

# 日語外來語的語彙試題之開發

魏志珍

中華大學應用日語學系助理教授

## 摘要

本研究嘗試開發外來語的語彙測驗，欲透過此語彙測驗簡便且確實地測量出日語學習者的外來語之語彙能力，以達到能客觀地掌握日語學習者的外來語熟習度之目的。本研究製成的外來語語彙測驗總共 36 題，內容是依據難易度、使用頻度、詞性這 3 個觀點所架構而成。本研究以台灣日語學習者 335 位為對象進行本語彙測驗的施測，並針對測驗結果進行分析。結果顯示，本語彙測驗除了具備極高的信度之外，試題不僅難易度適中還具備高度的辨別力，且選項設計也適當。另外，本研究透過 T 檢定的檢測，證實本語彙測驗能確實的分辨出學習期間不同的日語學習者於外來語語彙能力上的差異。由上述檢測可證實，本研究製作的外來語語彙測驗不僅兼具信度與效度，且還具備辨別日語學習者的外來語熟習度之辨識力。日後，本語彙測驗可提供台灣日語教師在掌握日語學習者的外來語熟習度時之參考依據。

關鍵詞：外來語、語彙測驗、信賴性、日語學習者、項目分析

受理日期：2016.03.18

通過日期：2016.05.13

# **A preliminary study for developing a Japanese loanwords test**

Wei Chih-Chen

Assistant professor, Department of Applied Japanese,  
Chung Hua university

## **Abstract**

This study attempted to develop a vocabulary test for loanwords, aiming to measure Japanese language learners' vocabulary competence in a convenient and accurate way. The vocabulary test for loanwords designed by this study contains 36 items based on the structure composed of 3 perspectives: difficulty level, frequency of use, and part-of-speech. The research subjects of the vocabulary test designed by this study are 335 Japanese language learners in Taiwan. The data from the test were analyzed. According to the analysis result, the vocabulary test was highly reliability. The items of this test were not only of appropriate difficulty level but also of very high discriminating power. The options were also properly designed. In addition, this study proved through t-test that the vocabulary test could positively identify Japanese language learners' differences in vocabulary competence of loanwords during their learning period. Based on the analysis results mentioned above, the reliability and validity of the vocabulary test for loanwords developed by this study were both high. The test was also discriminatory in terms of Japanese language learners' familiarity with loanwords. In the future, this vocabulary test can be used by Japanese teachers in Taiwan as a tool to measure Japanese language learners' proficiency in loanwords.

Keywords: loanwords, vocabulary test, reliability, Japanese language Learners, item analysis

# 日本語における外来語の語彙テストの開発への試み

魏志珍

中華大学応用日本語学科助理教授

## 要旨

本研究は、日本語学習者の外来語の習熟度を客観的に把握できるよう、外来語の語彙力を簡便で的確に測定できる外来語の語彙テストの開発を試みた。難易度、使用頻度、品詞の3つの観点に基づいて構成した36問からなる語彙テストを作成し、台湾人日本語学習者335名に実施した。測定した結果について検証を行ったところ、本語彙テストは高い信頼性を有しているのみならず、項目難易度のバランスが良くかつ比較的高い弁別力を持っており、さらに設問の選択肢も概ね適切であることが明らかになった。また、本語彙テストは日本語学習期間による外来語の語彙力の差を有効に弁別ができていることが  $t$  検定による検証で確認できた。このように、本研究で作成した外来語の語彙テストは、信頼性と妥当性を兼ね備えており、かつ日本語学習者の外来語の習熟度に対する測定力を有していることが実証できた。今後、本語彙テストは台湾の日本語教育現場において外来語習熟度の測定への活用が期待できる。

