，英文 1000 多万词。

句子级对齐的，65000 句对

短语级对齐的，50000 多条（一部分短语还附带属性信息）

通常能够收集到的对齐双语语料只是篇章级的。对于应用来说，单位过大，句子级对齐的更适用。为了实现较小的单位的对齐，需要开发辅助工具，也需要制订有利于统一管理和数据交换的标记。现在，不仅有了一套便于双语语料库开发和应用的工具软件（包括实现两种语言间不同单位的对齐、索引和检索、翻译等价单位提取等等），而且也有了覆盖从篇章到句子各级语言单位的 XML 置标规范。

这样的双语语料库对于汉英对比研究和教学也是十分有价值的。北大计算语言所曾参与

开发日汉机器翻译系统，也积累了一些日语资料。在条件成熟时，拟开发包含汉、英、日、韩、蒙、藏等多语对照的语料库。

专业领域语料库

如同一个人做工作，不仅要有语言能力和常识，还要具备专业知识一样，要最终实现自然语言理解，也要给计算机配备专业领域的知识。术语库 (Term Bank) 是最基本的专业知识。由于很多学科都在迅速发展，静态的术语库是跟不上应用需求的。需要建立从语料库中及时提取新出现的术语的机制。北大计算语言所同中国标准研究中心合作于 2002 年启动了一个新的研究课题，即“科技术语辅助提取与术语库建设”。并先选择信息科学与技术领域作为实验对象。

到 2002 年底，已收集信息科学与技术领域的语料约 1 亿字，并完成了篇章级标注（包括分支学科类别、标题、作者、摘要、关键词、来源、作品类型、地域分布等等），还要在 2000 万字精心选择的核心语料库中完成本领域术语的标注。这样，计算机辅助提取术语的软件不仅可以利用术语本身的结构知识，而且可以利用术语使用的语境知识。

3. 北京大学计算语言学研究所的其他语言资源

除上述语料库外，北京大学计算语言学研究所十多年的基础研究和应用研究中，还积累了其他类型的语言资源。

《现代汉语语法信息词典》

最重要的、而且最有影响的一项资源是自 1986 年便开始研制的《现代汉语语法信息词典》。这是一部面向语言信息处理的电子词典。当前的收词规模已超过 7.3 万[8]。构建这部大型语法信息词典的最具创新性的思路是在依据语法功能分布原则建立一个现代汉语词语分类体系并实际完成 7.3 万多词语的归类的基础上，进一步采用关系数据库技术对属于同一类的词语详细描述它们的语法属性。词典中有一个总库，吸纳各类词语共同的十几项属性信息。对每一类词语再各自建立一个数据库，描述其专有属性。以动词库为例，设立的属性信息字段包括：是不是联系动词、助动词、趋向动词，能不能受“不”、“没”、“很”、“在/正在”等副词修饰，能不能带“了”、“着”、“过”等助词，能不能带宾语，如果能带宾语，还要描述能带什么样的宾语，体词性的、谓词性的还是准谓词性的，能不能带双宾语，等等，共计 40 多项。为了既能更细致地描述体宾动词、谓宾动词、双宾语动词等动词子类的属性，又尽可能减少词典的冗余信息，对动词又设立了 6 个分库。总括起来，《现代汉语语法信息词典》所描述的动词属性达 120 多项。当然，在各类词中，动词的属性字段是最多的。各类词语的语法属性字段平均约有 50 项。规模是足够大的。而 7.3 万词语的归类以及这些属性的值的确定更是不容易的。以“分配”和“接受”这两个常用动词为例，也许人们平常说“分配部下一个任

务”和“接受上级一个任务”这样的话。从这个层次看，“分配”和“接受”都可以带双宾语。但《现代汉语语法信息词典》认为“分配”是双宾语动词，而“接受”则不是。为什么？“分配部下一个任务”可以变换为“给部下分配一个任务”，而“接受上级一个任务”不能这么变换。反之，“接受上级一个任务”可以变换为“从上级接受一个任务”，而“分配部下一个任务”又不能这么变换。除了上面的解释外，还可以从另外一个角度考察“分配部下一个任务”和“接受上级一个任务”。实际上，“接受上级一个任务”可以在“上级”后加上一个“的”字，得到“接受上级的一个任务”，仍符合原义。但在“分配部下一个任务”的“部下”之后就加不上“的”。原来，“接受上级一个任务”是省略了“接受上级的一个任务”中的“的”。在更深刻的认识层次上，这两个动词带宾语的属性是不同的。《现代汉语语法信息词典》的规格说明书限定只对“给予”义的动词描述它能否带双宾语。

这部词典是制订语料库加工规范和开发自动切分标注软件的基础。反过来，大规模加工语料库的存在又推动了《现代汉语语法信息词典》的发展。现在的《现代汉语语法信息词典》关于词的语法属性的描述都是定性的，属性值的类型是可否型（或逻辑型）的。有了《人民日报》标注语料库之后，可以将其改造为概率型的。仍以动词的若干属性为例（下表中的数据是从2600万字标注语料库中统计出来的）：

