

コーパス利用による中国語教育(2)

－可變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

I 語彙教学の指標

加速する国際化と、情報通信技術の進歩がもたらす学習環境の多様化を背景に、第二言語〔以後L2と略称〕習得の手段が大きく変化しつつある。現地への留学や旅行、インターネット情報、E-mail、あるいは携帯電話、TV会議と、母語話者との接触機会は格段に増した。海外生活経験を有する学生も増え、実践に耐える第二・第三外国語を在学期間中に習得することは、学生にとって現実的な関心事である。

大学の語学教育の理念は、外国語学習を通じて国際社会に対する深い教養を持つ人間性の涵養であり、実用会話や試験対策だけを達成目標とするわけではない。しかし一定の語学スキルを身につけなければ、高邁な教養主義も絵に書いたモチに終わろう。

人対人の伝統的コミュニケーション能力の向上が叫ばれる一方、昨今は電子辞書や、Web検索、自動翻訳、対話システム技術の精度向上により、人間の言語運用をコンピュータプログラム上で実現する動きも加速している¹。大学のL2教育も学習環境の変化に適応したカリキュラム改革を実践に移し、語学力低下を食い止めなければ、軽少短薄なテキストの横行、文法・講読の偏重、日本語多用の外国語授業など、悪循環を断ち切ることは難しい。

¹ 砂岡和子2004年「网络汉语口语课本和词汇分级」

言葉の運用に必要な情報は、語彙とディスコース²理解であろう。中でも語彙学習は言語習得の第一歩であり、L2習得においても基本知識とされる。中国語は文法的形態特性に乏しく、語彙結合のルールが多く意味に依拠して形成され、しかも文構造規則が語彙結合のルールと相似する特徴を有する。中国語習得に占めるディスコースと語彙学習の重要性は、他の言語のそれに勝ると言ってもよい。市場に溢れる速習用単語集は、語彙力増強を願うL2学習者の多さを物語っているが、実際には日本人学習者のポキャブラリー不足が、聞き取りや読解力に支障をきたしていることは、多くの研究者が指摘するところである³。

では、日本におけるL2中国語の場合、どの程度の語彙力が必要なのか？本来、語彙力とは単に語彙数の多少ではなく、実践に耐える運用力を備えた獲得語彙数であるべきだが、検証が難しいため、ひとまず語彙サイズで見てみることにする。TECC (Test of Communicative Chinese)⁴のデータで、中国語コミュニケーション能力と習得語彙サイズの相関を見ると、語彙数約1,000語が初級、語彙数3,000語以上で日常会話レベル、5,000語以上から現地での留學生活が可能水準とされる〔表1参照〕。同様の相関傾向は、英語など他言語でも観察されており、L2の一般的語彙指標といえよう⁵。

² discuss。言葉がコンテクストの中でどう使用され、どのような意味を持つかを分析する概念。ディスコーステキストとも言う。

³ 尹景春1995

⁴ 中国語コミュニケーション能力検定 <http://www.tecc-jp.com/about/>

⁵ 史有為2004年

表1 中国語レベルと習得語彙数⁶

語彙数	HSK	TECC 得点	コミュニケーション能力指標
8,000語	7級、8級 300～336	A 1,000-900	専門家レベル。微妙なニュアンスの理解・伝達ができ、ビジネスや生活のあらゆる面でネイティブと十分なコミュニケーションができる。
5,000語	6級 263～299	B 899-700	ビジネス上の交渉・説得も可能なレベル。中国の文科系大学の学部入学最低基準。
3,000語	5級、4級 210～300	C 699-550	日常生活や一般交際に必要な基礎中国語レベル。中国の大学の理、工、農、西洋医学の本科に留学する最低限の中国語能力レベル。
2,000語	3級 155～209	D 549-400	挨拶はある程度でき、日常生活の基本単語が聞いてわかるレベル。
1,000語	2級 100～154	E 399-250	相手が配慮してくれる場合、ある程度の会話ができるレベル。
500語以下	1級	F 249-	簡単な依頼や買い物、旅行など、特定の場面や限定的な範囲でのコミュニケーションができるレベル。

II 中国語履修者の習得語彙サイズ

では日本における中国語履修者の語彙習得の実態はどうであろうか？上級者、中級レベル、そして圧倒的多数を占める初級学習者について、実際の試験結果をもとに、習得語彙数を比較してみよう。

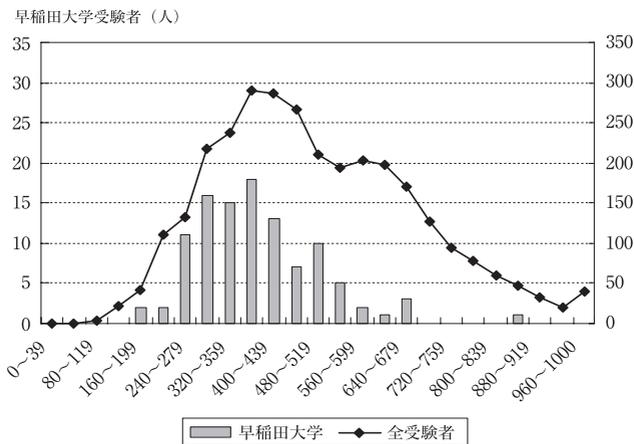
上で紹介したTECC受験者は、中国語実務経験をもつビジネスマンや、中国語能力をアップしたいと考える就職活動中、もしくは卒業間近の学生など、モチベーションの高い受験者が多く、日本のL2中国語学習者のなかでは高めの得点層を占める。学習目標がそれほど明確ではない一般の日本人L2中国語学習者は、1,000語どころか500語を使いこなせない場合が多い。2003年7月実施のTECC試験における早稲田大学および全受験者の得点別分布データを見ると、習得語彙500語以下が受検者の15%、1,000語以下は40%強を占めている〔表2、表3参照〕。