キーワード：外来語、語彙テスト、信頼性、日本語学習者、項目分析

# 日本語における外来語の語彙テストの開発への試み

魏志珍

中華大学応用日本語学科助理教授

## 1. はじめに

日本語には、和語、漢語、外来語、混種語の4つの語種がある。そのうち、外来語に関しては、その使用は近年著しい増加を見せている。国立国語研究所（1964）が実施した語彙調査によると、1956年に出版された90種類の雑誌で用いられた語彙のうち、外来語が占める割合は9.8%であった。しかし、国立国語研究所（2005）が実施した調査では、1994年に出版された70種類の雑誌で用いられた外来語は、全語彙の約33.8%を占めるようになっている。このように、外来語の使用率は、38年間でおおよそ3倍にも増え、増加の一途をたどっている。外来語は、今後も日本社会の国際化に伴い、様々な分野において増え続けていくと予想される。このような情勢から、日本語母語話者だけでなく、日本語学習者も外来語に触れる機会が多くなると考えられる。それに伴い、日本語学習者にとって外来語の学習はますます重要になってきており、避けて通れないものになっていると言える。また、この現状を受けて、教える側である日本語教師は、日本語学習者に対して外来語の学習をサポートするために、彼らの外来語がどの程度まで習得されているのかを的確に把握することが重要である。つまり、日本語学習者の外来語の語彙力をどのように捉え、それをどのように測定するかが外来語の学習指導において課題の一つとなってきている。

そこで、本研究では、日本語における外来語の熟達度を客観的に把握するための指標となる外来語の語彙テストの開発を行った。

## 2. 言語テストの基本と研究目的

外国語教育では、学習者の言語能力を測定するために、これまで

に様々な言語テストが開発されている。テスト開発において最も基本でかつ重要なのは、「何を測るか」というテストの目的を明確にすることである。

言語テストは、テストの目的によって、大まかに適性テスト (aptitude test)、プレイスメントテスト (placement test)、到達度テスト (achievement test)、熟達度テスト (proficiency test) の4種類に分けられることが一般的である(石田 1992、中村 2002)。適性テストは、外国語学習に対する適性があるか否かやその学習に要する期間の長短などといった言語学習の潜在能力を測定することを目的としたテストであり、プレイスメントテストは、学習者をその言語能力に最も適したカリキュラムに振り分けることを目的で実施するテストである。そして、到達度テストは、学習者が一定の学習期間内に学習した内容をどの程度身につけたかを測定することを目的としたテストであり、熟達度テストとは、学習者の持っている言語能力がある時点でどのレベルに達しているかを測定することを目的としたテストである。これらの言語テストはそれぞれ異なる目的を持つものであるが、いずれもテストとしての条件を備えていなければならない。

石田 (1992) では、テストの具備すべき条件として、(1) 妥当性、(2) 信頼性、(3) 問題の包括性、(4) 実施容易性、(5) 採点の客観性、(6) 採点の容易性、(7) 解釈の容易性、(8) 利用の容易性、(9) 経済性、の9つを挙げているが、中でも特に(1) 妥当性と(2) 信頼性が評価の測定として最も重要な要素であると述べている。妥当性とは、テストが測定しようとしているものを的確に測れるかどうかのことを意味しており、すなわち測定道具としての適切さを表す指標である。信頼性とは、同じ対象者に対して誰が測定しても何回測定しても同じ結果が得られることを意味しており、すなわち測定道具としての測定結果の一貫性、安定性を表す指標である。テストは妥当性と信頼性を兼ね備えることが重要であるが、実施の所要時間が長すぎたり、採点が難しかったり、実施に膨大な費用がかかっ

では、実用性を欠くものになる。とりわけ、テストの所要時間が長くなると、受験者の体力に大きな負荷がかかることにより、疲労の度合いが増し、結果的にテストの妥当性と信頼性にも影響してくることになる。したがって、テスト開発においては、これらの条件間の関係を見極め、妥当性と信頼性が許容される範囲内でテストに課せられた目的を最大限に果たすべく、各条件の重み付けを勘案して調整していくことが重要である。