动词属性频次表

动词	总频次	不	没	很	在	正在	了	着	过
吃	3125	103	24	1	7	2	177	25	85
发	2261	9	8	0	0	0	194	5	12
发展	21021	27	0	0	72	16	161	20	1
进	4435	38	7	0	1	2	584	15	0
进行	19813	32	1	0	70	193	3572	79	77
睡	289	7	5	0	0	0	5	9	12
说	34356	96	29	0	12	0	141	96	262
听	1669	44	4	0	2	0	210	74	21
想	4339	262	19	38	31	1	38	89	34

上表列出了9个常用动词的若干属性的频次。总频次表示这些动词在2600多万字的一年语料中出现的总次数。“很”所在列的值表示它们左接副词“很”的次数。从“发”到“听”，中间7个的值都是“0”，验证了《现代汉语语法信息词典》关于这些动词不能受“很”修饰的描述。“想”在语料中共出现4339次，受“很”修饰38次，通常认为“想”可以受“很”修饰，也是得到了证实的。问题出在“吃”上。通常认为“吃”不能受“很”修饰。但这里出现了1次（“拷寇”是一出唱、念、做、舞都很吃功的戏）。以往常常用一个反例否定一个命题（二值逻辑理该如此），但语言现象又不宜如此简单地处理。 $1/3125=0.00032$ ，如果用概

率值描述, 可以近似地认为“吃”受“很”修饰的概率为 0.032%。而 $38/4339=0.0088$, 即“想”受“很”修饰的概率为 0.88%。如果仍希望使用“可否型”的值, 不妨确定一个阈值 $\delta=0.5\%$, 因 $0.88\%>\delta$, 重新确认“想”可以受“很”修饰, 而 $0.032\%<\delta$, 也重新确认“吃”不能受“很”修饰。尽管同原有的语法词典的属性值相同, 但所依赖的知识背景却大不一样。顺便还可以讨论, “吃功”这种说法是在很专门的领域(戏曲艺术)中使用的, 不妨认为“吃功”就是一个词, 即形容词, 它能受“很”修饰又是完全符合规律的。

现代汉语语义词典

适应机器翻译系统消解句法结构歧义和词语翻译歧义的需要, 北大计算语言所与中科院计算所自 1994 年至 1998 年联合开发了一部基于广义配价理论的《现代汉语语义词典》[9], 完成了 4.9 万汉语常用实词(名词、动词、形容词)及部分成语、习用语的语义分类和配价信息描述。自 2001 年起, 计算语言所又与北大中文系合作, 在中国国家 973 项目“面向新闻领域的汉英机器翻译系统”和“现代汉语动词论旨结构系统和汉语词语语义分类层级系统研究”的支持下, 再次发展《现代汉语语义词典》。预计 2003 年初, 《现代汉语语义词典》的规模将达到 7 万个记录, 词典的质量也有较大的提高。它既可以在中文信息处理应用系统中独立使用, 也可以通过“词语、词类、同形”这 3 个主关键字(primary key)与《现代汉语语法信息词典》进行链接, 相互配合, 从而使系统获得更加完备的语法、语义信息。

中文概念词典

北大计算语言所自 2000 年 9 月开始构建中文概念词典(Chinese Concept Dictionary, CCD)。CCD 是与 WordNet 兼容的现代汉英双语概念词典, 同时提供汉英双语概念的知识表达[10]。在词典设计上, 它参照 WordNet 的框架, 从关系语义学的观点出发, 用同义词集合(synset)来描述概念, 用同义关系、反义关系、上下位关系、整体部分关系来描述词汇的语义; 同时, 针对中文的语言文化特点, CCD 也在概念的内容和概念间的关系上对 WordNet 进行了一定的调整和发展。CCD 可以直接应用于概念检索、信息提取、自动文摘、文本分类和机器翻译等领域, 为其中的语义理解任务提供丰富的语义资源。中文概念词典和现代汉语语义词典各有侧重, 相辅相成。

CCD 课题组将计算词典学的诸多新思想引入研制全过程。在各个开发阶段, 都设计并实现了合适的词典辅助开发工具。特别是基于 CCD 构造模型的可视化辅助词典构造软件 VACOL (Visualized Auxiliary Construction of Lexicon), 极大地提高了开发效率, 并保证了词典质量。CCD 构造模型强调双语词典(Bilingual Lexicon)构造中的继承(Inheritance)和转换(Transformation)的思想, 逐步实现由 WordNet 到 CCD 的计算性转换和词典结构的自动维护, 水到渠成地得到一个汉英对应的双语概念词典, 从而大幅度提高词典编纂的质量和效率。现在已完成了约 60,000 个双语概念的描述。2003 年将完成 7 万个中文常用概念的描述。

信息科学技术领域术语库

根据北大计算语言所同中国标准研究中心于 2002 年签订的“信息科学与技术领域术语辅助提取及术语库的建设”项目的任务书，现在已初步建成了一个“信息科学技术术语库”。其中术语条目已超过 15 万，对其中的 6 万条术语已收集了相关定义，正在校对中。更完整的信息科学技术术语库在建成后，将提供社会服务。