⁶ 同上ベネッセ(株)TECCデータHP

表2 2003年7月TECC試験得点別分布データ

得点	早稲田大学 (人数)	全受験者 (人数)	得点	早稲田大学 (人数)	全受験者 (人数)
0～39	0	0	520～559	5	194
40～79	0	0	560～599	2	204
80～119	0	3	600～639	1	197
120～159	0	21	640～679	3	170
160～199	2	42	680～719	0	127
200～239	2	111	720～759	0	94
240～279	11	133	760～799	0	78
280～319	16	218	800～839	0	59
320～359	15	237	840～879	1	48
360～399	18	291	880～919	0	33
400～439	13	287	920～959	0	20
440～479	7	266	960～1,000	0	39
480～519	10	211	計	106	3,083

表3 TECC試験得点分布比較 2003年7月実施



大学一般教養レベルの履修者の習得語彙数となると、TECC受験者よりさらに少ない。中国の対外漢語教学大綱⁷では、履修時間と習得語彙数を下記〔表

⁷ 2002年「高等学校外国留学生語教学大綱」

4] のように定めている。しかし日本の中国語教育では、個別の機関や教員を除いて、カリキュラムに漢語教学大綱準拠を明示する機関は少ない。

表4 中国対外漢語教学大綱の語彙数

対外漢語大綱			参考 (HSK大綱) ⁸	
レベル	履修時間	習得語彙指標	レベル	語彙数
高級2級	400時間	1,397語	丁	3,562語
高級1級	400時間	1,400語		
中級4級	200時間	750語		
中級3級	200時間	700語		
中級2級	200時間	700語	丙	2,202語
中級1級	200時間	700語		
初級4級	200時間	700～750語以下	乙	2,018語
初級3級	200時間	650語		
初級2級	200時間	562語	甲	1,033語
初級1級	200時間	500語		
	2,400時間	合計 8,042語		合計 8,822語

カリキュラムでは語彙獲得指標が明示されないため、Tutorial Chinese プロジェクトが開発中のTACCレベル分けテスト⁹の得点結果から、大学一般教養中国語履修者の語彙数を測定してみる。著者が授業を担当する早稲田大学政治経済学部的一般教養中国語履修者（中国語Ⅰ、中国語Ⅱグループ）、Tutorial Chinese受講者グループ（中国語Tutorialグループ）、および中国語を母語とするネイティブ群（ネイティブグループ）にTACCを受験してもらい、相互に比較した。Tutorial受講生は、語学力により数十クラスにレベル分けされているが、本サンプルでは初級レベルの学生の語彙得点を抽出してある。TACCは以

⁸ 2001年「語水平詞彙与漢字等級大綱（修訂本）」

⁹ Test of Ability for Chinese Communication。早稲田大学Tutorial Chineseのクラス編成と成果測定に使用するため開発した。

下の3種類の設問があり、TOTAL100点満点で評価する¹⁰。

Sec 1 中国語の単語（音声）を聴いて意味を選択する問題（40問）

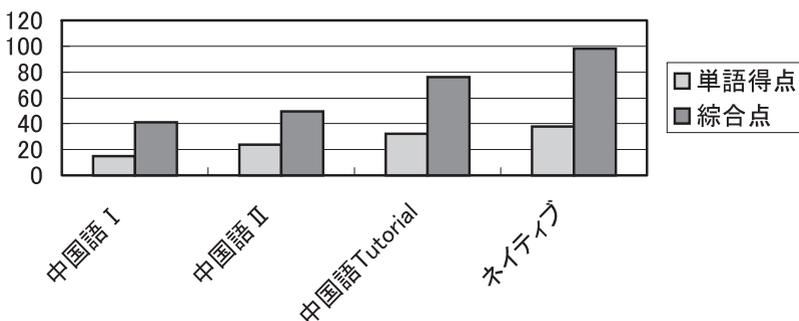
Sec 2 中国語（音声）を聴いて、続きとして適当な語句を選ぶ問題（40問）

Sec 3 中国語（音声）を聴いてリピーティング（復誦）する問題（10問）

2004年9月TACC試験結果を、一般教養中国語履修者（政経学部中国語Ⅰ、中国語Ⅱグループ）と比較すると、Tutorial Chinese受講者の語彙力が高いことが分かる〔表5参照〕。また総得点とSec1単語聞き取り問題には相関が見られることから、Tutorial Chinese初級クラスは教養クラスに比べ、相対的に中国語力も高いといえよう〔表6参照〕¹¹。一般教養中国語の履修時間は実質年120時間程度であるから、〔表4〕の中国対外漢語教学大綱語彙数で換算すると、習得語彙数は500語を大幅に下回り、300語程度ではないかと危惧される。

表5 TACC総合と単語得点の相関

クラス別得点



¹⁰ 2005年春からはCAD（項目応答理論を利用した加除式試験となり、設問数を各Section10題とした）。

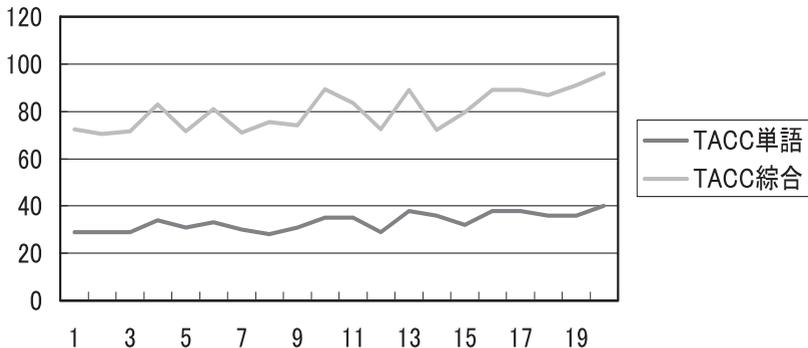
¹¹ TutorialグループはTECC試験も受験しており、TECCとTACCの得点の相関も認められる。

