

# 中国語語彙能力テストの開発 —HSK 三級レベルの日本人中国語学習者のデータによる評価—<sup>1</sup>

張婧禕<sup>2</sup>

玉岡賀津雄<sup>3</sup>

勝川裕子<sup>4</sup>

DOI: 10.18999/stul.31.21

**要約:** 中国語の語彙能力を測定するためのテストの開発を目的として、HSK 三級レベルの日本人中国語学習者にテストを実施し、その結果および評価を報告した。この語彙テストは、名詞、動詞、形容詞および副詞の4つの下位分類から構成されている。HSK 三級に準ずる語彙表から各下位分類のテスト問題のターゲット語を12語ずつ抽出し、合計48語とした。これら48語を使用頻度の高低で統制し、四者択一形式の48問を設けた。この語彙テストを42名のHSK 三級相当の日本人中国語学習者を対象に実施した。その結果、得点は20点から45点まで分布し、平均は33.36点、標準偏差は5.69点であった。クロンバックの信頼度係数( $\alpha=0.80$ )は非常に高く、信頼性は高かった。さらに、テストの各項目について、項目応答理論に基づいて、項目困難度、項目実質選択肢数、項目弁別力の3つの指標を計算した。これらの指標からも、この語彙能力テストの信頼性が高く、実用的であると判断された。この中国語語彙能力テストが、今後、中国語教育において活用されることを期待する。

**キーワード:** 中国語語彙能力テスト 中国語教育 HSK 三級 日本人中国語学習者  
項目応答理論 信頼性 項目困難度 項目実質選択肢数 項目弁別力

---

<sup>1</sup> English title: Development of a Chinese lexical knowledge test: Evaluation by data from native Japanese speakers learning Chinese at the HSK3 level

<sup>2</sup> Jingyi Zhang, Nagoya University, Japan; E-mail: star\_zhangjy@yahoo.co.jp

<sup>3</sup> Katsuo Tamaoka, Nagoya University, Japan; E-mail: tamaoka@lang.nagoya-u.jp

<sup>4</sup> Yuko Katsukawa, Nagoya University, Japan; E-mail: yuko-k@lang.nagoya-u.ac.jp

## 1. 研究の目的

語学教育においては、語彙と文法が並行して教えられている。そのため、語学教育の教授・学習効果を検証する手段として、テストを用いて学習者の語彙と文法の能力が測定される。たとえば、中国語を第2外国語として学ぶ場合、既習の中国語の能力を測定するために、新漢語水平考試<sup>5</sup>(以下、HSK)と中国語検定の2種の資格試験がある。とりわけ新漢語水平考試は中国政府が認定する中国語の資格試験であり、多くの学習者に利用されている。中級レベルの新漢語水平考試三級(HSK 三級)を例として挙げると、この種の資格テストは、リスニング、リーディングおよびライティングの3つの部分から構成されている。これらの各部分の問題の形式は、HSK 公式ホームページ(<http://www.chinesetest.cn/userfiles/file/dagang/HSK3.pdf>)に載せられた『新漢語水平考試三級(HSK 三級)の綱領』(ホームページでは「新HSK 三級考試大綱」と記されている)で調べることができる。それによると、リーディングの全問題数の30問のうち、語彙に関する問題が10問ある。これらの10問の設問文の出題形式は、未完成の短文と短い会話文の2種に分けられている。そして、5つの問題に対して6つの語が選択肢として初めに提示されている。

未完成の短文の場合は、初めに“A 刻, B 双, C 音乐, D 其他, E 声音, F 干净”の6つの語が選択肢として提示されている。その後、“除了数学, 弟弟的( )成绩都不错。(数学を除き, 弟は( )の成績も悪くない。)”のような5つの未完成の短文が続く。それぞれの短文を完成するために、先頭で提示された6つの語の中から最も適当な語を重複しないように1つだけ選ぶ形式である。

また、短い会話文の場合は、まず“A 明白, B 角, C 被, D 爱好, E 办法, F 甜”のような6つの語が提示される。その後、“A: 这个问题你们准备什么时候解决?(A: あなたたちはこの問題をいつ解決するつもりですか。)”の問いに対して、“B: 我们正在想( ), 今天应该可以解决。B: 私たちはちょうど( )を考えているので、今日中には解決できるはずです。”というような形式の5つの未完成の短い会話文が続く。そして、それらの短い会話文を完成するために、先頭に提示された6つの語の中から最も適当なものを重複しないように1つだけ選択する形式である。

<sup>5</sup> 本研究の「新漢語水平考試」は、「漢語水平考試」から2009年以降に新しく開発されたテストを指す。なお、略称のHSKは、ピンイン表記のHanyu Shuiping Kaoshiに由来し、新旧に関係なくHSKと呼ばれる。