中国古代诗词资料库

北大计算语言研究所开展古诗词计算机辅助研究，已有近十年的历史[11]。已收集整理古代诗词（全唐诗、宋代部分名家诗、全宋词）语料 1100 万字，其中 900 万字完成了题名、作者、体裁等篇章信息的标注，640 万字完成了注音、切分和词性标注。还构造了一部含 4 万多条目的多字词词典，并建立了 1400 名诗人的属性库（包括生卒年、性别、籍贯以及诸如皇帝、官僚、平民、僧人、道士等身份）。

短语结构规则库

现代汉语的短语(phrase, 又叫词组)是由两个以上的词或短语按照一定的规则构成的、能在更大的句法结构中承担主语、谓语、述语、宾语、定语、状语、补语、中心语等功能的语法单位。针对《现代汉语语法信息词典》中的 18 个基本词类，全面地考察了 $18 \times 18 = 324$ 种两两组合（例如名词和名词、名词和动词、动词和名词、动词和动词等）构成合法短语的情况，然后再用短语（名词性的，动词性的，等等）替代合法短语中的词类，递归地构造出更复杂的短语，如此得到 600 多条短语结构规则[12]。

适应计算机自动处理的需要，这些规则是形式化的，严格地给出规则成立的条件（包括《现代汉语语法信息词典》中词语的语法属性）及执行规则所产生的结果。例如，对于动-名组合具体地指明什么样的动词和什么样的名词可以构成什么样的短语，该短语是体词性的还是谓词性的，它继承了构成成分（特别是中心成分）的哪些属性，失去了哪些属性，又派生了哪些新的属性等等。

4. 语言知识库在汉语教学中的应用

Internet 的迅速扩展引起了教育模式的变革，“网上教学”成了时尚。同传统的教育模式相比较，“网上教学”的确有其优势。学校代替私塾曾经是中国教育模式的一个重大变革。在私塾里，学生只跟一个老师学，因而有“良师出高徒”之说。到了学校，学生可以跟许多老师学，好比蜜蜂采百花蜜，这是一个重要进步。不过，不论私塾还是学校，学生基本上是在被动地接受知识，主动学习的空间相当狭小。“网上教学”就不同了，学生可以根据自己的需要和时间安排主动地学习，也可以超越地理界限，跟远在外地、外国的老师学。这里所说的“老师”是广义的，包括各个教育网站以及它们所提供的各种教育资源。在以“教”为主导的模式中，好的老师已经主张“教给学生一杯水，自己要准备一桶水”。当教育模式转变为以

“学”为主导时，作为教育资源的核心和各种知识库的建设和质量成为网上教育成败的关键。将传统的教科书、百科全书等所包含的知识重新组织和包装，成为电子书籍（E-book），自然是重要的知识库。网上语言教学还有另外一个知识库来源。那就是为计算机处理乃至“理解”自然语言而开发的各种知识库，这些知识库具有基础性和完备性等特征，作为网上语言教学的资源有其独特的优势。汉语教学的对象可划分为两类：第一类学生的母语是汉语；第二类学生的母语是其他语言。针对不同的对象，利用这些知识库的方式应该是不一样的。本文主要以第二类学生为对象，讨论如何利用这些知识库。

词汇教学是语言教学的基础。相当多的外国朋友学习汉语是为了能够阅读中文资料以便直接了解当代中国的社会、政治、经济、科技、文化、教育、体育、旅游等各方面的实际情况。《人民日报》是反映当代中国实际的知识宝库，也是传播中国社会各个领域的信息的主渠道。对于外国人来说，阅读《人民日报》相当于打开了认识中国的一扇窗户。甚至一张简单的《人民日报》词频表也能折射出中国社会生活的某些侧面。而且，李行健先生在解释为什么选《人民日报》作为词频统计的语料时，明确地指出：“专家和有关方面都认为，《人民日报》最能代表当代汉语的书面语言风格”。词典是词汇教学的主要工具书。编词典碰到的第一个难题就是收词问题。根据使用频度选词最合理，可是过去却不便操作。这是因为中文文本只适合做字频统计，而词频统计就绕不开词语切分的难题。现在有了2600多万字的《人民日报》标注语料库，得到“切分单位”的出现频度（这是“词频”的一个最重要的参照数据）是轻而易举的事。新词语是词汇教学和词典编纂必须面对的另一问题。虽然使用频度仍是决定一个新词语该不该收入词典的重要依据，但是在文本中确定新词是中文信息自动处理的一个更困难的课题。当注意到2600多万字的《人民日报》标注语料库是由1998年全年的语料加工得到的，这个问题的解决有了希望。从这个标注语料库得到了1998年的词表，以此为基础，处理1999年的《人民日报》，准确度会很高。新词基本上只会在那些“单字碎片”中出现。需要特别关注的范围大大缩小，人与机器配合，可望解决新词发现问题。这样便又得到了1999年的词表，如此递进，不断发现和确认新词，词典编纂就可以跟上时代的步伐。当然，也可以先从2002年做起，两头夹击，也许可以更快地解决这个问题。

