

中国語コーパスを活用した中級語彙3,000語の選定

小川 利康

キーワード：中国語教育、コーパス、語料庫、HSK、中国語検定、基本語彙、中国語常用語彙、現代漢語

【要旨】我々早稲田大学商学部中国語教室は、2000年10月から毎年「中国語Ⅰ基礎B」（1年次秋学期配当）履修者を対象に単語統一試験を実施してきた。この試験において、学生は「中国語基本単語帳」（750語）を完全に習得していることが求められている。もしこの試験に合格できなければ、彼らは「中国語Ⅰ基礎B」の単位を認定されることはない。この試験によって、学生たちの中国語の基礎学力は大いに向上している。

しかし、もう一方で、中国語中級クラス（2年次）における中国語の学力は十分に向上しているとは言い難い。これは主として専任教員が不足し、非常勤教員に依存せざるを得ない状況に起因するが、やはり学生たちの中国語力を向上させる適切な方法が存在しないことにも起因している。そこで我々は「中国語基本単語帳」を基礎としつつ、新たな「中国語語彙集」を提供することによって、2年次の学生の学力を向上させたい。

中国語中級クラスにおいては、最低でも日本中国語検定三級に合格できる学力を期待しており、もしくは旧HSK（漢語水平考試）6級合格程度を目指したい。なぜなら、この学力レベルに達すれば、学生たちは中国大陆（そして他の中国語圏）に留学できる十分な学力を備えたと思わせるからだ。しかし、現実にはHSKの提供する語彙を全て学習させることは出来ない。なぜなら、HSK語彙（甲乙丙丁級）は語彙数5,000を遙かに上回り、2年次の学生には負担が大きすぎ、また現実生活に必要な新しい単語、例えば携帯電話（手机）などが欠けているからである。

新しい語彙集を編むため、我々は北京大学が提供する中国語コーパスおよび2010年より始まった新HSKの語彙表を調査した。その結果に基づき、我々は1,500単語を満足の行く形で選定することに成功した。

はじめに

早稲田大学商学部中国語教室（檀山健介、宇野和夫、尹景春、中村みどり、小川利康）では『中国語基本単語帳』（語彙750語、DTP出版）を編み、同書に基づく統一単語試験を「中国語Ⅰ基礎B」（秋学期配当、右図参照）履修者全員を対象に実施し、その合格を単位認定の条件としている。基礎学力の担保を目的とした試みは概ね成功しており、学力格差の是正に役立っている。

だが、その一方で初級から中級へ移行する2年次で伸び悩む学生が目立つ。その解決策として、我々は「中国語基本単語帳」をベースとして、中級レベルまで対応可能な語彙集

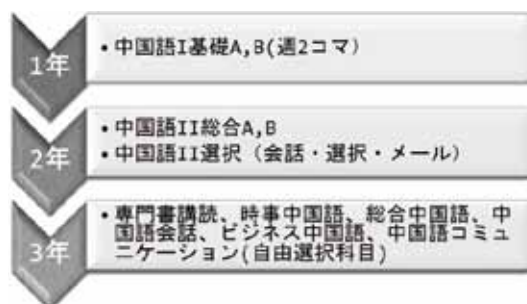


図1 カリキュラム概念図

を作成し、それに対応する統一試験を実施することで、解決の端緒を見いだしたいと考えた。だが、従来の語彙集にありがちな、経験論的な語彙選定ではなく、客観的なデータに基づく選定を行うために、中国語コーパスの統計データを利用する試みに挑戦することにして、教育総合研究所より2年間にわたりご支援を頂き、「中国語コーパスを活用した中級語彙の選定および教材開発」というテーマのもとに研究を進めてきた。本稿では、これまでの10年間の単語試験の実施状況の報告を踏まえつつ、中級レベルまで包摂する単語集の編制に係わる試行を報告する。

1. 基礎学力の担保のために

1.1. 中国語学習における語彙力

大学における学習者は、例外なく中国語（以下、中国における共通語である普通話を指す）入門学習において、その言語的特徴である声調を習得してから基本文型の学習に入る。その学習過程で声調を含む基礎発音の学習は、発音表記文字としてのピンインと表意文字たる簡体字（漢字）を結びつけるもので、その後の学習効率を決定づける大切な礎である。

一般的に週1～2回（90分）の授業では、5月までには発音学習を終えるが、現実的には模範朗読なしに正確な発音が出来なくなるまでになる学生は多くない。このため声調の安定化のため、文法学習を展開する一方で、継続的に発音矯正を行いつつ、オーラルテストを定期的に課してゆく必要がある。