現在、日本語に関する代表的な公的試験としてよく知られている日本語能力試験（JLPT）、BJT ビジネス日本語能力テスト、日本留学試験（EJU）は、全てが熟達度テストに当たるものである。これらはいずれも妥当性と信頼性を兼ね備えたテストであるが、大規模な言語能力測定テストであるため、日本語教育の現場において簡単に実施することができないという実施容易性を欠く点がある。この点を解消するために、これまでも、いくつかの日本語学習者の日本語に関する言語能力を短時間で測定できる簡便な熟達度テストが開発されている。例えば、フォード丹羽・小林・山本（1995）は、日本語学習者の総括的な日本語能力を簡単かつ短時間で測定できるよう、日本語能力簡易試験 SPOT（Simple Performance-Oriented Test）を開発した。また、宮岡・玉岡・酒井（2011、2014）は、日本語学習者の語彙能力と文法能力を測定するために、設問をいくつかの要因で多面的に統制した日本語の語彙テストと文法能力テストをそれぞれ開発した<sup>1</sup>。これらのテストは、いずれも必要最小限度のテスト項目で妥当性と信頼性を満たす上に、学習者の言語能力を短時間で測定できるように作られている。日本語教育の現場での実用性が高く、さらに言語習得の調査や研究を行う際の言語能力測定用の付随テストとしても使えることから、これらのテストは既に複数の研究で使用され（岩崎 2002；魏 2010a、2010b；斉藤・玉岡・母 2012；初・玉岡 2013；大和・玉岡 2013 など）、日本語能力を測定する簡便なテ

---

<sup>1</sup> 宮岡他（2014）で開発された文法テストは、その後、従来の項目に改良を加えたものが、改訂版として発表されている（早川・玉岡 2015）。

ストとして一定の評価が得られている。しかしながら、現在に至るまで、日本語学習者の外来語に関する語彙力を測定するための実施容易性の高いテストは、未だ開発されていない。

上記のことを踏まえ、本研究では、日本語学習者の外来語の語彙力を簡便で的確に測定でき、さらに妥当性と信頼性が保証された実用性のある外来語の語彙テストを開発することを目的とする。

### 3. テスト項目の構成

#### 3.1 設問の構成

本研究では、日本語学習者の外来語の語彙力を測定する熟達度テストとして、高い弁別力を実現できる外来語の語彙テストを作成するために、設問に使う語彙を複数の観点から統制する必要があると考えた。宮岡他（2011）によれば、設問には下位分類が設けられることによって、テストの弁別力を高めることが可能となり、さらに結果の分析を多方面から分析することも可能となるという。

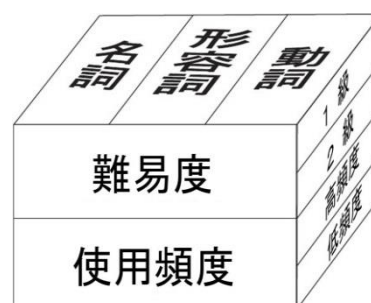
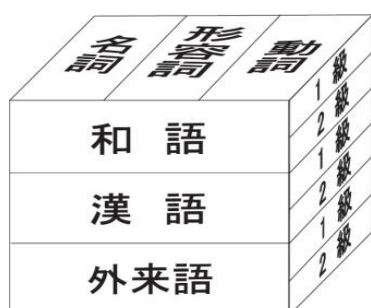


図1 宮岡他（2011）の設問構成モデル      図2 本語彙テストの設問構成モデル

宮岡他（2011）が開発した日本語の語彙テストでは、図1に示したモデルのように、語種について品詞と旧日本語能力試験配当級の双方から語彙を選定しており、語彙を語種（和語、漢語、外来語）、品詞（名詞、形容詞、動詞）、旧日本語能力試験配当級（1級、2級）の3つの観点から捉え、統制している。しかし、語種のうち、外来語は近年急速に増加している上、日本人が日常生活でよく使用している語が必ずしも日本語能力試験にも出ているとは限らない。その

ため、外来語に絞って語彙テストを作成する場合に、下位分類が宮岡他（2011）の枠組みそのままでは不十分であると考えられる。そこで、本研究で作成する外来語の語彙テストは、設問に使う語彙に対し、図2のように品詞、難易度、使用頻度の3つの下位分類を設けることにした。具体的には、品詞（名詞、形容詞、動詞）ごとに、難易度および使用頻度という異なる観点からそれぞれ語彙を選んだ。

まず、難易度においては、本研究では宮岡他（2011）に倣って、『日本語能力試験出題基準』（国際交流基金2002）に準拠し、1級と2級に割り当てられている外来語から品詞ごとに語彙を3語ずつ選定した。日本語能力試験は、2010年から改定された新試験を実施することとなったが、難易度の対応において旧試験の1級と2級が新試験のN1とN2と対応している<sup>2</sup>ため、難易度を判断するための参考基準として用いるのに適切であると言える<sup>3</sup>。