在1980年代，中国也曾在2000万字语料的规模上进行过词频统计[13]。但同一个“词形”却可能是不同的“词”。例如，“在”既是介词，又是副词和动词，意义和用法彼此不同。因此，为了更准确地、更细致地把握汉语词汇的全局，需要做带词性的词频统计。在汉语语言学研究史上这是从来没有做过的事。由于2600万字的语料库不仅完成了切分，而且进行了词性标注，这样就可以统计带词性的词频或概率。现在，已经有了这样的数据：“在”作为介词、副词和动词的概率分别是95.4%，2.4%，2.2%；“把”（读ba3）作为介词、量词和动词的概率分别是95.9%，2.8%，1.1%，而“把”作为数词（助数词）和名语素的概率接近于零。带注音的语料库可以帮助学生掌握词语的正确读音（如连词“和”、介词“和”都读he2，动词“和”却有4种不同读音：he2, hu2, huo2, huo4）。显然，参照信息处理基础研究得到的这些数据实施汉语词汇教学，会收到事半功倍的效果。

名词和特定的个体量词搭配是汉语的特点之一，也是外国人学汉语感觉麻烦的地方。中

国人说“三头牛、五匹马”，为什么不问“三牛、五马”、“三匹马、五头牛”或“三个牛、五个马”，很难说出什么道理。因此，已故语言学家朱德熙先生在1987年出版的《现代汉语名词量词搭配词典》（作者：刘学敏、邓崇谟）序言中说：“现代汉语的名词大都有固定的量词跟它搭配，对于外国人、外族人或中小學生来说，学习一个新的名词的时候，必须同时记住跟它搭配的量词是什么。因此理想的词典应该在名词条目下注明适用的量词”。不过，《现代汉语名词量词搭配词典》只收了8000词条。《现代汉语语法信息词典》所收的名词有3.5万之多，量词也在400以上，并对量词划分了子类。对每个名词都按量词的子类具体给出可以搭配的量词，如“笔”可以搭配的个体量词有“支、管”，容器量词有“盒”，种类量词有“种、类”，成形量词有“把”，不定量词有“些”。可以说，《现代汉语语法信息词典》实现了朱德熙先生的遗愿。

在汉语教学过程中，相互参照《现代汉语语法信息词典》和标注语料库是十分有意义的。《现代汉语语法信息词典》中的名词-量词搭配知识是根据研制者的认知并参照一些词典和语法著作给出的，难免有失偏颇。现在有了大规模标注语料库，可以在其中检索名词“笔/n”前面所搭配的量词，从而发现在真实语料中“笔”到底与哪些量词搭配，也许“管”这个量词现代已经不使用了，这样可以更有根据地改进词典的质量。反过来，如果外国学生在标注语料库中检索“尼龙/n”所搭配的量词或“双/q”所搭配的名词，检索出“五/m 双/q 尼龙/n”，感到困惑，可以去查《语法信息词典》，就会发现“尼龙”是不能与“双”搭配的，而与“双”搭配的典型名词也没有“尼龙”。这会促使学生扩大观察窗口，得到“五/m 双/q 尼龙/n 袜子/n”。就是说，这里的“双”实际上是与“尼龙袜子”的中心词“袜子”搭配的，这个认识在《语法信息词典》中又可以得到验证。人的认知不断加深的过程也为智能化的语言信息处理技术提供了启示。

当然，在标注语料库、《语法信息词典》这样的语言知识库上进一步开展数据挖掘和知识发现研究，还可以得到更深层的知识，如词语搭配知识、短语结构规则、句型及其分布等等。例如，吴云芳研究“‘是’字句主语和宾语的自动界定”[15]和“现代汉语的联合结构”就是有价值的尝试。

在中国人的语文教学中，语法教学是被“淡化”的环节。究其原因，固然与汉语的某些特点（组词造句无形态变化，“意合”起的作用或许比其他语言更多一些）有关，但是，汉语语法研究历史不长，理论不够成熟，体系不够完备，甚至系统地掌握第一手资料的工作也做得不够，语法研究的成果及其在汉语教学中的价值还没有能够充分显示出来，可能也是一个原因。通常的语法教材只论述属于同一类的词有哪些共同的语法属性，很难详细列举属于同一类的词有哪些不同的语法属性，《现代汉语语法信息词典》正好做到了这一点。而且可以进一步了解具有某种语法属性的词到底是哪些，同语料库相结合，还可以找到具体运用的实例。显然，这些资料对在教学中权衡语法知识的取舍，以提高教学效果是很有用的。在3.1中曾讨论了“分配”和“接受”是否能带双宾语的问题。外国学生理解“汉语中存在能带双宾语的动词”不困难，但要判断具体的一个动词（“分配”或“接受”）是不是双宾语动词就难了。特别是遇到“接受上级一个任务”这样的句子，就更难理解它为什么不是双宾语的句子了。《现