表6 日本の大学L2中国語履修者の習得語彙実態

クラス	学 習 歴	受験者(人)	年間履修時間	単語得点	総合点	予測語彙数 [≒]
中国語Ⅰ	政経教養中国語1年	74	120 h	15.09	41.17	300
中国語Ⅱ	政経教養中国語2年	12	120 h	24	49.5	500
中国語 Tutorial	Tutorial Chinese 初級	68	120h (注)	32.16	76.02	2,000
ネイティブ	中国語を母語とする被験者	4		38	98	8,000以上

(注) 中国語 Tutorial は半期科目で、周1回クラス〔半期60 h〕と、周2回クラス〔半期120 h〕の2コースがある。Tutorial履修生の受講時間は数字の上では一般教養中国語と同じであるが、実際は予習復習時間に授業の前後、平均1～1.5時間かける学生が多い。

表7 TACC総合と単語得点の相関



以上、各レベルのL2中国語学習者の語彙サイズを調査し、特に初級者や大学一般教養語学履修者の語彙不足を確認した。日本人学習者の語彙不足の原因は、①カリキュラム編成上の絶対的履修時間不足、②L2必修化による学習動機欠如など、制度に起因する問題、③日本社会での中国語、ことに音声言語としての認知度の低さ、④講読・文法中心の学習、⑤漢字を介した視覚的記憶に固執する日本人L2学習者の特性、などを挙げることができよう。履修時間

の大幅な不足、学習意欲の低い、単位取得目的の履修者を相当数抱えていることから、L2の目標値を設定しても達成が見込めず、教学の大綱化が進まない¹²。こうした不毛な現状は改善すべきである。統一的教学指標がなければ、カリキュラム編成や、試験の効果が見えにくい。

しかし実際に学習大綱を定める段になると、それぞれの教育現場で困難に直面する。語彙サイズは、学習時間に比例して上掲「中国対外漢語教学大綱」の、初級1,000語、中級3,000語以上、上級5,000語以上を受容できよう。ところが具体的に語彙範囲を決めるとなると難しい。専攻・地域・年齢・職業など学習環境と目的によって、「常用語彙」の内容に大きな差異がある。HSKは歴代各種語彙統計に基づき、語彙の選定とレベル分けを行っているが、しかし“因特網（インターネット）”“手机（携帯電話）”“超級市場（スーパー）”“打工（アルバイト）”のような、既に「新語」とはいえないほど普及した日常語彙が未登録なのは勿論、日本人にとって常用語彙である“棒球（野球）”“地鐵（地下鉄）”がともに難度の高い丁級語彙になっている¹³。“日本（日本）”“美国（アメリカ）”など固有名詞をまったく含まないなど、現代中国語を学ぶ上で問題が多い。

HSKを代表とする中国製常用語彙教本のより根源的な問題は、これら常用語彙の資源が、依然古典中心である点である。常用語彙選定作業には、当初から、歴史的啓蒙書や唐宋詞中の漢字が資料として用いられた。最近の統計によれば、「三字経」「千字文」など童蒙書は、現代漢語常用漢字の70%以上をカバーしており、中国国定教科書が採用する「千家詩」の唐宋詩詞に至っては、現代漢語常用漢字の90%以上を網羅するという¹⁴。漢字を唯一の文字とする中国では、初等国語教育は識字教育に費やされる宿命にあり、膨大な漢字学習を

¹² 2004年史有為「中国語教育に於ける語彙の位置付け－中国語教育研究その4」

¹³ 2001年「汉语水平词汇与汉字等级大纲（修订本）」卷末索引参照

¹⁴ 鄭国民他2003年「小学校文常用読物的字種与字量研究」

通じて語彙も同時に習得する。中国の伝統的国語教育を振り返れば、「三字経」「百家姓」「千字文」「大公家教」など児童向け啓蒙書は、すべて識字教本であった。中国の国語教育は古代から一貫して漢字学習と平行して語彙教育が行われて来たと言える。これらテキストは単漢字の出現率が圧倒的に高く、この資料に基づいた「常用語彙」使用調査は、実際は「常用漢字」の統計に等しい。加えて中国の児童向け啓蒙書は儒教的思想を反映し、唐宋詞の語彙も韻律や時代的制約がある。収録語彙1万語規模の教育用常用語彙表は、事実上HSKが唯一で、単語本やCALLドリルにはHSK準拠を謳うものが多い。しかしHSKは1990年代初頭の制定で、かつ静態データのため、日常生活語彙の欠落は否めず、語彙の個別カスタマイズにも対応が難しい。近年、現代語の語彙調査成果を反映し、修訂を重ねてきたが、依然として書面語の語彙が多く、現実の言語生活との乖離が大きい¹⁵。

語彙学習の占める重要性は大学語学授業でも不変であるが、ディスコース理解を離れた語彙学習が非現実的であることを知る所以に、日本の教学現場では自由な語彙教育を求める傾向が強い。市販教材の収録語彙基準も科学的根拠に欠けるため、語彙学習を正規語学カリキュラムの中心に据える科目は少なく、現状は正規授業の補完的学習に止まる。

Ⅲ 可変的語彙教育を支援する中国語語彙コーパスの構築

早稲田大学Tutorial Chineseプロジェクトにとって、質量ともにいかに効果的な語彙教育を行うかは、重要な教学課題である。Tutorial Chineseは対面式少人数会話レッスンと、北京・台湾とのライブ遠隔授業を組み合わせた、新しいL2学習モデルである(図1、図2参照)。異文化コミュニケーション体験を通し、学習者のモチベーション向上と語学力の強化をはかることを主要な教学目