表1 本語彙テストにおける難易度別の抽出語彙一覧

	旧1級	旧2級
	ジャンル	スケジュール
名詞	キャリア*	トップ*
	サイズ*	ポスター*
形容詞	ナンセンス	スムーズ
	ドライ*	スマート*
	ルーズ*	モダン*
動詞	リード	カバー
	アップ*	ストップ*
	マッサージ*	スタート*

注：\*は宮岡他（2011）から援用した語彙のことを表す。

<sup>2</sup> 新試験のN1、N2、N4、N5は、それぞれ旧試験の1級、2級、3級、4級のレベルと対応しているが、新試験で新設されたN3は旧試験の2級と3級の間のレベルとなっている。

<sup>3</sup> 旧日本語能力試験配当級が難易度の判定基準として用いられた研究には、邱（2012）、大和・玉岡（2013）、早川・玉岡（2015）などがあり、さらに日本語読解学習支援システム「リーディング・チュウ太」(<http://language.tiu.ac.jp/>)も、それを基準に、文章に含まれる全ての語彙についての難易度判定を行っている。



表 1 は、難易度の観点から選定した語彙の一覧<sup>4</sup>であり、その中の一部は宮岡他（2011）で用いられた外来語に関する 1 級と 2 級の語彙を援用したものである。

次に使用頻度において、本研究では『毎日新聞』2000 年から 2010 年までの 11 年分の語彙の中から、旧日本語能力試験配当級の 1 級と 2 級に当たる外来語を全て除外した後、使用頻度が 1,500 以上の語を高頻度語、使用頻度が 800 未満の語を低頻度語として分類し、それらの中から品詞ごとに語彙を 3 語ずつ選定した。

表 2 本語彙テストにおける使用頻度別の抽出語彙一覧

	高頻度		低頻度	
	語彙	頻度数	語彙	頻度数
名詞	テーブル	2606	スキル	410
	コンサート	9324	コンプレックス	502
	コート	3100	サーカス	489
形容詞	ワイド	5514	ハード	460
	シンプル	2334	デリケート	243
	クール	1516	ゴージャス	221
動詞	アピール	13267	アタック	531
	リストラ	6470	オーダー	739
	チェック	10875	リフレッシュ	508
	平均	6112	平均	456
	標準偏差	4205	標準偏差	147

新聞はその国のその時点における縮図のようなものであり、かつ幅広い読者を持っているため、記事に用いられている語が一般に比べ保守的で、概して社会でもよく接し、ある程度知られている語である（中山 2001、関根 2003）。したがって、本研究では、使用頻度で外来語を抽出するためのデータベースとして、新聞コーパスを用いることが適切であると判断した。表 2 は、使用頻度の観点から品

<sup>4</sup> サイズ、ポスター、スケジュールなどの語彙は、日本語学習において比較的早い段階に提示されるものと思われるが、『日本語能力試験出題基準』（国際交流基金 2002）によれば、旧 1 級または旧 2 級に配当級されている語彙である。その配置が適切かどうかについては、本稿では議論しないこととする。

詞ごとに選定した語彙の一覧である。

以上のように、本研究では、語彙を統制するため、難易度と使用頻度の観点により、設問に使う語彙をそれぞれ旧日本語能力試験と新聞コーパスの2種類のデータベースから品詞(名詞・形容詞・動詞)ごとに3語ずつ均等に合計36語抽出し、1語につき1問の問題を割り当てた。これにより、本語彙テストの設問を品詞の観点から見ると、名詞が12問、形容詞が12問、動詞が12問の計36問となる。一方、難易度と使用頻度の観点から見ると、旧1級と旧2級が同数の9問ずつ、高頻度と低頻度が同数の9問ずつとなる。

### 3.2 設問文の作成

本研究では、テストの実施と採点の容易性を考慮し、設問の出題形式は文中の括弧内に入れる語として最適なものを4つの選択肢から1つ選ぶという四者択一の選択形式にし、前掲した表1と表2の抽出語彙を設問の正解として用いた。設問文の作成にあたって、テストの信頼性と妥当性を確保するため、特に以下の点に配慮した。