代汉语语法信息词典》以及标注语料库等知识库提供了准确、丰富的语法知识，如果能加以运用，对汉语教学会有很大的帮助。

5. 结语

本文介绍了北大计算语言所开发的面向信息处理的汉语知识库，重点介绍了《人民日报》标注语料库以及《现代汉语语法信息词典》。本文也探讨了这些知识库在语言教学中的应用。但是，由于笔者缺乏汉语教学的实际经验，这些探讨是十分肤浅的。汉语信息处理研究实际上是在教最笨的“学生”——计算机“学”汉语。之所以说计算机笨，是因为它缺乏人的认知能力，好在它存储量大，“记性”好，计算快，让它多记住一些基础的、系统的、全面的知识，也许可以变得“聪明”些，可以为人类多做一些事。这样的研究工作显然也可以从汉语教学的理论与实践中得到启发。实际上，第二语言教学模型对机器翻译的理论模型就有很大的影响。因此，期望同从事汉语教学的老师以及学汉语的学生进行交流，学到如何教汉语、如何学汉语的学问。

笔者衷心感谢北京大学计算语言学研究所全体同仁、学生以及研究所以外的合作伙伴，感谢他们的努力和奉献。衷心感谢中国的各类科研项目以及国内外合作单位对这些研究给予的支持。

衷心感谢日本早稻田大学砂冈和子教授。由于她的推荐和邀请，俞士汶才得以参加此次“汉语国际远程教育的实践与前景”盛会，并且获益匪浅。

参考文献

- [1]冯志伟，计算语言学基础，北京：商务印书馆，2002年8月
- [2]吴立德，大规模中文文本处理，上海：复旦大学出版社，1997年7月
- [3]周强、俞士汶，一个人机互助的汉语语料库多级加工处理系统 CCMP，见陈力为、袁琦主编《计算语言学进展与应用》，清华大学出版社，1995年，P50-55
- [4]Qiang Zhou and Shiwen Yu, Annotating the Contemporary Chinese Corpus, International Journal of Corpus Linguistics, Volume 2, Number 2, P239-258, 1997
- [5]周强、张伟、俞士汶，汉语树库的构建，《中文信息学报》，1997年第4期，42-51
- [6]俞士汶、段慧明、朱学锋、孙斌，北京大学现代汉语语料库基本加工规范，《中文信息学报》，2002年第5期和第6期连载
- [7]俞士汶、朱学锋、王惠、张芸芸 著，《现代汉语语法信息词典详解》，北京：清华大学出

版社，1998年4月

[8]俞士汶、朱学锋、王惠，《现代汉语语法信息词典》的新进展，《中文信息学报》，2001年第1期58-65

[9]王惠、詹卫东、刘群，《现代汉语语义词典》的概要及设计，《1998 中文信息处理国际会议论文集》，清华大学出版社，1998年

[10]于江生、俞士汶，CCD的结构与设计思想，《中文信息学报》，2002年，第16卷第4期，P12-20

[11]胡俊峰、俞士汶，唐宋诗之计算机辅助深层研究，《北京大学学报（自然科学版）》，2001年，第37卷第5期，P725-733

[12]俞士汶，关于计算语言学的若干研究，《语言文字应用》，1993年，第3期，P55-64

[13]刘源等著，《信息处理用现代汉语分词规范及自动分词方法》，北京：清华大学出版社，1994年第1版

[14]朱德熙，《现代汉语名词量词搭配词典》序，杭州：浙江教育出版社，1989

[15]吴云芳、段慧明、俞士汶，“是”字句主语和宾语的自动界定，《中文信息学报》，2002年，第16卷第2期，P40-46

原載「中国語国際遠隔教育の实践と展望」論文集（2002年2月）

基於社會互動語言觀點的遠距教學原則

信世昌

國立臺灣師範大學 華語文教學研究所

透過視訊會議方式的遠距教學是目前最熱門的教育發展方向，將之應用於語言教學也是必然的趨勢。遠距教學不僅在教學地點與傳遞方式異於傳統課堂教學，在師生互動、同儕互動、內容安排及教學方法等方面亦產生不同的狀態，尤其又牽涉到跨國異文化的交流，使得遠距教學更需要慎密的規劃設計。

壹、語言教學的三種理念

一般在設計教學之前，應先行訂定教學的目標，再決定教學方法，但是在擬定教學目標及教學方法之前，更應先行思考一些關於語言的基本理念，例如語言是什麼？語言的作用為何？什麼是語言教學？理想的語言教學為何？這些最基本的理念將影響教學目標及教學方法的制定。

Richards & Rodgers (1986) 將語言的觀念概分為「結構觀」(Structural view)、「功能觀」(Functional view)與「社會互動觀」(Social interactional view)。「結構觀」認為語言是一種有系統的知識，語言由語素(語音、單字、詞彙)所構成，語言教學就是教授這些語素及其構成的法則(拼音、構詞、語法)，因此特別注重語言結構的分析及形式的操練，並且遵照結構的難易而循序漸進；「功能觀」則強調語言是一種具有功能性質的工具，是發揮功用及表達含意的載具，因此語言教學應著重語言的實用性，以語意理解與溝通能力為其根本，偏向於聽說讀寫的實用技能訓練；而「社會互動觀」視語言為人際關係及社交表現的工具，也是開創與維持人際關係的工具，語言的習得是在真實而有意義情境中發生，語言教學著重在維持人際關係與社會互動的實際應用，並著重文化理解的層面，並且強調語言使用的場合、分際、情境等語用的層面。