ライティングの全問題数 10 問のうち、語彙に関する問題は5問ある。これらの5問の設問文は、短文の中で漢字1文字が抜けたところがピンインで提示されており、ピンインで提示された部分を簡体字で正しく書く形式の問題である。たとえば、“没（ ）系，别难过，高兴点儿。”という設問文であれば、括弧の上にピンインで“guān”が提示されている。提示されたピンインを漢字(簡体字)で表記する問題形式である。

以上のように、リーディングとライティングの 40 問のうち、語彙の意味と書字から中国語の語彙能力を測定するのは 15 問(37.5%)でしかない。そのため、全問題数に占める語彙項目の割合からみれば、このテストでは、語彙能力を独立した1つの変数として測定していないと考えられる。そこで、本研究では、中級レベルの中国語の語彙教育における指導指針とするため、中級レベル(HSK 三級)に相当する語彙能力を測定するテストを開発して、日本人中国語学習者に実施することにした。そのデータを、項目応答理論(IRT: Item Response Theory)に基づいて開発されたプログラム(TDAP: Test Data Analysis Program)を用いて、本テストの信頼性と妥当性を検証した。また、最少のテスト問題数で、つまり最短の所要時間で、最大の信頼性を保証できれば、最も効率の良いテストであるといえる。この意味で、本テストが真に実用的であるかどうかも検討した。

## 2. 中国語語彙能力テストの作成手順

テストを作成するという目的からみると、テストは、クラス編成・プレースメントテスト、進度測定用テスト、到達度確認用テスト、熟達度測定用テストと総合的な診断テストの5つの種類に分類される(オルダーソン他, 2010; ブラウン, 1999)。また、テストの方法には、多肢選択、空欄補充、組み合わせ法、書き換え、短文解答、小論文などの多様な問題形式がある。本研究では、中国語学習者の語彙能力が HSK 三級に達しているかどうかを測定することを目的とするため、テストにおける内容の妥当性、測定の信頼性および実用性の3つの観点を考慮したうえで、四者択一の多肢選択の問題形式を用いて、熟達度テストを作成する。新漢語水平考試(HSK)の公式ページには、最低級の一級から最高級の六級までの合計5,000語が掲載された語彙リスト(公式ページでは「新汉语水平考試(HSK)词汇(2012年修订版)」と記されている)が公開されている([www.chinesetest.cn/godownload.do](http://www.chinesetest.cn/godownload.do))。

本研究は、中級レベル(HSK 三級)の学習者を実施対象とするため、この語彙リストの三級語彙(計 600 語)のみをテスト対象の語彙とした。また、この語彙テストで測定する対象語

彙について、品詞の観点から「名詞」「動詞」「形容詞」「副詞」という4つの下位カテゴリーを設定し、三級語彙の 600 語をこれらの4つの下位カテゴリーに基づいて分類した。さらに、分類した後の語に対して、「語料庫在线」という中国語コーパス([www.cncorpus.org](http://www.cncorpus.org))で公開されている語彙使用頻度(以下、使用頻度)一覧を利用して、使用頻度を検索した。なお、ホームページで公開されている語彙頻度表の名称は「現代汉语語料庫詞頻表」(<http://www.aihanyu.org/cncorpus/resources.aspx>)である。さらに、使用頻度の高低別でこれらの語を統制したうえで、各下位カテゴリーについて 12 語ずつ(高頻度6語, 低頻度6語)選んだ(合計 48 語)。

高使用頻度 24 語と低使用頻度 24 語は、別々の語彙群であるため、独立したサンプルの  $t$  検定を使って、両群の使用頻度の平均が有意に異なっているかどうかを検証した。その結果、高使用頻度 24 語 ( $M=18,565, SD=9,095$ ) と低使用頻度 24 語 ( $M=1,416, SD=535$ ) は有意に違っていた [ $t(46)=9.03, p<.001$ ]。また、2つの頻度(高使用頻度・低使用頻度)と4つの下位カテゴリー(名詞, 動詞, 形容詞, 副詞)の  $2 \times 4$  の二元配置の分散分析で使用頻度の違いを検証した。その結果、高低の使用頻度の主効果は有意であった [ $F(1,40)=70.94, p<.001, \eta_p^2=.64$ ]。一方、4つの品詞分類の主効果は有意ではなかった [ $F(3,40)=0.007, ns, \eta_p^2=.001$ ]。また、両変数の交互作用も有意ではなかった [ $F(3,40)=0.006, ns, \eta_p^2=.001$ ]。したがって、4つの語彙分類で使用頻度が適切に高低で統制されていた。