その学習過程において、基礎語彙学習は発音練習（発音表記文字としてのピンイン習熟）と文法学習を結びつける要となる存在である。英語など欧米系言語学習では動詞など語形変化の学習に費やすが、中国語では常に漢字（簡体字）と発音の対応関係の把握に費やす必要があり、その語彙学習なくして基本文型習熟は成立しない。従って、初級文法に必要な基本動詞などを過不足なく収録した語彙集による学習は基礎学力を担保するうえで必要不可欠であり、早稲田大学商学部では1999年に初版を、2005年には現行の改訂版を刊行し、以来毎年10月末に1年次中国語履修者14クラスを対象に「単語統一試験」を実施してきた。

1.2. 単語統一試験概要

単語統一試験はマークシート問題50題（60点）および筆記（ヒアリング書き取り）10題（40点）で構成され、試験時間30分である。マークシート問題は5択問題だが、意味理解だけでなく、発音習熟（ピンイン表記理解）に重きを置いた内容となっており、以下の形式で毎年作題している。

1. 簡体字単語に対応する正しいピンイン選択
2. ピンイン単語に対応する正確な日訳選択
3. 日本語に対応する正確な簡体字単語選択
4. 声調が異なる単語（2音節）を選択
5. 文中空欄に当てはまる単語を選択
6. （筆記）読み上げられた単語の簡体字、ピンイン、日本語の意味を書き記す

この試験を通して試されるのは、発音表記としてのピンインを自ら正しく発音でき、音声を聞いて正しいピンイン表記で書き取れること、またピンイン表記を簡体字と正確に対応させられるかであり、秋学期始まって早々の10月末に実施されるため、実質的には半年間の勉学の成果が試されることになる。毎年500名前後が受験し、筆記試験部分は専任教員4名が即日採点し、マークは機械処理を行い、一週間後に結果を通知する。マーク及び筆記の合計が60点以上の者を合格とし、不合格者は各クラス担当教員のもとで合格するまで再テストを受け、最終的に合格できなかった者には秋学期の単位を認定しない。

1.3. 試験の実施と評価

これまで10年間試験を実施してきたなかで、効果を上げたのは、学生達が合格できるまで再テストを受けさせる指導を徹底した点にある。与えられた課題に自学自習で対応できる学生が減りつつある今日、重要なのは不合格になった学生をどのように導き、再テストで合格させるかにあると言っても過言ではない。

単語統一試験を実施する準備段階での議論では、基礎学力の担保のため、より総合的な学力を測る試験を実施するべきだとの意見もあったが、総合的な学力試験の場合、使用教科書を統一しなければ作題は不可能であり、教科書が統一されても、担当教員ごとに異なる教授法が行われていけば、統一的成績評価も容易ではない。このため「担当教員の個性、教授法を生かすためには教科書を統一すべきではない」（樋山健介）という考え方を基本に、まずは最低限の共通項として単語という基礎学力だけを測ることにした。わずか750語の単語帳を完全に習熟させる作業は愚直とも言えようが、これまでの十年間のデータを比較してみると、2000年開始当初は合格率が全クラス平均で50%を切っていたのに対して、近年は70%を確実に上回るようになり、基礎学力向上については確実な成果を上げている。それとともに平均点も確実に上昇しており、09年の例を示すと、図2^[1]のように概ね70点前後の平均点に達し、その後の再テスト実施によって最終的には合格率は90%以上に達している。

この指導プロセスにより、クラスの中下位層の底上げに顕著な効果を上げることが出来た。毎年年度末には、授業担当教員から毎年の授業内容総括を提出いただいているが、その内容からも2年次の授業で最初から落伍する学生が減ったことが明らかとなっている。

むろん単語試験の結果がクラスごとの総合的な中国語の学力とイコールとはいえない。だが、基礎学力としての重要な意味を持っていることは疑えない。我々は、今後もクラスごとの担当教員の個性や教授法を極力尊重する方針をすることを大前提としつつ、基礎学