¹⁵ 史有為2004年「关于日本中国語教育初級词汇的统计研究和語篇研究」、張延瑞2004年「若干大学教材的初步统计报告」参照。

的とする。多地域にまたがる異なるレベル・異なる内容の複数クラスを教学管理し、母語話者との実践的言語コミュニケーションを通じて学力を向上させるためには、通常の教場授業に必要な基本スキルに加え、動態教学情報の管理が必要となる。情報の迅速なフィードバックを保障する通信環境と、習得事項の定着を支援する学習環境が不可欠である。Tutorial Chineseプロジェクトでは、専用のWeb教学管理サーバーを設置し¹⁶、レベル分けWebテスト、ユーザー登録、レッスン管理、授業コメント配信、Web作文添削指導、BBS投稿、テキスト学習サポート情報など、HPと連動して実践体験を学力として蓄積する各種教学支援を行っている（図3参照）。



図1



図2

Tutorial Chinese授業の最優先課題は、語彙量の拡大である。最下級クラスの学生の登録時点の語彙数は500語程度で、ネイティブスピーカーの中国語を聞き取るのに到底及ばない。とにかく聞き続け、5、6回授業に出席すると、ほぼ半数以上が話しの内容をおおよそ理解できるまでに語彙数を増やしてゆく。その後は、語彙の運用法を学び、自ら発言する段階へ移行する。数回で聞き取り能力が急速に向上するのは、Tutorial Chineseがディスコース主導の語彙教育を行うことによる。言葉は目標言語の言語体系の中で学んでこそ習得できるという考えに立ち、実際の場面と結びつけた実践学習を行っている。一般の教場外国語学習では、受講生が多く、レベルや学習動機もまちまちで、個別能

¹⁶ <http://www.tcwaseda.net/tc/>

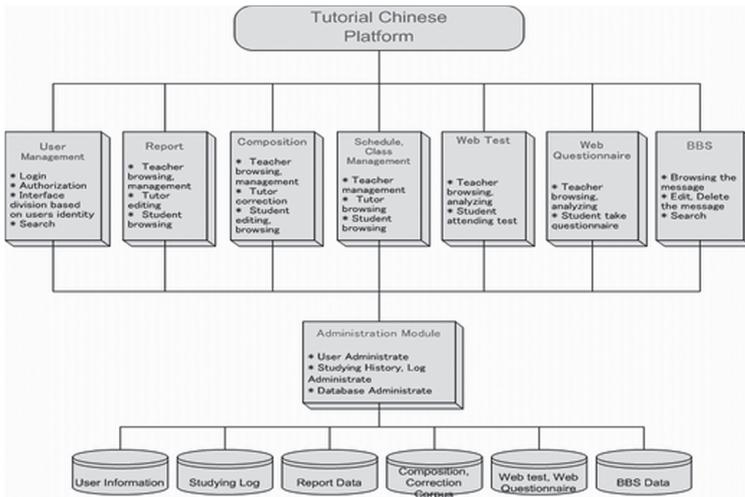


図3 TCP Module Structure 劉松 [GITS] 作成

力と関心に応じた教学のカスタマイズができず、往々にして範文暗記を強要する句型練習に傾りがちである。弾性に欠ける句型練習は語彙産出力の向上に寄与しない。ことにTutorial Chineseのように母語話者との協調型ライブレッスンでは、場面の展開が予測不可能であり、定量化した語彙や文型項目提示だけでは学習支援とならない。

以下、実際に授業で使用した語彙と、テキスト指定の語彙が、どの程度差異があるのか、Tutorial Chineseの授業記録で具体的に比較してみよう。

(表8)は2004年10月下旬の1週間、Tutorial Chinese各クラスで使われたテキスト所出以外の「新出単語」データである¹⁷。Tutorたちはそれぞれ担当クラスの授業で使った語彙を、専用HPの授業報告欄に手書き入力する。(図4「Tutorial Chinese HP TutorReport 新出単語揭示欄」参照)。

この週の討論テーマは「出事」「トラブル発生」であった。折りしもこの時

¹⁷ 実際は、各Tutor誤認により、テキスト既出の語彙も多く含まれる。(表9)と対照比較のこと。またテキストの他課にも15課の関連語彙が多数掲載されており、受講生は学習の機会がある。

期、日本で台風23号による豪雨災害や新潟中越地震が発生し、遠隔授業交流先の台湾でも、水害被害が報じられた直後であったため、授業は討論テーマの災害関係に話題が集中し、通常レッスンと比べて、テーマ密着型の語彙使用が顕著であった。他の授業では、語彙の拡散がより大きい。

(表9)「第15課所出語彙一覧」と比較すると、ライブレッスンで使用された語彙群には、一目して以下の使用特色を見て取れる。

- 1 クラスごとに使用語彙が分散している。〔例：対面-Rと対面-Lクラス〕
- 2 クラスの語学レベルと、使用語彙の難度には、明確な差が見えない。〔例：対面-L初級と対面-M上級クラス〕
- 3 対面クラスの使用語彙のほうが、遠隔クラスのそれより、広範囲、かつ話題密着タイプである。〔例：対面-A、Mクラスと、台湾遠隔-Cクラス〕
- 4 実際の対話によって生まれた語彙もしくはフレーズ形式で提示され、教科書のように系統的に現れない。使用文例とともに提示するTutorもあり、ディスコースと結びついた新出単語提示は受講者にとって記憶に留めやすい。〔例：対面-Gクラス〕

表8 ライブレッスン新出単語一覧

共通討論テーマ「“出事”トラブル発生」 2004年10月26日—11月1日授業実施

クラス名	Tutor	レベル	レッスン中の新出単語 (Tutor 記入ママ)
対面-R	A	初級	滑, 瘫痪, 直升飞机, 劫机, 海啸
対面-L	B	初級	干燥, 燃烧, 倒塌, 道路, 预告, 根据, 震灾, 备用, 坠毁, 空难, 裂开, 情味, 扒手, 酷暑, 出海, 洪水, 敞篷车, 面包车, 小卧车, 小货车, 猜测, 波浪, 海啸, 劫机
対面-E	C	中級	摇, 倒, 窄, 行人, 斑马线, 流氓, 强盗, 扒手, 犯罪, 医药费
対面-F	D	中級	摇醒, 流血, 墙壁, 缝 (了五) 针, 瓦斯, 躲到, 手枪, 派出所, 电脑中毒, 凶