また、コーパスで検索した使用頻度は、基本的にはベキ分布に従う傾向があるので、正規分布に近づけるために底を2とする対数に変換した。対数変換した使用頻度を基に、独立したサンプルの  $t$  検定を行った。その結果、高使用頻度 24 語 ( $M=4.22, SD=0.20$ ) と低使用頻度 24 語 ( $M=3.11, SD=0.24$ ) は有意に違っていた [ $t(46)=17.64, p<.001$ ]。また、2つの頻度(高使用頻度・低使用頻度)と4つの下位カテゴリー(名詞, 動詞, 形容詞, 副詞)の  $2 \times 4$  の二元配置の分散分析で対数変換した使用頻度の違いを検証した。まず、高低の使用頻度の主効果は有意であった [ $F(1,40)=274.77, p<.001, \eta_p^2=.87$ ]。一方、4つの品詞分類の主効果は有意ではなかった [ $F(3,40)=0.08, ns, \eta_p^2=.006$ ]。また、両変数の交互作用も有意ではなかった [ $F(3,40)=0.12, ns, \eta_p^2=.009$ ]。以上のように、4つの語彙分類において、使用頻度が適切に高低で統制されていた。

これらの 48 語の対象語の品詞, 高低使用頻度, 品詞分類, 使用頻度, 錯乱肢は表1に示すとおりである。

表1 テスト対象語の高低使用頻度と品詞分類別の使用頻度

4つの 品詞分類	使用頻度 の高低	テスト 対象語	使用頻度	四者択一形式問題の錯乱肢				
				形態類似	意味類似	音韻類似		
名詞	低	比赛	1,146	比较	故事	笔袋		
		报纸	903	服装	新闻	包子		
		天气	845	大衣	季节	星期		
		朋友	1,930	明天	要求	碰头		
		机会	1,348	几何	作用	几位		
		会议	2,076	合计	影响	回忆		
	高	马	12,014	鸟	汽车	妈		
		家	45,991	宝	花	假		
		题	14,565	堤	路	体		
		国家	10,987	回家	世界	活下		
		信	11,572	言	字	行		
		门	14,813	问	床	梦		
		動詞	低	欢迎	857	观看	选择	安静
				喜欢	1,496	参观	参加	奇怪
同意	953			因果	解决	统一		
洗	1,670			汗	喝	急		
玩	1,935			现	穿	完		
明白	1,512			明年	完成	并不		
高	可以		17,444	可能	会	客气		
	做		13,912	故	放	作		
	听		10,786	斤	读	停		
	教		21,550	烤	授	交		
给	16,738	答	向	割				
开	29,338	干	站	改				
形容詞	低	累	2,105	思	阴	类		
		饱	1,018	跑	久	保		
		认真	1,363	人身	容易	忍者		
		热情	936	凉快	好吃	热气		
		健康	1,112	建筑	容易	想看		
		简单	1,961	筒子	聪明	健谈		
	高	近	9,565	斤	忙	亲		
		新	21,526	薪	慢	晴		
		热	9,860	熟	冷	人		
		难	10,546	准	贵	男		
		高	29,423	亮	低	告		
		好	33,058	舒	忙	靠		
		副詞	低	一共	161	一只	一边	有空
				多么	1,137	怎么	可能	都是
马上	1,231			天下	就是	目前		
几乎	1,742			几点	越	奇怪		
一直	2,035			一样	特别	一起		
非常	2,507			观赏	已经	赔偿		
高	最		19,092	量	太	虽		
	一定		7,829	一会儿	必须	异地		
	先		15,198	元	更	再		
	只		26,560	口	才	旧		
没	27,717	投	不	每				
才	15,485	寸	就	猜				

対象語に形態的・音韻的・意味的に類似する語彙を3つの錯乱肢として設けた。学習者に分かりやすい設問文を作成するために、専門分野の知識を要する語彙や、難しい表現や方言の表現などを避けた。錯乱肢の詳細は表1に示した。たとえば、名詞の“比赛(試合)”に対して、“比较(比較)”(形態的類似)、“故事(ストーリー)”(意味的類似)、“笔袋(ペンケース)”(音韻的類似)という3つの錯乱肢を用意し、ランダムに提示した。設問文の“明天我想请小赵去体育场看一场( )。”を完成するため、それらの4つの選択肢から、最も相応しいものを1つ選ばせる形式である(設問文の詳細は補記を参照)。

### 3. 中国語語彙能力テストの実施対象

本中国語語彙能力テストは、2016年12月に、日本の4年制大学で中国語を学んでおり、日本語を母語とする42名の大学生を対象に実施した。これらの42名は、最年長が21歳6ヶ月、最年少が19歳10ヶ月で、平均年齢は20歳5ヶ月、標準偏差は6ヶ月であった。これらの中国語学習者の学習期間は平均1年9カ月であった。これらの学習者は、第2外国語として中国語を履修しており、中国語を専門としておらず、1年間を通して週2回90分の授業を60回受けている。新漢語水平考試(HSK)がネットで公開した各級の学習時間の一覧表を参照すると(<https://www.hskj.jp/level/>)、本研究が調査対象とする中国語学習者は、テスト実施の当日で計算して、二級から三級レベルの間であると想定される。