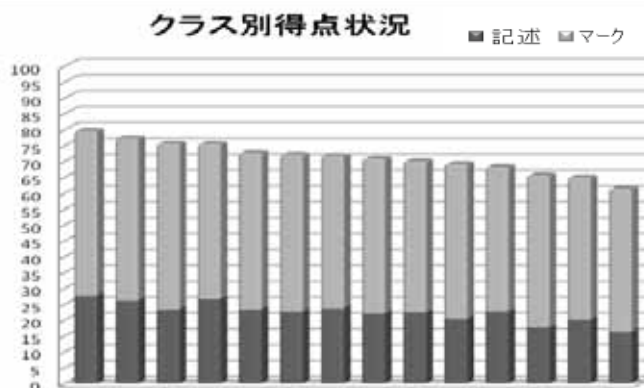


図2 09年度中国語単語統一試験得点状況

力を担保する方法論として、単語試験は有効な解の一つではないかと考えている。

またここで強調しておかねばならないことは、近年の合格率上昇は、学生の基礎学力向上を意味すると言うよりも、単語統一試験の定着とともに、クラスごとの担当教員が相当の熱意をもって学生指導を行った結果に外ならないという点である。これは自画自賛ではなく、他大学と同様、本学も中国語履修者急増のなかで専任教員増員が追いつかず、全14クラス中6クラスを非常勤講師に依存し、2年次に至っては全クラスを非常勤講師に依存している。従って、この試験の成果の過半は多々不利な状況下でも真摯に教育に取り組む非常勤講師の力に与ってのものであり、今後、更なる授業改善を進めるには担当教員の専任ポストを増やすことが喫緊の課題であると付言しておかねばならないだろう。

1.4. さまよえる中級クラス

1年次の基礎学力向上では一定の成果を上げていると評価できるものの、2年次で伸び悩むものが目立つ。その結果、3年次の中国語選択履修者には少数の中国留学経験者など高い学力を持つ学生がいる一方で、相当数の学生が「読めない、書けない、聞き取れない」という状態にある。彼らの中には2年次に力を伸ばすどころか、1年次に身につけた基礎学力すら維持できなかった者も見受けられる。学習のモチベーションを見失ってしまい惰性で学習を続けている点で、まさしく「さまよえる中級クラス」である。今後、あらゆる分野で日本と中国の連携が必要となるなかで、対等に中国と渡り合える人材の育成は急務である。そのためにも、2,3年次の授業でコアとなる教材を新たに編む必要がある。

2. 経験的語彙論からの脱却

2.1. 中級向けの「語彙集」編纂のために

現行の「中国語基本単語帳」は1年次秋に実施する試験に照準を合わせているため、語彙数は750と限られている。しかし、2年次も射程に入れた語彙集となると、語彙数も相当増やす必要がある。おおよその目安として、日中の代表的な中国語能力検定試験の語彙表^[2]を目安とすると、以下ようになる。

HSKの丙、丁は中級の範疇を超えるので、甲乙に限れば3,000語であり、大学2年（教養課程）修了程度という定義のある中国語検定三級が2,640語で、これに四級語彙を加えるなら、3,000語から3,740語が中級修了程度と見なされてきた。だが、日本の大学の中国語教育課程の現状を考えれば、この語彙数は余り現実的であったとは思えない。

そのうえHSKが正式に制定されたのは20年近く前の1992年で、中国語検定も1996年の改訂版から数えても十年以上経過している。統計的データを踏まえて制定されたHSKにしても、近年の中国のめまぐるしい変貌によって語彙は相当陳腐化してしまっている。中国語検定についても然りである。

HSK（漢語水平考試）甲乙	3,051語
HSK（漢語水平考試）丙	2,202語
HSK（漢語水平考試）丁	3,569語
中国語検定四級	約1,100語
中国語検定三級	約2,640語

図3 主要検定試験語彙数一覧

私たちが編んだ語彙集についても言うま

でもない。諸家の語彙集を参照したとはいえ、日頃の共学経験という極めて経験的な「カン」に拠って編まれたものであり、科学的な根拠に乏しく、1,000語以上の語彙集を編むための方法論としては薄弱すぎる。そこで現代中国社会での利用頻度の高い語彙を計量的に把握するための方法として、中国語コーパスを利用することにした。

2.2. 中国語コーパスの現況

大規模中国語コーパス（語料庫）として、現在良く存在が知られているのは、「北京大学CCL語料庫検索系統」(http://ccl.pku.edu.cn:8080/ccl_corpus/index.jsp)と「国家語言資源監測与研究中心」(<http://202.114.40.172:9090/cqs/>)であろう。このほかにも台湾の中央研究院、香港城市大学も語料庫を提供しているほか、日本の関西大学では中国語教科書の例文データをコーパスとして提供している。

このなかで先駆的な存在は北京大学であり、俞士汶教授を中心とする研究グループによって『現代漢語語法信息詞典』（清華大学出版社）が刊行されている。北京大学コーパスは現代中国語で利用される文例語彙を幅広く収集するために、その素材を新聞雑誌から文学作品に至るまで広く網羅し、テキストデータでおよそ3億字（2009年7月20日現在）を蓄積している。