這三種語言觀各有其切入點，也各有其論據，所引申出的教學法及教學重點也就各自不同。從早期的文法翻譯法(Grammar-Translation Method)、口語教學法(Oral Approach)、聽說教學法(Audiolingual Method)、情境學習法(Situational Learning)，到暗示教學法(Suggestopedia)、身體反應法(Total Physical Response)，溝通教學法(Communicative Language Teaching)、以及自然學習法(The Natural Approach)與全語言學習法(Whole Language Approach)等等，都是依照者不同的語言學習理念而產生的。一般的教材也都依據著某種的觀念而設計成不同的面貌，例如採用結構觀的教材，就特別強調語法點及構詞的拆解分析；而功能觀的教材就很強調實用的會

話場景；若是社會互動觀的教材，往往跟本沒有語法解釋或詳細的生詞表，反而附有文化註解及語用情境解釋。但也經常有理念交雜而產生自相矛盾的教材，例如一方面強調功能語言的觀點，卻又注重語法點的分析教學。

貳、遠距教學最適宜的語言觀

上述的三類語言教學觀對於遠距教學有何啟示？由於遠距教學的成本高，並且必須經過特別的安排，因此在設計教程及教材時，必須優先考慮遠距教學的目的及合宜性，更必須釐清背後的語言教學之理念。

在前述的三種語言觀之間，筆者認為與遠距教學最搭配的理念應是基於「社會互動語言觀」的教學理念，功能語言觀次之，最不適合的是傳統的結構語言觀。

社會互動的語言觀認定語言的使用往往是因交際的話題而產生的，不同的交際情境產生不同的詞彙及用語，因此並不預先設定出單字、詞彙或句型，也不精確計算字彙量和句型範圍，即使在評量方面，也是依據語用的原則，測試在某些情境下是否能妥善表達己意及瞭解雙方的觀點和立場，並不單獨測驗拼音、字彙及語法。這些都迥異於一般以「結構觀」為主的教學方法及評量標準，甚至與功能觀的教學也略有差異。

透過網路或視訊會議的遠距教學，尤其是跨文化的語言交際，不宜以結構語言觀為主，基於結構語言觀的語素分析教學，如單字拼音、字詞解釋、語法說明等，最適宜的場合應是課堂上的面授教學。儘管語言形式的操練或許是語言學習過程必經的步驟，但應在課餘做反覆練習即可，而不必通過電視會議來練習，透過遠距教學來進行語言分析或語言形式的操練是一種誤用和浪費。

進行跨國與跨文化的遠距教學，應以「社會互動觀」為主體，並可與「功能觀」的教學相互搭配，例如採用電視會議(Video Conferencing)方式的教學，因為不是直接的教學，而是雙方共同討論一些主題，宜以「社會互動觀」為主導，並可兼及「功能觀」；而採用遠距家教式(Tutorial Method Training)的教學，因為具有教學指導的關係以及固定的學習內容，則宜以「功能觀」為主導，兼及「社會互動觀」。而透過現場網路的教學(例：CUSeeMe)，或是非同步的文字對談(例：Web, BBS)，因為可進行文字對談，則可兩種理念兼顧。

參、引起發言的方式

電視會議的主題及內容是整個教學的關鍵，討論的題目必須具有充分的文化價值及引起發言的興趣。通常電視會議往往從比較通俗或流行的話題入手，以淺顯的內容交談，但是淺顯通俗的話題卻不足以使雙方瞭解彼此的思想與文化。此外，發言的頻率與學習的效果成正比，電視會議最大的困難在於如何激起雙方有意義的互動，促使學生踴躍發言，又能討論具有深度的內容，是遠距語言教學的一大挑戰。

此外，跨國的遠距語言教學必定牽涉到異文化的交流，Byram (1989)提出文化教學的目的

是為了促進對於不同民族及文化的瞭解，並且以減少對於異文化的偏見。基於在有限的電視會議時間內能夠觸及異文化的風格，在會談的主題選擇與內容擬定應持有幾個原則：

一、不妨討論具有爭議的主題：

討論具有爭議的事務最容易激發學生發言的慾望，讓學生覺得有意見想表白，有想說話的衝動，不吐不快。這些有爭議的事務往往與社會制度、政治及文化有關，也最能顯示雙方在文化及思想的差異。某些漢語教材即以此為設計重點（例如美國普林斯頓大學出版的「華夏行」），台灣師大與早大的電視會議曾討論一些有爭議的主題，例如：兩方持不同立場的問題（日本教科書對於近代史的寫法）、兩國都同樣面臨的社會問題（智慧財產權與仿冒的問題）、熱門的政治話題（台灣的選舉與黨派）。這些議題都會引起雙方熱烈的發言，甚至相互辯論，充分達到引起發言的效果。