### 4. 中国語語彙能力テストの結果と信頼性

本中国語語彙能力テストは、満点が48点である。4つの品詞別の下位分類の平均、標準偏差、最高および最低点およびクロンバックの信頼度係数は表2にまとめた通りである。日本人学習者42名に行った48問の語彙テストの全体の平均は33.36点、標準偏差は5.91点であった。最高点は45点、最低点は20点であった。そのうち、名詞については、12点満点のうち、平均が6.93点、標準偏差が2.12点、最高点は11点であり、最低点は3点であった。動詞の平均は10.38点、標準偏差が1.53点、最高点は12点であり、最低点は6点であった。副詞では、平均が5.98点、標準偏差が2.34点、最高点は11点であり、最低点は2点であった。形容詞の平均は10.07点、標準偏差が1.47点、最高点は12点であり、最低点は5点であった。

表2 テスト対象の48語の下位分類(品詞)ごとの記述統計と信頼度係数

下位分類	満点	平均	標準偏差	最高点	最低点	信頼度係数
名詞	12	6.93	2.12	11	3	-
動詞	12	10.38	1.53	12	6	-
副詞	12	5.98	2.34	11	2	-
形容詞	12	10.07	1.47	12	5	-
全体	48	33.36	5.91	45	20	$\alpha=0.80$

注：N=42.

その後、クロンバックの信頼度係数( $\alpha$ )を計算し、0.80 という非常に高い値を得た。クロンバックの信頼度係数とは、複数の質問項目で構成されるある概念を測定するためのテストについて、それらの項目間の内的整合性を調べることで、テストの測定精度を表す指標である。項目の測定誤差が小さいほど信頼度係数が大きくなる。係数は、0から1の間で変化し、1に近いほど信頼性が高くなる。0.80 以上であれば、信頼性が非常に高いと判定される。さらに、各項目を削除した場合のクロンバックの信頼度係数をみると、0.79 から 0.81 までの範囲内であった。つまり、測定内容の一貫性がよく保たれた信頼性の高いテストであることが示された。

## 5. 項目応答理論に基づくテスト評価

項目応答理論 (Item Response Theory) とは、項目の回答を基に、項目難易度、実質選択肢効果、項目弁別力と3つの指標の合計などの指数を計算し、テスト項目の特性および学習者の能力を同時に推定する理論である。古典的テスト理論と比べれば、学習者がテスト項目の信頼性を改善することに有効的な情報を提供し、学習者依存性・テスト依存性に拘束されず、学習者の能力とテスト項目の難易度の不変性を求められるという利点がある(大友, 1996; 大友・中村, 2002)。本研究では、この項目応答理論に基づいて開発した分析ソフト TDAP ver.2 (Test Data Analysis Program; 大友・中村 2002) を用いて、テストにおける各項目の項目難易度、実質選択肢効果(以下、実質選択肢数)、項目弁別力および3指標の標準適切度の合計をテスト評価の有効な指標として算出し、正答者数と共に、表3に品詞別に報告した。