中国語コーパスは基本的に動詞などを中心とする語法用例の収集が目的であり、語彙頻度に着目する研究は必ずしも多いとは言えなかったが、近年はコーパスデータに基づく『現代漢語常用詞表』（商務印書館）が2008年に、『中国語言生活狀況報告』（商務印書館）が2005年より毎年刊行されるようになった。こうしたデータも援用することによって、よりアップデートな語彙集を編むのが私たちの狙いである。



図4 北京大学CCL語料庫検索系統検索画面

2.3. 用例収集及び自動推薦システム (詹善斌)

以上の意図を実現するために、私たちはHSKの語彙集を検証するところから、作業を始めた。北京大学のコーパスはウェブ上で公開されている。本学の理工学術院の詹善斌氏（博士後期課程在学）の助力を得て、用例収集作業をPerlによって自動化し、HSK単語の用例数

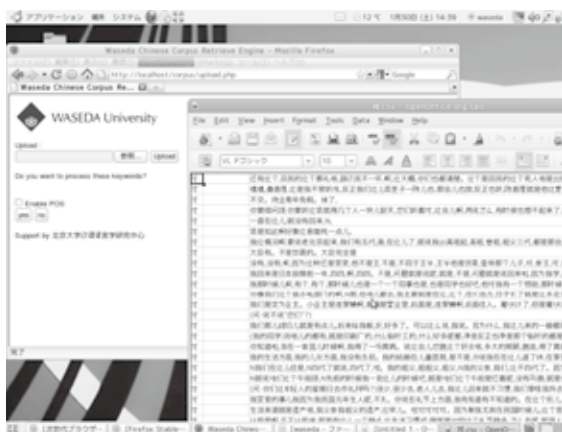


図5 Ubuntu910上での自動用例収集画面

を記録できるようにした。

当初は甲乙のみに限定する予定であったが、最終的には丙、丁も追加して調査したところ、丁の方にも用例頻度の高い単語が多数見つかり、選定語彙の陳腐化がデータからも裏付けられた。

さらに自動収集された用例は1,437万に上るため、その例文の検討に当たっては詹氏の助力により、自動品詞分析を行ったうえで適切な用例が上位に検索表示されるシステムを開発し、その用例検索システムに基づいて用例の検討を行っている^[3]。

2.4. HSK語彙にない新語の取り込み

以上の調査を通して、HSK単語を利用頻度順にソートし、上位語彙を選定することが出来たが、それだけでは新たに生まれた語彙を全て取り込みきれない。そこで、当初は現代中国社会での利用頻度に沿った語彙集として『現代漢語常用詞表』から頻用度の高い語彙を取り込む予定であったが、頻度の高い語彙にも学習者向けに適切とは言えないことが明らかになった。

そこで初級から中級者に提供する学習語彙として、一定程度安定的に利用される語彙が選別された語彙表として、『中国語言生活狀況報告』(2007年版、商務印書館)に収める「語文新課標教材3,000基本詞語表」(以下、語文基礎語彙)を利用することにした。

このデータは、「語文」教科書(日本「国語」に相当する科目)で学習する語彙(小学校1年次から中学三年次まで)について、中国国内で刊行される北京、上海の主要教科書四種類について集計調査し、利用頻度の高い語彙を3,000位まで示したものである。このデータを頻度順HSK及び中国語検定の単語と照合し、欠落部分を補った。言語動態を最もリアルに反映する北京大学コーパスと教科書という規範性の高いデータによる「語文基礎語彙」との双方を参照することで、データとしては恐らく十二分であると考えられる。



3. 新HSKの登場

初年度の研究において、ここまで検討を進めたところで、2010年春より中国では新HSKが実施された。当初、旧来と余り変わらないと仄聞していたが、現実には資料を入手してみると、かなりドラスティックに改訂されていることが明らかになった。何よりも大きな変化は受験者全員が同一問題を受験し、聴力、閲読、語法ごとの合計点に応じて級が判定されるスコア形式の試験であったのに対し、新たなHSKはCEF基準(ヨーロッパ言語共通参照枠組)に基づいた級別試験になった点である。つまり、学力レベルに応じた級を受験すればよくなり、特に初中級レベルの学習者の負担が大幅に軽減された。これまでは初級者も難易度の高い設問を大量に含む試験を長時間受験せねばならず、負担が大きかったが、これで受験者にとっては受けやすい試験になった。また級別に必要な語彙が細分化されて明示されているので、準備もしやすくなった。従来の甲乙だけでも3,000語に及ぶ語彙集は明らかに中級レベル以上の学習者でないと手に負えなかったが、