対面-G	E	中級	爬到树上, 强盗, 一艘轮船, 倒塌(房屋倒塌了), 劫持, 乘客, 燃烧(火燃烧起来了), 报警, 架(一架飞机), ~级(日本发生了6级地震), 人质(把乘客作为人质), 救命啊!, 损失, 察觉(他察觉到有人在偷他的钱包)
対面-K	F	中級	洪水, 劫持, 告示牌, 人质, 失火, 空难, 生命危险, 遇难
対面-N	G	中級	楼房, 倒了, 倒塌, 裂了, 地裂, 跑, 逃, 体育, 男的, 女的, 冲走了, 防震棚, 卧轨, 卧室, 飞行员, 枪, 恐怖分子, 吵嘴, 打架, 爬, 爬树, 掉下来, 飞机失事, 坠落
対面-M	H	上級	坠毁, 劫机, 空难, 规划, 海啸, 扒手, 天干物燥, 天灾人祸, 趁火打劫, 雪上加霜, 雪中送炭, 倾斜, 倾倒, 便道, 非机动车
対面-A	I	上級	倒塌, 冲走, 洪涝灾害, 天灾人祸, 逃生, 吃一堑长一智, 私了, 黑车, 瞌睡
対面-Q	J	上級	统计, 妥善, 搜索, 损失, 数字, 信号, 逮捕, 援助, 地铁, 一连串, 一则, 老年人, 赔偿, 无法, 协商, 解决, 案件, 讹诈
遠隔 台湾-B	K	初級	地震, 發生, 出事, 警察, 立刻, 安全, 危險, 失火, 失事, 救火, 滅火, 防火, 受災, 人數, 統計, 馬虎, 爆炸, 恐怖, 小偷
遠隔 台湾-C	L	中級	搖晃, 電線桿, 洪水, 意外, 消防隊員, 勇敢, 技術, 水桶
遠隔 台湾-D	M	上級	地震, 火災, 放鞭炮, 原來, 失火, 水災
遠隔 北京-G	N	初級	地震, 发生, 出事, 警察, 立刻, 安全, 危險, 妥善, 失事, 失火, 救火, 灭火, 防火, 受灾, 人数, 统计, 水灾, 丢, 马虎, 注意, 爆炸, 恐怖, 出现, 小偷, 车祸, 处理, 事故, 骑, 撞, 曾经
遠隔 北京-L	O	初級	地震, 发生, 出事, 警察, 立刻, 安全, 危險, 妥善, 失事, 失火, 救火, 灭火, 防火, 受灾, 人数, 统计, 水灾, 丢, 马虎, 注意, 爆炸, 恐怖, 出现, 小偷, 车祸, 处理, 事故, 骑, 撞, 曾经, 手表, 硕士, 研究生, 厉害, 体育馆, 藏, 跟, 一样

表9 テキスト第15課所出語彙一覧 部分 (ローマ字注音順に並べ替え)

安排	从	会	美国	塞车	通知	因为
安全	打	火灾	迷路	三级	统计	银行
安全帽	打算	给予	灭火	伤亡	突然	应该
把	大	技术	灭火器	上学	推销员	有
办	大家	交通工具	摩托车	生命	腿	遇到
保护	大学	交通警察	哪里	什么	脱离	原来
报道	逮捕	交通事故	那里	声音	妥善	原因
报警	到底	结果	哪儿	失事	外面	援助
爆炸	的	今年	呢	时候	危险	院子
被	得到	警察局	你	事	问题	在
比较	地点	救火	浓烟	是	我	造成
不是	地铁	卡车	朋友	事故	我们	怎么
不要	地震	考试	平时	事件	现场	着火
不要紧	都	可怕	破坏	事情	相信	找
布告栏	堵车	可以	骑	受害人	消防队员	这里
不可	对	恐怖	起火	受伤	消息	政府
不少	对方	恐怖主义分子	企业家	受灾者	小偷	知道
财产	多	快	汽车	树	小心	重地
仓库	而	那么	钱包	数字	新生	重要
曾经	发生	垃圾	前面	水灾	信用卡	住宿
常	飞机	里面	清楚	说	幸好	注意
常常	服务生	立刻	情况	死亡	学生	撞车
长江流域	刚才	练习	请	损失	学校	状态
超过	给	乱	请问	所以	严重	自行车
车祸	根据	轮船	确定	他	要求	最
车子	工作	麻烦	人民	她	要	最近
出	官方	马虎	人数	台风	要是	做
出来	关于	马上	人员	特别	一定	作弊
出事	还	满意	任务	天干物燥	以后	
出现	很	没	日本	天灾	已经	
处理	后来	没有	如果	听说	一点儿	

Ⅳ 中国語語彙コーパスデータの収集と設計

以上のように、教科書で提示する語彙情報は、実際のレッスンに必要な語彙をカバーできないことが明らかである。同様の状況は、通常教室授業でも起こっているが、教師が学習範囲外の新語彙使用を自制するか、学生の質問があれば、板書して終わることが多い。オーラルコミュニケーション重視の外国語学習では、こうした「番外」語彙の出現を疎んではならない。むしろディスコース理解と一体になった語彙学習ととらえ、積極的に習得すべきであろう。もともと自然な対人コミュニケーションでは、生きた言語との遭遇が避けられない。初級段階でも学習者は辞書にない語義や用法、新語や地域特有の表現、個人差による音声のバリエーションなど、従来の学校文法では教えない言語現象にぶつかる機会は多い。