二、不避諱較敏感的議題

異文化之間必然有許多敏感的事務，大都與雙方的歷史、政治、信仰、民族等有關，因雙方文化與環境的差異，對於事務的觀點及看法極為不同，甚至是雙方聞所未聞而感到震驚。大多數的語言教學場合都避而不談這些敏感話題，若從結構觀及功能觀的角度而言，避開這些議題能保持語言教學的「單純」，也不至於影響語言教學的效果，並且若從社會互動觀的角度來看，這些議題卻又是最吸引人發言之處，也是最值得雙方相互學習之處，因此為了增進異文化之間的相互交流，這些敏感的議題不必特別迴避，但是雙方容易因立場不同而產生對立或負面的情緒，所以教師必須以冷靜而客觀的方式加以處理，而參加的學生也要有心理準備，要能自我控制情緒。

三、討論的主題與內容不必畫框設限

國際文化的交流，就是要藉著討論而促進雙方的瞭解，討論的主題與內容愈寬愈好，只要不違反人類社會普遍的道德，任何的主題都是可以討論的，不必事先設定禁區，使任何一方產生說話的顧慮。

然而有時為了擔心觸犯某方面的禁忌而設法事先畫出框框，限制討論的範圍，例如：「這個話題不方便談」、「這問題已有定論，不必再談」、「這個已經有規範了，不須討論」、「我們今天不談政治」...等等，這些都是限制的方式，應儘量避免。事實上，話題與內容往往隨著兩方互動而不斷蜿蜒轉變，不易預先設定路徑及範圍，只要互動順暢，順其自然即可，只是參加的師生們都必須先有心理建設，必須以多元的觀點及開敞的心胸來面對這些話題即可。

四、以真實而自然的方式展現風格

每個社會都有各自有語言表達的方式及風格，無論是含蓄或直爽、冷靜或熱情，都形成各地有特色的文化差異，這些特色經常會從電視會議中不經意的展現出來，例如有些文化講求眾人和諧的氣氛，有些講求對外的口徑一致，有些講求個人獨立的想法，都是最真實

而鮮活的異文化展現，因此參加的雙方不必特別因電視會議而太過於客氣，特意隱藏這些自然的風格，過於客氣而生份的對話不足以深刻的相互理解，應鼓勵雙方保持自然的表現，只要維持應有的禮節即可。

伍、教學設計的過程

遠距教學的教學設計過程與一般課堂教學並無太大差異，但是設計所須顧慮的問題則較為複雜。下列是一般應具備的步驟：

一、瞭解雙方在交流方面的期待與限制

參加遠距教學的雙方必然對於合作關係及教學成果有所期待，在展開具體合作之初，必須先瞭解雙方期待能從遠距教學中達到什麼效果？對於雙方的助益有哪些？未來發展的前景如何？以及如何來滿足對方的期許？

此外，雙方也必然面臨一些交流限制，有些是制度上的限制，有些是資源的限制，例如人力資源、設備、及時間；有些是心理的限制，例如興趣與意願；有些是認知的限制，指雙方對於彼此社會文化的瞭解程度。

二、進行需求評估

需求評估(Needs assessment)是指為了推動遠距教學的各種需求，包括為何非要採取遠距教學的原因，並且評估各種影響交流的社會語言條件及因素，這些因素多半不是語言教學本身的問題，而是屬於社會文化的範疇。有些因素與教育體制有關，例如日本的學期與台灣的時間不同；有些是教學的習慣，例如日本每節課 90 分鐘，而台灣則習慣每節課 50 分鐘。有些是教學目標語的分歧問題，例如各地的語言標準不同。此外，雙方對於語言規範的處理觀念也須納入考量，以社會語言學的角度而言，語言是隨著社會的發展而不斷變化，必定因時間、地區及社會文化而有所不同，沒有固定不移的標準，因此跨文化交際的語言規範應是寬鬆而多元的。

三、訂定教學目標

依據社會互動的語言觀，教學目標不是學到了多少字彙量或語法點，而是實際的互動任務，因此一般的教學目標多為敘述式的，例如：

- 能夠面對漢語人士勇於發言；
- 能用漢語適度表達心裡的想法，並且使對方瞭解；
- 能藉由討論來瞭解雙方在文化及社會的異同；
- 能從語言的互動中，瞭解對方的立場及觀點；
- 能保持和諧的對話，並且啟發同理心。

至於特定的目標就必須與討論內容加以結合，例如：

能夠與台灣同學取得對於仿冒品的共識：
能夠瞭解台灣年輕哈日族的現象。

四、組織教學內容

社會互動觀強調有意義的溝通，為了表達意見而說話，不是為了練習語言而說話；是為了意義而討論，而非為了語言而討論；因此教學內容必須具有實際的意義及價值。此外，討論的內容雖然會隨著現場互動而流轉，但是仍應根據語言交際的活動階段來規劃說話的內容，先從暖身活動開始，到切入正題，到做結論，到最後收尾，將話題融入每一階段。