右図に示すように、最低級はわずか150語さえ学習すればよい。

このドラスティックな改革の背景について、その責任者として改革を担った張晋軍は難易度の高い旧来のHSKが外国人学習者にとっては大きな負担になっていたことを率直に認め、その改革が必要とされながら遅々として進まなかった

が、近年の国家的な海外進出（いわゆる「走出去」）のかけ声のなかで、HSK改革が急務となった経緯を紹介している¹⁴⁾。

その説明に拠れば、今回の改革のポイントは学習者のレベルに対応した難易度の適正化に尽きるが、改革の原則として「考教分離」（特定の教材に依拠した試験を作らず、考査と教育を分離する）の原則を維持しつつ、「考教結合」を目指したという。ここでいう結合とは試験合格を目指すことで学習が促進され、学習促進の結果として試験に合格するという好循環を指す。そのために試験の作題からも恣意性を排除し、内容に規範性を持たせることを目指し、初級レベルでは規範性よりもコミュニケーション側面での言語理解を重視したという。この結果、学習語彙も大幅に削減された。張晋軍は両者の違いを説明するために、次のような語彙を例示する。

旧HSK：愛、飯館、飯店、商店、天、喜欢、喜爱、星期、星期天

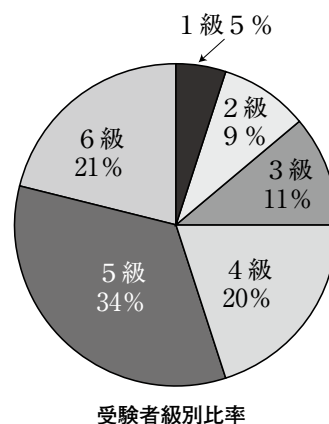
新HSK：愛、飯館、商店、天、喜欢、星期

つまり、張晋軍によれば「チョモランマのような仰ぎ見るような高山が実はよく見れば“水増し”が多く、単語としては多くても漢字として見れば採録しすぎ」であるという。確かに上記の語彙に含まれる「喜爱」「喜欢」（好きだ）は「喜欢」一つで兼ねても中級学習者には全く問題ないだろう。類義語は上級レベルで必要とはいえ学習語彙としては不要であり、漢字の字義で十分類推可能だというのが張の見解である。このほか聴解試験のウェイトが従来より高くなったこと、筆記試験の改善、初級レベルでの文字に拠らない図解写真を見て回答する設問の新設などHSKの作問内容も大きく変化しているが、ここでは語彙に係わる問題だけ見てきた。

以上の改革で生まれ変わった新HSKの各級レベルを日本の大学での標準的な中国語教育と対応させるなら、大学1年修了程度（週2回、90時間履修）ならHSK 3級を、2年修了程度（週2回、累計180時間履修）ならHSK 4級を目指すのが妥当だろう。単純に合格を狙うという意味では60%の得点で合格できるので、学力の高い学生なら更に一つ上の級を狙うことも十分可能であり、大学で2年間履修した学生ならHSK 3～5級までをターゲットとして考えられる。HSK

級 別	語彙数	旧HSKとの比較
HSK	5,000	HSK
HSK	2,500	HSK
HSK	1,200	HSK
HSK	600	上位級の語彙数は 下位級の語彙も含めた 数字である。
HSK	300 (Pinyin)	
HSK	150 (Pinyin)	

図6 新旧HSK比較表



5級でも必要とされる語彙数は2,500語だが、旧HSK甲乙3,000語より少なく、ハードルは確実に低くなっており、2年間で履修する単語数としても決して非現実的な数字ではないはずだ。旧HSKに代えて2010年から海外でも実施された新HSKの受験者数は初年度から2009年実施の旧来試験の受験者数(76,088人)を越えて、98,291人に達している。級別受験者比率は右の表^[5]に示したとおり、5級が中心を占めているが、今後は旧HSKでは全く相手にされなかった3,4級レベルの受験者が増えてゆくことが予想される。

3.1 新旧HSK語彙の比較

では、具体的に新HSK単語の内容はどのように変わったのだろうか。旧HSK単語と比べ、内容的にどこが違うのか。まず語彙数からして当然だが、旧HSK語彙との間で包摂関係を調べてみると、新HSK5,000で旧HSKに含まれていたのは37%(3,271単語)に過ぎなかった。旧HSKは甲乙丙丁(8,852語)と語彙数そのものが多いので単純に比較できないが、単なる削減ではなく、相当踏み込んだ取捨選択が行われたことを示しており、新HSKは旧版の改訂版ではなく、全面的改訂版という印象を受ける。