このような動態言語現象を合理的かつ効率よく習得するには、電子コーパスの利用が有効である。広範な日常生活の語彙を収集・蓄積したコーパスソースに、きめ細かく使用域レベルを表示すれば、コーパスがネイティブスピーカーに替わり、言語事実の適合性や語感の確認に辛抱強く応えてくれ、言語直観の働かない外国人学習者でも、平均的な語感と、実態に即した正確な言語事実を、自立的に発見可能となる¹⁸。

我々は中国語語彙コーパスを構築して、教場で出現した事例に結び付いた語彙学習と、学習者の理解度に応じた発話生成の規則習得ルールを提示し¹⁹、様々な状況に対応できる柔軟な言語能力の養成に役立つツールを提供したいと考えた。この中国語語彙コーパスをWTCC (Waseda Tutorial Chinese Corpus) と呼ぶ。

¹⁸ 砂岡和子、尹明2003年「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-」PP.250-253

¹⁹ 文型コーパスの資源はあるが、時間的な制約により、コーパス作成は後日とする。

a) WTCCのコーパス資源

最初に、WTCCのデータソースに関して述べる。WTCCは早稲田大学オープン科目、2005年度 tutorial Chinese のテキスト所載語彙約8,000語〔異なり語数約2,500語²⁰⁾〕を核に(表9参照)、機械言語処理用の北京大学計算語言学研究所 The Grammatical Knowledge-base of Contemporary Chinese corpus 10,000語²¹⁾、対外中国語教育用語彙大綱である中国漢語水平考試甲乙丙約8,000語、および日本の大学教育用基礎語彙コーパス約3,000語の語彙情報²²⁾から、それぞれ難度、品詞、日本語訳、音注などの言語情報をタグ付けした電子コーパスである。L2学習者の基本語彙を満たすと同時に、難度表示には、HSK(漢語水平考試)甲乙丙を基準に、これを数字の1~3度に置き換えて自動付与し、HSKにない語彙に関しては、北京大学計算語言学研究所通用度統計数値を参考にした²³⁾。

こうして既存コーパス資源をもとに基本語彙を蓄積したあと、WTCCは更に、①CALL-drill作成用データベース、②レベル分けWeb-test TACC²⁴⁾の問題作成用データベース、③語学分析用データベースとして、学習・教育・研究の

²⁰⁾ Tutorial Chinese テキストの語句は、北京大学計算語言学研究所編「現代汉语短语结构知识库 (The Phrase Structure Data-Base of Contemporary Chinese 句型からテーマ別に選別し、日本の学習環境に応じてリライトした文を使用している。例文提示に当たっては学習効果を配慮し、前後に2回から3回、同一語彙が重複して現れるよう編集した。テキスト所載語彙は約8,000語だが、実際の異なり語数が約2,500語と三分の一に減っているのは、このためである。

²¹⁾ 同研究所『現代漢語語法信息詞典』(The Grammatical Knowledge Base of Contemporary Chinese) 同名の辞書の電子版。50万語余の動詞、形容詞の属性記述(タグ)を収録。収録語7.3万語。

²²⁾ 史有為2004年1月「关于日本中国語教育初级词汇的统计研究和语篇研究」

²³⁾ 朱学峰、張化瑞、段慧明、俞士汶2004年8月『漢語高頻詞語法信息詞典』的研制

²⁴⁾ TACC (Test of Ability for Chinese Communication)、Tutorial Chinese クラス受講者レベル判定のためのコンピュータテストシステム名、開発元 Waseda University International

用に耐えるデータ資源を構築する。このうち、現在、実際に使用開始しているのは、①と②である。

WTCCはTutorial ChineseのTutor Reportと連動して、新語情報を更新し(図5参照)、さらに授業で生成される新出単語を自動的に取り込むプログラミング設計を目指している(図4参照)。TutorがReportに入力した新出単語は、先に登録済みの本文語彙情報と辞書情報に従って、単語に切り分けられる。それぞれ中国語、日本語訳、発音情報、難度、テキスト番号が振られて、既存コーパスに蓄積される仕組みだ。この拡張コーパス情報は、さらにCALLドリルの作問庫にストックされてゆく(図5参照)。

Report list(Teacher)	
The detail about report 2360	
Tutorial 班级	遠隔-L
日期	2004-10-28
时限	7限 19:40-21:10
授课方式	远程
《Tutorial汉语》	第15課 出事
辅导教师姓名	李益
辅导教师所属单位	なし
今天的生词 (中间请用“,”分割。 例: 天气, 气候, 下雨)	地震, 发生, 出事, 警察, 立刻, 安全, 危险, 妥善, 失事, 失火, 救火, 灭火, 防火, 灾难, 人数, 统计, 水灾, 丢, 马虎, 注意, 爆炸, 恐怖, 出现, 小偷, 车祸, 处理, 事故, 骑, 撞, 曾经, 手表, 硕士, 研究生, 厉害, 体育馆, 藏, 跟……一样
学生感兴趣的话题	1、请新同学与其他同学互相认识。 2、如果发生地震, 你怎么办? 3、在日本, 如果发生车祸, 怎么办? 4、在日本, 如果发生火灾, 怎么办? 5、看图说故事。
对课本的具体建议 (不包括对拼音注解错误的修改)	P86的某些生词在课文中没有体现, 如“灭火、防火、爆炸”等。

図4 Tutorial Chinese HP TutorReport新出単語揭示欄

注: エクセルに自動転記でき、作問資源として使いやすい形式になるよう、Tutorは新出単語Reportのさい、単語をカンマ(,)で区切るように指示がある(中间请用“,”分割。例: 天气, 气候, 下雨)。