表3 48 語の正答者数, 項目困難度, 実質選択肢数, 項目弁別力および適切度

ID	問題番号	品詞	正答	正答者数	DIFF	AENO	DISC	SATOT
1	39	名詞	比赛	15	0.357	3.773	0.609	1.774
2	6	名詞	报纸	18	0.429	1.980	0.293	1.275
3	9	名詞	天气	39	0.929	1.354	0.128	1.407
4	18	名詞	朋友	28	0.667	2.247	0.156	1.560
5	31	名詞	机会	38	0.905	1.512	0.228	1.498
6	5	名詞	会议	37	0.881	1.613	0.438	1.598
7	38	名詞	马	23	0.548	2.186	0.498	1.584
8	11	名詞	家	22	0.524	3.106	0.393	1.673
9	1	名詞	题	5	0.119	2.042	-0.098	0.964
10	15	名詞	国家	22	0.524	2.535	0.491	1.620
11	21	名詞	信	16	0.381	3.446	0.481	1.595
12	42	名詞	门	28	0.667	2.247	0.337	1.613
13	47	動詞	欢迎	41	0.976	1.119	0.197	1.334
14	19	動詞	喜欢	39	0.929	1.354	0.397	1.495
15	35	動詞	同意	33	0.786	2.055	0.062	1.552
16	28	動詞	洗	38	0.905	1.463	0.297	1.471
17	20	動詞	玩	24	0.571	2.958	0.605	1.926
18	14	動詞	明白	34	0.810	1.984	0.414	1.657
19	16	動詞	可以	37	0.881	1.529	0.186	1.428
20	25	動詞	做	38	0.905	1.463	0.311	1.477
21	13	動詞	听	40	0.952	1.252	0.128	1.399
22	2	動詞	教	40	0.952	1.211	0.320	1.375
23	8	動詞	给	41	0.976	1.119	-0.017	1.313
24	46	動詞	开	31	0.738	2.012	0.277	1.535
25	4	形容詞	认真	30	0.714	2.182	0.084	1.544
26	22	形容詞	饱	39	0.929	1.354	0.033	1.399
27	26	形容詞	简单	30	0.714	2.393	0.255	1.649
28	24	形容詞	热情	30	0.714	2.428	0.553	1.853
29	29	形容詞	累	25	0.595	2.554	0.216	1.598
30	12	形容詞	健康	39	0.929	1.354	0.033	1.399
31	17	形容詞	近	35	0.833	1.734	0.487	1.615
32	34	形容詞	新	42	1.000	1.000	0.000	0.885
33	36	形容詞	热	41	0.976	1.119	0.357	1.389
34	10	形容詞	难	39	0.929	1.293	0.270	1.363
35	43	形容詞	高	34	0.810	1.959	0.476	1.690
36	45	形容詞	好	39	0.929	1.293	0.080	1.325
37	37	副詞	一共	17	0.405	2.897	0.664	1.777
38	40	副詞	多么	14	0.333	3.681	0.406	1.526
39	33	副詞	马上	11	0.262	2.668	0.038	1.179
40	44	副詞	几乎	15	0.357	2.931	0.516	1.507
41	23	副詞	一直	20	0.476	3.548	0.342	1.661
42	32	副詞	非常	36	0.857	1.741	0.584	1.773
43	7	副詞	最	30	0.714	2.069	0.237	1.533
44	27	副詞	一定	12	0.286	2.848	0.629	1.570
45	30	副詞	先	30	0.714	2.365	0.129	1.612
46	41	副詞	只	22	0.524	2.557	0.270	1.502
47	48	副詞	没	35	0.833	1.680	0.344	1.493
48	3	副詞	才	9	0.214	2.012	0.147	1.037

注: DIFF=項目困難度, AENO=実質選択肢数, DISC =項目弁別力, SATOT=適切度.



## 5.1 各設問の項目困難度

項目困難度(item difficulty: DIFF)とは、テストに設けられた項目がどれくらい難しいのかを示す指標である。項目困難度は正答者数を学習者数で割る数値であるため、実質的には正答率と同じであり、0から1までの連続数値をとる。項目困難度の数値が0であれば、最も困難な項目であることを示し、1に近ければ近いほど、正答率が高くなり、項目難易度が低いと解釈される。また、大友・中村(2002)の最適項目困難度の計算式( $0.5+0.5*1/\text{選択肢数}$ )により、4つの選択肢をとるテストの場合、最適項目困難度は0.625となる。これを基準として、0.625よりも高ければ高いほど、簡単な問題、逆に0.625よりも低ければ低いほど難しい問題とされる。今回のテストの各項目困難度では、表3に報告した通り、0.625よりも値が小さいのは問題1の“題”(0.119)、問題3の“才”(0.214)、問題33の“马上”(0.262)、問題27の“一定”(0.286)と問題40の“多么”(0.333)などの17問があり、0.625よりも値が小さいのは問題34の“新”(1.000)、問題36の“热”(0.976)、問題8の“给”(0.976)、問題47の“欢迎”(0.976)と問題2の“教”(0.952)などの29問であった。0.625に極めて近いのは、問題18の“朋友”と問題42の“门”(0.667)の2問であった。このように、このテストは問題の難易度の配置に偏りがなく、HSK 三級の中国語の語彙を適切に測定できるといえよう。

## 5.2 各設問の実質選択肢数

実質選択肢数(actual equivalent number of options: AENO)は多肢選択形式(multiple choice format)の問題項目で、正答以外の魅力である錯乱肢があるか、準備した選択肢が実質的に機能したかを示す数値化の指標である(大友・中村, 2002)。1から選択肢の数までの連続的な数値をとり、本研究では四者択一の問題形式であるため、1から4までの値となる。実質選択肢数の数値が1であれば、全員が正答したことになる。4であれば、1つの正答に対して3つの錯乱肢を含む4つの選択肢がほぼ同数の回答者に選ばれたことを示す。基本的に、実質選択肢数は選択肢の分布状況を反映し、どれくらいの錯乱効果があるのかを数値化する。正答率が低くなると、項目困難度も0に近づき、正答以外の選択肢を選ぶ確率も高くなるので、実質選択肢数も高くなる傾向がある。

本研究では、全語彙テストの項目( $N=48$ )の項目困難度と実質選択肢数のピアソンの積率相関係数は、極めて高い逆相関を示した( $r=-.804, p<.001$ )。表3に示した通り、48問のうち、最も実質選択肢数が高かったのは問題39の“比赛”(3.773)で、最も低かったのは問