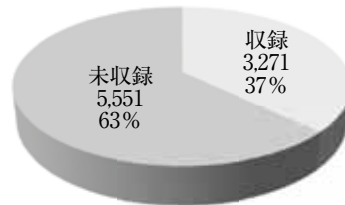


図7 新HSKの収録比率

次に旧版が新HSKにどれだけ単語が継承されたかを級別に分けて見てみると、以下のようなになる。ここで特徴的なのは低い級ほど単語の継承率が高いという特徴である。これは生活基本語が多数収録されている甲級については、余り変動の余地がなかった事を意味していると考えられる。1,033単語中719単語が一致し、継承率は69%に達する。続く乙級になると43%と半分以下になり、25%もダウンし、丙丁級はそれぞれ28%、29%更に半減しており、新旧HSKの方向性の違いが鮮明に現れている。すなわち、従来のHSKが単なる語学力に止まらず、発想レベルまで漢民族と過酷なまでに同化することを求めていたのに対し(華僑や中国少数民族

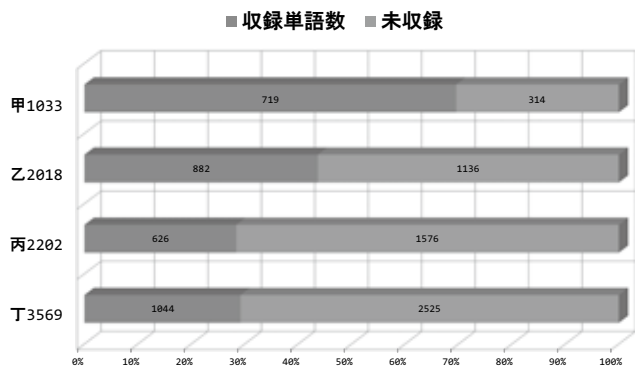


図8 甲乙丙丁級別収録状況

新HSK	旧HSK級別分布					旧HSK占有比率	新HSK単語数
	甲	乙	丙	丁	全体		
1級	134	5	3	0	142	94.67%	150
2級	127	13	3	0	143	95.33%	150
3級	195	55	9	3	262	87.33%	300
4級	175	249	32	17	473	78.83%	600
5級	79	458	183	73	793	61.00%	1,300
6級	9	102	396	951	1,458	58.32%	2,500

図9 新HSK級別占有状況

が多数受験していた)、新HSKでは外国人のコミュニケーションツールとしての中国語力を測る方向へと大きく舵を切った結果の反映である。このため新HSKでは口語能力の習得に重点を置くCEF基準に基づいて級別設定されており、例えば、1、2級は漢字を知らなくても受験できるように試験問題にはピンインが併記されている。その意図が明瞭に現れているのが新HSKに残っている甲級語彙の配置である。下の図に見るとおり、1級から4級までに傾斜的に配置されているのが分かる^[6]。旧HSK甲級語彙631語(甲級の63%)を4級までに学習させ、5級以降は中級レベルを目指すという意図が読み取れる。逆に興味深いのは丙丁級であった単語が1、2級に採録されているケースもある。これは6例と数が少ないので具体的に例を挙げてみよう。

电脑(パソコン:1級<=丙)、饭馆(高級レストラン、1級<=丙)、看(見る、1級<=丙)、进(入る、2級<=丙)、旅游(旅行する:2級<=丙)、生病(病気になる、2級<=丙)

動詞については、これほど平易な単語がなぜ丙丁に入っていたのか首をかしげざるを得ないが、名詞については時代の変化を反映するものと理解できる。これを3級までで丙丁に含まれていたものに対象を広げると、「菜单」(メニュー)、「打扫」(掃除する)、「地铁」(地下鉄)、「普通话」(標準語)、「照相机」(カメラ)、「周末」(週末)と、如上の傾向を明瞭に確認することができるだろう。週末という語彙も週休二日制が90年代半ばに定着してから使用頻度が上がった背景がある。

このように時代に対応した手直しがあったことは伺われるが、張晋軍が今後の改善点として級別語彙の配置の改善や新たに語彙を取捨選択する必要性を認めている^[7]ように、現代中国社会の言語実態を本当に反映しているかどうかは疑問が残る。その検証として、今回の研究プロジェクトで蓄積したデータを利用してみたい。