五、安排教學活動

教學活動宜減少單向表達的方式，而加強雙向互動式的活動，而且不僅是兩邊互動，也可以安排同一邊學生之間的互動，例如在家教式(tutorial method training)教學時讓學生進行角色扮演，或在電視討論會中，由同一邊的幾位參加者互相談論，將「兩邊對談」變成「多邊交談」。

六、準備教學材料

使用真實的材料(Authentic materials)是社會互動語言觀所強調的，因為電視會議是視覺與聽覺的溝通，因此所謂教材不僅是書面教材，也可包括圖片、音樂，甚至真實的物品，例如某次台灣師大與早大的電視會議中，談到一種台灣出產的水果稱為「釋迦」(一種形狀像佛頭的綠色水果)，而無論如何解釋也無法使早大同學明白，這時台灣師大的 Tutor 就必須拿來真正的水果，讓早大的學生觀看。

陸、電視會議參加者及遠距家教 (Tutor) 的認知

擔任遠距教學的教師或是家教(tutor)必須具備比一般的語言教師更多的技能與認知，下列數點是較為特別的問題：

1. 遠距教師的角色

在傳統的課堂教學中，語言教師都是直接的授課教師，多展現出語言訓練師的角色，但是在遠距教學的情境下，語言教師應扮演著輔助者、鼓勵者、及引導者。並且有時也成為文化的代言人，例如雙方的互動中經常會提出一些屬於集體文化的問題，例如：你們台灣人都喜歡什麼運動？你們台灣為何有哈日族？你們日本人為何喜歡泡溫泉？這時，教師或參加者就必須被迫扮演文化代言人的角色，以集體的身份來回答，而不是個人的立場了。

2. 語言的表達

社會互動語言教學觀強調真實而自然的會話，口語流利重於發音標準，表達意思重

於句法正確，因此以鼓勵對方能順暢的表達意思為主，Tutor 可以用正確的語調去重複對方的話語，但不必刻意去糾正對方的發音，以免製造對方說話的壓力。

此外，以社會語言學的觀點而言，在任何的國家或社會中，所謂口語標準之人數比例都極低，百分之九十以上的人士都帶有某些腔調，語言交際能力就是能在真實的環境中與帶有各種腔調的人士溝通，基於此，參與電視會議的人士也不必個個都字正腔圓，帶有各種口音也無妨，反而更有助於對方的學習。

至於口語的速度應儘量維持自然，以漢語為母語者的一般面對面快速交談可達每分鐘 250 字以上，正常速度也有 200 字，電視會議時可略降為 160 至 180 字，若在 150 字以下則太慢而不自然，因此應儘量維持正常速度，但每個句組之間可略為停頓，讓對方有思考的機會。

3. 異文化的理解

討論的內容既然與社會文化等方面有關，對於彼方的社會文化應有所理解，但是通常身處當地的教師往往容易理解當地的社會文化，例如身在日本的中國籍漢語教師很容易理解日本的社會文化狀況，但日本之外的遠距漢語教師就不見得瞭解日本，因此如何充實教師對於異國文化的認知及對於異國學生的瞭解，應是教師培訓的重點。目前參與早大遠距教學的台灣師大華研所的研究生仍是以具有日語背景者為主，未來若要納入非日語背景的教師，就必須設法加強他們對於日本社會文化的理解。

4. 後設認知(Metacognition)的能力

家教式的電視會議教師所需具備的技能很多，包括語言能力、教學能力、溝通能力、異文化認知、甚至網路電視會議的操作技術，但是最重要的能力卻是後設認知(Metacognition)的能力。家教式的電視會議教師必須具有一心多用的能耐，能夠一方面與對方學生討論，緊扣住談話的線索，一方面要能平衡照應學生，又要掌握時間，這些都是一種後設認知(Metacognition)的能力，類似電視新聞主播的角色，一方面要面對鏡頭播報新聞，另一方面又要隨時注意周邊的狀況，這種能力需要特別的訓練以及經驗的累積。

結語

遠距教學將原本遙遠的人群與文化直接帶進了課堂，使得語言學習有了全新的面貌，也使得語言教學法有著新的發展。而在語言學習之外，不僅開擴了學生的心胸及眼界，也促進了國際間異文化的理解，這將是遠距教學合作雙方最重要的收穫。

參考資料：

- Byram. M. (1989). *Cultural Studies in Foreign Language Education*. Clevedon, UK: Multilingual Matters.
- Richards J.C. & Rodgers, T.S. (1986). *Approaches and Methods in Language Teaching*. UK: Cambridge University Press.

- 砂岡和子、村上公一、小川利明、于洋（2000）早稻田大學中國語遠程教學演習。第二屆全球華文網路教育研討會論文集。頁 66-72。
- 信世昌（2002）網路華語文教學之實施經驗與探討。國際漢語教學研究 2002 年特刊。頁 81-83。
- 信世昌（2001）網路中文應用文寫作課程之設計與實施。第二屆全球華文網路教育研討會論文集。頁 117-123。
- 信世昌（1999）使用視訊會議系統做為遠距語言教學之探討。第三屆國際電腦多媒體語文教學研討會論文集，頁 387-394。

原載「中国語国際遠隔教育の実践と展望」論文集（2002 年 2 月）