第一に、設問文に使用する語彙の難易度を統制することである。本研究では、表3に示したように、ターゲットとして設定した外来語以外の語彙に対してその難易度を調整した。

表3 設問文に使用する全語彙の難易度による分類

難易度	旧日本語能力試験の範囲内					合計 (%)
	1級	2級	3級	4級	級外	
初級前半	0	0	0	38	0	38 (25%)
初級後半	0	1	17	29	1	48 (32%)
中級前半	0	15	8	2	4	29 (19%)
中級後半	1	25	2	0	6	34 (23%)
上級前半	1	0	0	0	1	02 (01%)
合計	2 (1%)	41 (27%)	27 (18%)	69 (46%)	12 (8%)	151 (100%)

具体的には、設問文に使う語彙の難易度は、全体の約6割以上が

旧日本語能力試験配当級の3・4級に当たるものである。なお、これらの語彙は、日本語学習辞書支援グループ（2015）によって開発された「日本語教育語彙表 Ver2.62」（<http://jreadability.net/jev>）の語彙難易度指標に従えば、全体の約8割近くが中級前半までの範囲内に収まる。このように、提示される設問文の語彙は難易度においてほぼ中級レベルまでになるように統制した。

第二に、選択肢の品詞を統一することである。本語彙テストは、前述したように全ての設問において正答1つと錯乱肢3つの四肢選択で構成されている。錯乱肢の語彙は、36問全て異なる外来語を用い、各設問においてそれぞれの品詞が正答と同じ種類になるように統制した。例えば、「外はとても寒いので、（ ）を着て出かけたほうがいい」という設問に対して、「コート」が正答であり、錯乱肢として用いられたのは「スカート」「ワンピース」「スーツ」のように正答と同じく名詞の語彙である。

第三に、設問文と選択肢の言語形式を統一することである。本研究では、外来語の語彙力を測定するのが目的であるため、語彙知識以外の知識を問う項目をなるべく排除するよう、設問文と選択肢の言語形式を統制した。設問文に関しては、使用する文体は普通体、丁寧体、敬語などの織り交ぜや複雑な文末形式を避けて、普通体を主とした。一方、選択肢に関しては、名詞の場合は単語そのままを選択肢として提示するが、活用のある動詞と形容詞の場合は活用の部分を設問文の中で示し、語幹のみを選択肢として提示するように統制した。例えば、動詞の場合、「監督にはうまく選手を（ ）してほしい」という設問に対して、「リード（して）」が正解であるが、選択肢として提示されるのは動詞の語幹「リード」のみであり、「～して」といった活用の部分を設問文に示すことにした。また、形容詞の場合、「～な」といった活用の部分を設問文に示すことにした。例えば、「ナンセンス（な）」という語彙を問う「この物語は（ ）なところが面白い」という設問では、選択肢として提示されるのは語幹「ナンセンス」のみであった。