3.2. コーパスデータ新旧HSK語彙の比較

右の図は北京大学コーパスデータで用例数上位5,000までの語彙と「語文新課標教材3,000基本詞語表」(以下、語文基礎語彙)との比較である。詳細は上述の通りで、いずれも現代中国社会の言語実態を忠実に反映していると考えられるが、コーパスデータは用例採集対象が多岐にわたることから規範性が希薄であるのに対して、後者は国語の教科書(小中学校6年間)で高頻出語彙3,000語を採録しているため規範性が高い語彙が多い。性質の異なる頻度順データとつきあわせることにより、新HSK語彙がどの程度まで言語実態を反映しているかを判断する材料になると思われる^[8]。ただし、後者については3,000語に限られるため、5,000語の6級は検討対象には含めないものとする。

この表でも明瞭に読み取れる特徴は1~4級までの採録比率の高さである。4級までの1,200語につい

新HSK	単語数	北大コーパス 5,000		語文教科書 3,000	
1級	150	141	94.00%	131	87.33%
2級	150	143	95.33%	124	82.67%
3級	300	258	86.00%	223	74.33%
4級	600	455	75.83%	329	54.83%
5級	1,300	717	55.15%	370	28.46%
6級	2,500	1,072	42.88%	202	8.08%

高頻度語彙の採録率

ではコーパスデータの上位3,000語に絞り込んで採録比率を調べてみても、それぞれ1級が138語(92.0%)、2級130語(86.67%)、3級238語(79.33%)、4級406語(67.67%)と比較的高率を保っている。この状況は語文教科書上位3,000位についても同様で、4級までは比較的採録率が高いものの、5級になると28.46%まで落ち込んでしまう。

以上のデータだけで新HSK語彙の信頼性を判断することは難しいが、様々なデータから徴する限り、4級までの最も基本的なミニマムの語彙1,200語については、相当精選されていることは結論として確認できるだろう。ただ、その先の3,000語、さらには5,000語までのデータは新HSKに限らず、様々な頻度順語彙データの間でも齟齬があり、これは語彙採集対象の設定、あるいは語彙認定の方法(如上の「喜欢、喜爱」を「喜欢」に絞る問題、あるいは単用可能な単語をどこまで認めるか)などによって生じるズレによってもたらされた結果と考えられ、今後更なる検討が必要である。

3.3. 新HSKにおける文法リスト

ここまで新HSKについては語彙の問題を中心に論じてきたが、文法事項についても触れておきたい。従来のHSKは語彙のみで文法には全く言及がなかったが、新HSKでは級ごとに学習すべき文法項目を明示し、単語と同様に累加方式で学習項目が追加されてゆく。どんな中国語単語集を編む場合でも所謂虚辞類をどこまで収録するかは悩ましい問題である。虚辞類には代詞だけでなく、量詞、前置詞、疑問詞など文法的に重要な役割を果たす単語が含まれるが、意味だけ覚えても、文法学習には使えない。このため語彙集から省かれることもあるが、新HSKでは大まかとはいえ文法項目を掲げ、例文を示したのは一つの見識といえるだろう。文法項目も難易度を配慮して傾斜配置を行っており、例えば、前置詞では以下のように配置されている。頻度的にも高い「跟」(～とともに)が3級まで出てこない一方で、「向」(～へ)が先に出てくる選定方法には疑問がつかうが、他は常識的な配置といえる。一部「为了、除了、关于」は昨今の大学1年次の中国語では取り上げられない語彙も含まれるが、初級文法の項目として見れば無理とはいえない。

1級：在

2級：在, 从, 对, 比, 向, 离

3級：在, 从, 对, 比, 向, 离, 跟, 为, 为了, 除了, 把, 被, 关于

助動詞についても、動詞、助動詞として2つの用法がある「想」は省かれ、3級で「应该」がようやく出てくるなど問題点は指摘できるが、その他の傾斜の仕方は常識的と言えるだろう。

1級：会, 能

2級：会, 能, 可以, 要, 可能

3級：会, 能, 可以, 要, 可能, 应该, 愿意, 敢

むしろ多くの中国語教員を失望させるのは、アスペクトの扱いが余りにもぞんざいなことである。3級でもアスペクトとして提示されているのは、以下の例文だけである。

- 1, 用“在…呢”表示动作正在进行 他们在吃饭呢。
- 2, 用“正在”表示动作正在进行 他们正在打篮球。
- 3, 用“了”“过”表示动作已经完成 他买了一斤苹果。
- 4, 用“要…了”表示动作(变化)将要发生 火车要开了。
- 5, 用“着”表示动作(状态)的持续 外面下着雨。