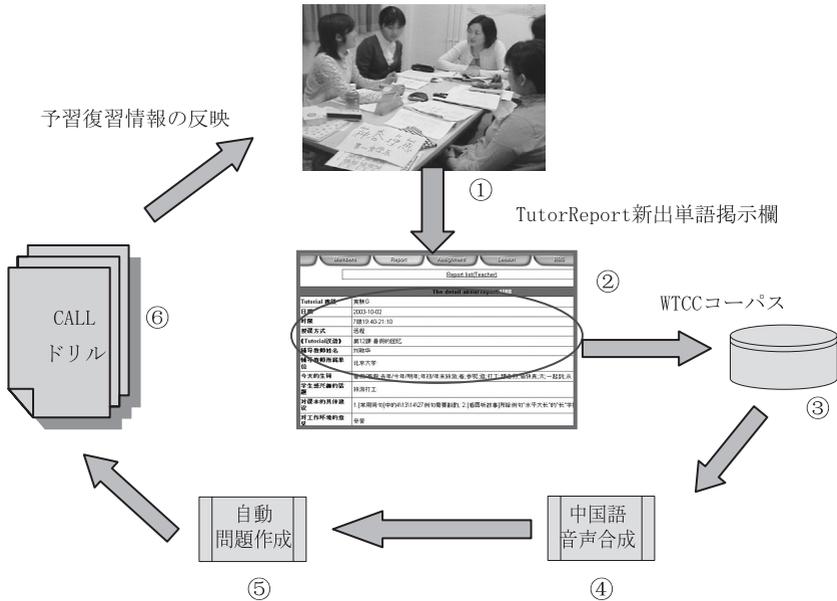


図5 WTCC コーパスプログラム設計図

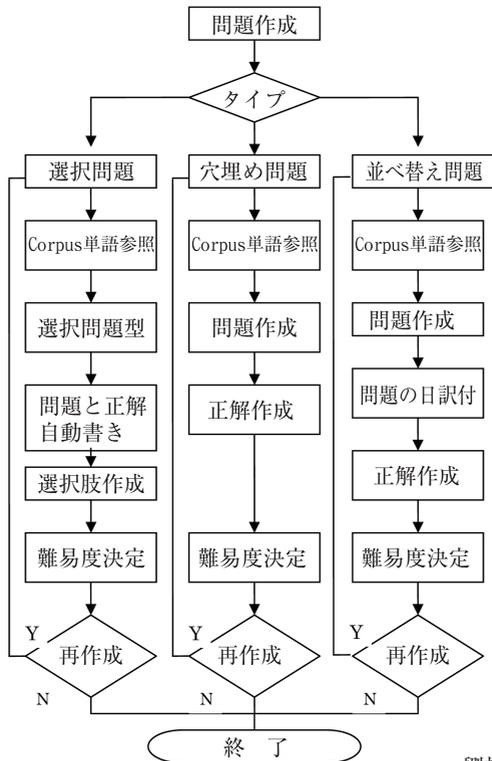
表10

课数	中文意	日文意	词性	拼音	声音	难度级别
15	出事	事故が起きる	vo.	chul shi4	102100070	3
15	到底	いつたい	adv.	dao4 di3	202300130	2
15	马上	すぐ	adv.	ma3 shang4	201300400	1
15	警察	警察	n.	jing3 cha2	202100320	2
15	迷路	道に迷う	vo.	mi2 lu4	202300040	3
15	警察局	警察署	n.	jing3 cha2 ju2	202200060	3
15	丢	なくす	v.	diu1	102200200	2
15	生命	生命	n.	sheng1 ming4	202100010	3
15	财产	財産	n.	cai2 chan3	102100440	3
15	危险	危険だ	adj.	wei1 xian3	206100120	2
15	小路	路地	n.	xiao3 lu4	2011005170	3
15	脱离	脱する	v.	tuol li2	201200120	3

V PC、携帯端末対応の中国語ドリルの開発

本文テキスト語彙と新出単語を加えた作問庫をベースに、クラスや難度別に応じたCALLドリルが出問される。ドリルの種類は、現在3種類が稼動中であり、問題の種類は10種類のバリエーションを用意している。問題作成フローは（表11）の通りである。受験者の学習履歴も自動的に残され、随時検索して参照できる。

表11 問題作成フロー



劉松〔GITS〕作成

WTCはコンピュータベース主体の開発が先行したが、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の環境整備が進むことを想定し、携帯電話でのCALLドリル配信実験にも着手した。これが、すでに実用実験

の段階に達している²⁵ (図9、1参照)。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、将来は一台の携帯電話で世界と交信できるユビキタス・ラーニングが可能となろう。



図6 CALLドリルの半自動化作問

劉松〔GITS〕作成

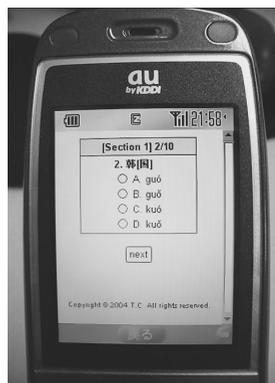


図7 携帯ドリル受験画面

現在、Tutorial ChineseのPC版練習問題は「穴埋め」、「選択」、「並べ替え」の三タイプを用意するが、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、パスワードを入力して認証後、操作を開始する。

携帯電話版CALLシステムもPC版同様、多言語対応である。日本語、簡体字中国語、繁体字中国語など、多種類の文字コードをサポートしており、プログラムで自動的にUnicode文字を画像データに変換する。伝送には画像圧縮技術を利用し、約1/8～1/10の圧縮率で携帯画面に表示できる。現在の携帯の伝送速度能力であれば、画像データ閲覧に支障はない(図7参照)。

CALLドリルシステムは「練習問題作成」、「練習ペーパー作成」「練習画面」、及び「コーパス管理」の四部分から構成され、それぞれの機能分担は以

²⁵ ただし機種限定。劉松、砂岡和子、浦野義頼2005年3月「PC、携帯端末に対応するコールドリルシステムの開発」

下のようなのである。

「練習問題の作成」：テキスト単語コーパスと連携し、半自動で作問する。

「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を行う²⁶（図6参照）。

WTCCはCALLドリルシステムの作問エンジンとして稼動し、上で紹介した新出単語情報と連携して常に更新を繰り返す。現在は、主に語彙情報を集積するが、今後はこれに北京大学「現代漢語短語結構規則庫」(The Phrase Structure Data-Base of Contemporary Chinese)の短語(詞組)結合関係データベースを加工して追加し、フレーズ情報を追加してゆく²⁷。語法情報の付加により、語彙・語法分布や、前後の単語連鎖の分析が可能となり、L2教学研究データとしての価値を高めることができると信じる²⁸。これら新規開発については、次回以降紹介してゆく。