題 34 の“新” (1.000) であった。選択状況を詳しくみると、問題 39 の“比赛”では中国語学習者 42 名のなかで、15 名 (35.71%) が正答を選び、7 名 (16.67%) が錯乱肢の“故事”，7 名 (16.67%) が錯乱肢の“笔袋”，13 名 (30.95%) が錯乱肢の“比较”を選択しており、錯乱肢効果がみられる。それに対し、問題 34 の“新”では、42 名の学習者に錯乱肢の魅力が発揮できず、全員が正答を選んだ。このように、問題 34 の“新”を除き、他の項目の錯乱肢は機能していたことが分かる。

### 5.3 各設問の項目弁別力

項目弁別力 (item discrimination power: DISC) とは、学習者の能力をテストで弁別できるかどうかについての指標であり、0 から 1 までの数値をとるものである。項目弁別力を用いて、各項目に対して行ったテストの得点に基づいて分けた 3 群間で正答者数を検討するものである (大友・中村, 2002)。各項目に対しては、TDAP がテストの得点に基づいて設定した上位群の正答者数が下位群より多ければ多いほど項目弁別力の数値が高くなる。逆に、下位群の正答者数が上位群よりも多ければ、項目弁別力がマイナスの値をとる。つまり、この数値は項目における学習者の能力を弁別することを示す。大友・中村 (2002) の提案では、項目弁別力が 0.30 以下なら、能力を弁別する項目として不適切である。表に示したように、0.30 を満たしていないのは問題 1 の“題” (-0.098)、問題 8 の“给” (-0.017) などの 25 問があり、テスト問題の 5 割を占めた。しかし、今回、テストを実施した学習者のほぼ全員が HSK 三級に相当し、中国語の語彙能力がほぼ均等であるため、これらの学習者によるデータからは、テスト項目の弁別力が高くならなかったのではないかと考えられる。

### 5.4 項目分析結果の総合評価

どのテスト項目が最も有効であるかを示す指標として、項目困難度、実質選択肢数、項目弁別という 3 つの指標による標準適切度の合計 (standard appropriateness total: SATOT) を計算し、表 3 に報告した。SATOT 値が大きければ、大きいほど優れていると解釈される。表 3 に示したように、問題 20 “玩” (1.926) の SATOT 値が最も大きかったため、最も良好な項目となった。それに対し、問題 34 の“新” (0.885) の SATOT 値が最も小さかったため、最も不良な項目とされる。つまり、今回の学習者は、問題 34 の“新”に対して、全員正答を答え、天井効果があったと考えられる。

## 6. おわりに

本研究は、HSK 三級レベルの中国語語彙能力テストを開発し、その評価を報告した。HSK 三級に相当する日本人中国語学習者に対しテストを実施、回答に基づいて、テストの信頼性、有効性および実用性を報告した。結果と今後の検討は、以下のようにまとめられよう。

第1に、今回のテストの語彙項目は、HSK 三級に準じて抽出し、語彙使用頻度を厳密に統制した。日本人中国語学習者42名に実施した結果、信頼度係数は0.80となった。これは極めて高い値といえる。本テストは、テスト問題全体として測定の精度を高く保っており、信頼性が高いと評価できよう。

第2に、項目応答理論に基づいて開発された分析ソフトである TDAP ver.2 を使って、項目困難度、実質選択肢数と項目弁別力の3つの指標から本テストの問題の適切性および特徴を総合的に考察した。まず、今回実施した48問のうち、最適項目困難度の0.625をやや超えたのは29問(60.42%)であり、0.625に近かったのはわずかに2問(4.17%)であった。0.625以上を示した語彙は全体の64.59%で、学習者にとっては本テストはやや簡単であったかもしれない。しかし、48問の項目困難度の平均は0.694であり、最適項目困難度の0.625に非常に近い数字である。本テストの問題は、難易度に適度な分散があり、語彙能力を効果的に測定しているといえよう。

また、学習者のデータからは、“新”(100%)、“热”(97.6%)などの形容詞と“给”(97.6%)、“教”(95.2%)などの動詞の正答率が高いことが分かった。これらは、高い使用頻度の語彙であり、普通の教科書やカリキュラムに頻繁に出現するため、よく習得されていると思われる。このように、使用頻度による習得効果はテストの結果からもうかがえる。次に、実施選択肢数では、問題34の“新”には天井効果が見られ、実施選択肢数が1.000となり、3つの錯乱肢の魅力は全く発揮されていなかった。これについては、今後、より適切な項目と差し替え、テストを改定する必要がある。

大友・中村(2002)は、項目弁別力が0.30以上を良好な問題項目の基準としている。本テストでは、項目弁別力が0.30を超えない項目が23問もあった。しかし、今回の学習者では、中国語の学習歴や習得内容がほぼ同等であるため、本テストでは、弁別力に大きな違いが出なかったと思われる。今後の課題として、中国語学習者の学習歴で分けるなどして、本テストを実施する必要がある。