完了の「了」と経験の「过」を同一項目で説明して事たれりとする一方で、なぜか進行形を1、2で分けて説明するのは不可解である。4で将然相を挙げ、(変化)と指摘しておきながら、「了」が完了だけでなく、変化を示す助詞であると説明は抜けているなど、問題点だらけである。また、特殊句形として「被」「把」兼語文、存現文を挙げながら、その句形に必須の方向補語、結果補語などは全く言及がない。文法項目を示したのは画期的だが、今後の改善が大いに望まれるところであろう。

4. 結語に代えて

初級から中級レベルまでをカバーする新たな語彙集編製の構想から出発した一連の研究は、コーパスデータの頻度順用例などを各種データを参照し、更に新HSKの単語データも取り込むことで、中級レベルに相応しい語彙数はどれくらいか、また、どのような語彙を盛り込むべきか、ようやく結論に達しつつある。すなわち、

- 1, 旧HSKの甲乙級では3,000語が目安とされていたが、明らかに大学2年生までを主たる対象とする学習者には負担が大きすぎる。
- 2, 負担軽減への解として新HSKが示す語彙量は適切なものと言える。新HSK 4級を目安とすれば1,200語となり、2年間の学習で十分習得可能な語彙量である。
- 3, 新HSK 4級の1,200語が本当に必要十分かどうかは検討の余地がある。その不足を補うためには、補助的に頻度順語彙集を用いる必要がある。
- 4, 頻度順語彙集としては、北大コーパス上位2,000位以内、及び語文教科書3,000語以内の単語で上記1,200語から漏れている単語を取捨選択して追加すべきだろう。
- 5, 現実的要請として受験希望者が多い中国語検定試験3級単語についても一部採録するが、最低限にとどめる。
- 6, 現行1年生向け語彙集が700語前後であることから、2年次での学習語彙も800語を限度都とし、全体で1,500語とする。

残念ながら、まだ定稿を見るに至っていないが、上記の方針で初中級学習者向け語彙集1,500語を完成させ、ネット上で公開したいと考える。

注記

- 【1】個人情報保護への配慮から、表中では平均点順にデータをソートしたうえで、クラスが特定できないようにして表示した。
- 【2】HSKは同委員会が提供する語彙表による。中国語検定の単語数は上野恵司著『標準中国語辞典』に拠る。
- 【3】口頭報告：詹善斌、小川利康「検索エンジン（Lucene）による中国語用例抽出最適化—北京大学中国語語料庫を利用して」（「漢字文献情報処理研究会」第十二回大会、2009年12月20日於花園大学）
- 【4】新HSK語彙の編成過程については、その改革責任者となった張晋軍のブログ「张晋军的BLOG」（<http://blog.sina.com.cn/zhangjinjun>）に詳しい。「HSK（基礎）の語彙は3,000で、HSK（初中等）は5,000で、（双方合わせた）HSK（高等）は8,000となり、このような難易度は大多数を占める海外の初級中国語学習者にとっては手の届かない高みであり、HSK受験をすると強い挫折感を味わい、中国語を勉強する意欲をなくしてしまう」（「新汉语水平考试（HSK）诞生记」2011-03-23）と指摘している。
- 【5】「新汉语水平考试（HSK）海外实施报告」（2011-02-22、「张晋军的BLOG」）に示されたデータに基づき、小川が作成した。
- 【6】「新汉语水平考试（HSK）诞生记」（同上）及び「新汉语水平考试（HSK）研制报告」（2010-06-02）いずれも「张晋军的BLOG」（同上）による。なお、新旧HSKの語彙比較を試みた論考として吕禾「新HSK一、二、三级词汇大纲用字情况研究」（『黑龙江社会科学』2010年5期总122期）があり、「新HSK一级、二级、三级词汇用字和旧HSK汉字大纲的甲级字有很高的一致性。新HSK一级、二级词汇用字中的98%左右属于旧甲级字，新HSK三级词汇用字中的90%左右属于旧甲级字。」と指摘する。
- 【7】「新汉语水平考试（HSK）研制报告」（同上）
- 【8】北京大学コーパスデータの用例数調査対象は旧HSK語彙について実施したため、その語彙は甲乙丙丁級8,822語に限定されており、新HSK語彙の妥当性を測るデータとしては一定の限界があると言わざるを得ないが、参考に掲げた。