〔追記〕本稿は、2002年4月—2004年3月科学研究費補助金研究「中国語チュートリアルシステムの国際共同開発」(基盤研究(B)(2)課題番号：14310136

研究代表者：砂岡和子)、および2001年度早稲田大学特定課題研究(国際共同研究)「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—」(国際(一般)課題番号2003C-002代表：砂岡和子)の研究成果の一部である²⁹。

²⁶ ただし機種限定。劉松、砂岡和子、浦野義頼2005年3月「PC、携帯端末に対応するコールドリルシステムの開発」

²⁷ 朱学峰、張化瑞、段慧明、俞涎汶2004年8月『漢語高頻詞語法信息詞典』的研制

²⁸ 和泉絵美、内元清貴、井佐原均2004年『日本人1200人の英語スピーキングコーパス』

²⁹ 劉松、砂岡和子、浦野義頼2005年3月「PC、携帯端末に対応するコールドリルシステムの開発」

〔参考文献〕

- 国家对外汉语教学领导小组办公室2002年1月「高等学校外国留学生汉语教学大纲」北京
語言文化大学出版社
- 国家汉语水平考试委员会办公室考试中心2001年6月「汉语水平词汇与汉字等级大纲（修
订本）」经济科学出版社
- 和泉絵美、内元清貴、井佐原均2004年10月『日本人1200人の英語スピーキングコーパス』
（株）アルク
- 尹景春1995年10月「漢語聽弁的特殊認識過程及其規則」『中国語学』242号、日本中国語
学会pp153-157
- 尹明、砂岡和子、成田誠之助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」（ICICE 2003）
- 史有為2004年1月「关于日本中国語教育初级词汇的统计研究和语篇研究」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回報告会資料
- 史有為2003年3月「中国語教育に於ける語彙の位置付け－中国語教育研究その4」明海
大学大学院応用言語学研究科紀要『応用言語学研究』No5
- 史有為・周剛2004年3月「日本第二外国語漢語教育之探討－以教学大綱和教材為中
心－」明海大学外国語学部論集第16集
- 朱学峰、張化瑞、段慧明、俞士汶2004年8月『漢語高頻詞語法信息詞典的研制』『語言
文字应用』第3期
- 砂岡和子、李利津、王玉、劉敬華2005年4月『Tutorial Chinese Basic』早稲田大学CCDL
研究所中国語部会発行
- 砂岡和子、李利津、王玉、劉敬華2004年4月『Tutorial漢語課本2004年版』東方書店
- 砂岡和子2004年4月「网络汉语口语課本和词汇分级」同上漢語コーパス研究会第4回《汉
語教学的核心词汇》報告会資料
- 砂岡和子・伊明・劉松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」
早稲田大学遠隔教育中国語部会漢語コーパス研究会第4回報告会
- 砂岡和子、村上公一2003年10月「以提升交際能力為目的之漢語教学以及水平測驗雜議
－早稲田大学國際漢語遠距教学四年實線見証」The 3rd International Conference on
Internet Chinese Education (ICICE 2003)『Dialogue between Technology and Chinese
Language & Culture』
- 砂岡和子、尹明2003年3月「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－」早稲田大学政経紀要『教養諸学研究』114号
- 砂岡和子2002年11月「当前日本漢語教学对言語信息处理技術的需求」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專門家委員会（CJNLP2002）Proceedings

88-95頁

- 砂岡和子2002年8月「早稲田大学漢語遠程教学的技術与理論」『E-Leraning与対外漢語教学』清華大学出版社(中国)436-442頁
- 砂岡和子2001年12月「語彙&語法ヘルプ機能付き中国語CALLドリル“ゴーツー(構詞)”」
- 張延瑞2004年1月「若干大学教材の初歩統計報告」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
- 張化瑞2004年1月「汉语词汇的頻度和通用度統計研究」同上
- 鄭国民・劉彩祥・王元華・陳双新2003年11月「小学校文常用讀物的字種与字量研究—“三、百、千”“四書”、古詩80首等六種讀物適用字」『語言文字应用』第4期
- 比企静雄・砂岡和子・楊立明・今泉一哉「合成音声利用による中国語声調の学習システム」2004 PC Conference 研究大会予稿集2004年8月
- 村上公一・砂岡和子・劉松2003年10月「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
- 俞士汶、段慧明、朱学鋒2002年12月「汉语语料库的建设及其在汉语教学中的应用」『早稲田大学CCDL中国語國際遠隔教育シンポジウム論文集』47-55頁好文出版
- 俞士汶、段慧明、朱学鋒、孙斌2002年2月「北京大学现代汉语语料库基本加工规范」《中文信息学报》第5期和第6期连载
- 俞士汶、朱学鋒、王惠、張芸芸1998年4月《现代汉语语法信息词典详解》清華大学出版社北京
- 劉松、砂岡和子、浦野義頼2005年3月「PC、携帯端末に対応するコールドリルシステムの開発」『ケータイ・カーナビの利用性と人間工学シンポジウム予稿集』31-32頁
- 劉松、砂岡和子、浦野義頼2004年8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」142-145頁2004 PC Conference 研究大会予稿集
- 劉松・張悦・倪萌・砂岡和子・村上公一・浦野義頼2004年7月「遠程網絡漢語口語教學管理平台的開發」『数字化對外漢語教學理論与方法研究』(清華大学出版社)