[参考文献]

- 大友賢二(1996)『項目応答理論入門: 言語テスト・データの新しい分析法』, 大修館書店.  
大友賢二(監修)・中村洋一(著)(2002)『テストで言語能力は測れるか—言語テストデータ  
分析入門—』, 桐原書店.  
オルダーソン, チャールズ・クラッフアム, キャロライン・ウォール, ダイアン(著)・渡辺良典  
(訳)(2010)『言語テストの作成と評価—あたらしい外国語教育のために—』, 春風社.  
ブラウン, J. D.(著)・和田稔(訳)(1999)『言語テストの基礎知識—正しい問題作成・評価の  
ために—』, 大修館書店.

[Web 参考サイト]

新漢語水平考試三級の綱領(《新 HSK 三级考试大纲》):

<http://www.chinesetest.cn/userfiles/file/dagang/HSK3.pdf>

HSK 各級の一覧表:

<https://www.hskj.jp/level/>

HSK 一級~HSK 六級の語彙リスト(《新汉语水平考试(HSK) 词汇(2012 年修订版)》):

<http://www.chinesetest.cn/godownload.do>

中国語コーパス(《语料库在线》):

[www.cncorpus.org](http://www.cncorpus.org)

《现代汉语语料库词频表》:

<http://www.aihanyu.org/cncorpus/resources.aspx>

[補記]

中国語一語彙能力テスト

この語彙テストは48問で構成されています。各文の( )を埋めて文を完成させるのに、最も適切なものをA, B, C, Dの中から1つ選んでください。

1. ( ) 我不会这道( )。  
A 路 B 堤 C 体 D 题
2. ( ) 这学期, 王老师( )我们汉语。  
A 交 B 教 C 授 D 烤
3. ( ) 我昨天晚上12点( )睡觉。  
A 就 B 猜 C 才 D 寸
4. ( ) 他是个( )的好学生。  
A 容易 B 忍者 C 人身 D 认真
5. ( ) 明天早上公司要开一个重要的( )。  
A 会议 B 回忆 C 影响 D 合计
6. ( ) 爷爷每天上午都去图书馆看( )。  
A 新闻 B 报纸 C 包子 D 服装
7. ( ) 这次, 他的数学成绩是班里( )好的。  
A 太 B 最 C 虽 D 量
8. ( ) 王老师( )了我一本汉语字典。  
A 向 B 给 C 答 D 割
9. ( ) 快看, 出太阳了! 今天的( )真不错!  
A 季节 B 星期 C 大衣 D 天气
10. ( ) 你觉得汉语会话( )吗?  
A 难 B 贵 C 男 D 准
11. ( ) 她一个人在国外上学时, 很想( )。  
A 假 B 花 C 家 D 宝
12. ( ) 我爷爷奶奶的身体都很( )。  
A 容易 B 健康 C 想看 D 建筑
13. ( ) 我在中国( )过两次音乐会。  
A 听 B 读 C 斤 D 停

14. ( ) 你 ( ) 这个字的意思吗?  
A 完成 B 明白 C 并不 D 明年
15. ( ) 你都去过哪些 ( ) 旅游?  
A 世界 B 国家 C 回家 D 活下
16. ( ) 请问, 我 ( ) 在这里抽烟吗?  
A 会 B 可以 C 客气 D 可能
17. ( ) 弟弟的学校离车站很 ( )。  
A 厅 B 近 C 亲 D 忙
18. ( ) 我想明天请 ( ) 吃晚饭。  
A 明天 B 要求 C 碰头 D 朋友
19. ( ) 我很 ( ) 边听音乐边写汉字。  
A 喜欢 B 奇怪 C 参加 D 参观
20. ( ) 今天我去邻居家 ( ) 游戏了。  
A 玩 B 穿 C 完 D 现
21. ( ) 放学后, 我要去邮局给中国的同学寄 ( )。  
A 信 B 言 C 字 D 行
22. ( ) 奶奶: “多吃点儿, 都是你平时爱吃的。”  
孙子: “奶奶, 我真的吃 ( ) 了。”  
A 久 B 饱 C 保 D 跑
23. ( ) 他 ( ) 在这家公司努力地工作着。  
A 一样 B 一起 C 特别 D 一直
24. ( ) 那里的人们很 ( ), 送了我们很多礼物。  
A 好吃 B 热气 C 热情 D 凉快
25. ( ) 妈妈正在家里 ( ) 晚饭呢。  
A 做 B 作 C 故 D 放
26. ( ) 这件事看上去很 ( ), 可是做起来很复杂。  
A 筒子 B 简单 C 健谈 D 聪明
27. ( ) 我想这次她 ( ) 不会和我再见面了。  
A 一定 B 异地 C 必须 D 一会儿
28. ( ) 吃完饭, 妈妈叫我 ( ) 碗。

- A 洗 B 急 C 喝 D 汗
29. ( ) 昨天我睡得很晚，上完课后，感觉很 ( )。
- A 类 B 累 C 思 D 阴
30. ( ) 外婆让我回家后 ( ) 吃饭，然后再写作业。
- A 更 B 再 C 元 D 先
31. ( ) 将来要是 ( )，我想去中国旅游。
- A 作用 B 机会 C 几何 D 几位
32. ( ) 我觉得这个电影 ( ) 有意思。
- A 非常 B 已经 C 赔偿 D 观赏
33. ( ) 请您等一下，我 ( ) 帮您查一下。
- A 就是 B 天下 C 马上 D 目前
34. ( ) 过生日时，妈妈给我买了一台 ( ) 电脑。
- A 慢 B 新 C 晴 D 薪
35. ( ) 刚才，爸爸 ( ) 暑假带我去北海道。
- A 统一 B 同意 C 因果 D 解决
36. ( ) 我：“ ( ) 死我了！”
- 小赵：“天气预报说今天最高温度 40 度呢！”
- A 热 B 冷 C 熟 D 人
37. ( ) 中午，我去食堂点了一碗饭和两盘菜， ( ) 花了 45 元。
- A 一只 B 一边 C 一共 D 有空
38. ( ) 我小时候的兴趣爱好是骑 ( )。
- A 马 B 汽车 C 鸟 D 妈
39. ( ) 明天我想请小赵去体育场看一场 ( )。
- A 比赛 B 故事 C 笔袋 D 比较
40. ( ) 我 ( ) 希望明年我能顺利毕业啊！
- A 可能 B 怎么 C 多么 D 都是
41. ( ) 每天早上我 ( ) 喝一杯咖啡。
- A 只 B 口 C 旧 D 才
42. ( ) 李奶奶每天早上 5 点出 ( ) 去买菜。
- A 问 B 门 C 梦 D 床

43. ( ) 他弟弟的身高没有我( )。  
A 低 B 亮 C 高 D 告
44. ( ) 他( )天天去公园练习汉语会话。  
A 越 B 几乎 C 几点 D 奇怪
45. ( ) 我：“小李，我通过这次 HSK 考试了！”  
小李：“真的太( )了！恭喜你！”  
A 好 B 忙 C 靠 D 舒
46. ( ) 老李每天下班后，( )车去学校接女儿。  
A 站 B 开 C 干 D 改
47. ( ) 开学了，让我们热烈( )新来的同学。  
A 欢迎 B 选择 C 观看 D 安静
48. ( ) 我昨天生病了，所以( )去学校学习。  
A 不 B 没 C 每 D 投



Development of a Chinese lexical knowledge test:

Evaluation by data from native Japanese speakers learning Chinese at the HSK3 level

Jingyi Zhang (*Assistant Professor, Graduate School of Humanities, Nagoya University, Japan*)

Email: star\_zhangjy@yahoo.co.jp

Katsuo Tamaoka (*Professor, Graduate School of Humanities, Nagoya University, Japan*)

Email: tamaoka@lang.nagoya-u.jp

Yuko Katsukawa (*Professor, Graduate School of Humanities, Nagoya University, Japan*)

Email: yuko-k@lang.nagoya-u.ac.jp

**Abstract:** The present study aims to develop a comprehensive Chinese lexical knowledge test for learners of Chinese and evaluate its reliability based on the data collected from native Japanese speakers learning Chinese at the HSK3 level of the Chinese language proficiency test. This Chinese lexical knowledge test used 48 target words selected from the HSK3 vocabulary list and consisted of four sub-categories: Namely, nouns, verbs, adjectives, and adverbs. Each category had 12 target words which were further divided into two groups of six based on word frequency. This test used a multiple-choice question style, selecting one correct answer out of four possible choices. The test was administered to 42 native Japanese speakers learning Chinese at the HSK3 level. The scores of the 42 learners (maximum of 48 points) ranged from 20 to 45 points with an average of 33.36 and a standard deviation of 5.69 points. The reliability index revealed a Cronbach's coefficient alpha ( $\alpha$ ) of 0.80, which indicated a very high reliability for the test. The current study also evaluated each item from the test using (1) item difficulty, (2) actual equivalent number of options, and (3) item discrimination power, which also supported the test as an excellent inventory for evaluating Chinese lexical knowledge of native Japanese speakers learning Chinese as a foreign language. In conclusion, this test was found to be a highly reliable measurement for Chinese lexical knowledge.

**Key words:** Chinese lexical knowledge test, Chinese language education, HSK3, native Japanese speakers learning Chinese, Item Response Theory, reliability, item difficulty, actual equivalent number of options, item discrimination power