表 4 外来語の語彙テストの設問文と正解

下位分類	品詞	正解	問題番号	設問文	
難易度	旧1級	名詞	サイズ	30*	この靴は（ ）が小さすぎて、私には履けない。
		名詞	キャリア	20*	父は医者として40年の（ ）がある。
		名詞	ジャンル	13	本棚の本を（ ）ごとに分けて整理した。
		形容詞	ドライ	32*	姉は（ ）な性格だから、感情に左右されずにいつも合理的に物事を判断する。
		形容詞	ルーズ	12*	田中さんはいつも時間に（ ）だから、今日も待ち合わせの時間に遅れて来るだろう。
		形容詞	ナンセンス	17	この物語は（ ）なところが面白い。
		動詞	アップ	33*	最近、うちの会社は儲かっているので、今月から給料がわずかに（ ）した。
		動詞	マッサージ	16*	ひさしぶりに山登りしたら、後で足がだるくなったので、自分で足を（ ）した。
		動詞	リード	35	監督にはうまく選手を（ ）してほしい。
	旧2級	名詞	トップ	36*	水泳で2位だった選手が1位の選手を追い抜いて、（ ）に立った。
		名詞	ポスター	1*	図書館の掲示板に、館内飲食禁止を呼びかける（ ）が貼ってある。
		名詞	スケジュール	27	今週の（ ）はもういっぱい、他の予定は入れられない。
		形容詞	スマート	10*	太っている私と違って、妹はとても（ ）だから、体にぴったりした服がよく似合う。
		形容詞	モダン	8*	駅前にあるあの建物はとてもおしゃれで（ ）だ。
		形容詞	スムーズ	19	問題の（ ）な解決を目指し、関係者が集まって協議した。
		動詞	ストップ	29*	台風のため、電力の供給が一時的に（ ）した。
		動詞	スタート	6*	来週の水曜日から新しい新聞小説が（ ）した。
		動詞	カバー	34	職場の優しい同僚は、いつも私の失敗を（ ）してくれる。
使用頻度	高頻度	名詞	テーブル	9	（ ）の上に置いてあったケーキを半分食べた。
		名詞	コンサート	3	大学時代の親友はピアノの先生で、（ ）を開くそうだ。
		名詞	コート	24	外はとても寒いので、（ ）を着て出かけたほうがいい。
		形容詞	ワイド	26	新製品の特徴は、画面が（ ）になったことだ。
		形容詞	シンプル	21	母に機能が（ ）で使いやすい携帯電話を買ってあげた。
		形容詞	クール	2	（ ）な彼は普段からあまり笑わないし、感情も顔に出さない。
		動詞	アピール	15	昨日の面接は時間が短くて、自分をうまく（ ）できなかった。
		動詞	リストラ	5	父は32年間勤めていた会社を（ ）されてしまった。
		動詞	チェック	11	記入した内容に間違いがないか、もう一度（ ）してください。
	低頻度	名詞	スキル	28	英会話の（ ）を上げるために、大学の集中講義に参加した。
		名詞	コンプレックス	18	中学生の頃は、背が低いことに（ ）を持っていた。
		名詞	サーカス	7	子供の頃、両親に連れられて（ ）を見に行ったことがある。
		形容詞	ハード	4	勉強もアルバイトも忙しくて、今週は結構（ ）だった。
		形容詞	デリケート	14	これは（ ）な問題で、慎重に扱わなくてはならない。
		形容詞	ゴージャス	31	この店なら、（ ）な雰囲気です。食事が楽しめるよ。
		動詞	アタック	22	好きな人には勇気を出して（ ）してみたほうがいい。
		動詞	オーダー	25	自分の体型に合わせて、デパートでスーツを（ ）した。
		動詞	リフレッシュ	23	先週末は友達と一緒に運動して、気分を（ ）させた。

注：\*は宮岡他（2011）の設問文を基に修正して作成したことを表す。それ以外の設問文は筆者が作成したものである。

以上の基準で作成した外来語の語彙テストの設問文とその正解は、表 4 に示した。なお、作成する際、本語彙テストの旧 1 級と旧 2 級に関する設問の一部は、宮岡他（2011）の設問文を基に修正して使用した。また、作成した後、日本語教育を専門とする日本人研究者 3 名に設問文の内容および選択肢の適切さを確認してもらった。

## 4. テストの実施と分析

### 4.1 調査協力者

本研究は、2014 年 11 月～12 月に、台湾の北部にある 3 つの大学で日本語を専攻する 3 年生 125 名と 4 年生 210 名の合計 335 名（男性 90 名、女性 245 名）の日本語学習者に対して本語彙テストを実施した。彼らの平均年齢は 22 歳 1 カ月で、平均日本語学習歴は 3 年 9 カ月であった。また、調査時の日本語能力は、日本語能力試験 N1 合格が 55 名（3 年生 10 名、4 年生 45 名）、N2 合格は 154 名（3 年生 34 名、4 年生 120 名）、N3 合格は 24 名（3 年生 14 名、4 年生 10 名）、未受験が 102 名（3 年生 67 名、4 年生 35 名）であった。

### 4.2 テストの結果と信頼性係数

日本語学習者 335 名に対して本語彙テスト（36 点満点）を実施した結果、全体の得点の平均が 21.73 点（標準偏差は 5.63、最高点は 36 点、最低点は 6 点）であった。また、テスト項目が一貫した特徴を有しているかを検証するため、テスト全体の得点についてテストの内容的一貫性の指標であるクロンバックの  $\alpha$  信頼性係数を算出した。信頼性係数は、0.000 から 1.000 の数値となり、1.000 に近いほど信頼性の高いテストであることを示す。一般的には、0.700 を越えれば信頼性が保証されていると解釈される。本研究では、語彙テスト 36 問のクロンバックの  $\alpha$  信頼性係数は 0.816 となり、高い信頼性を示した。

### 4.3 テストの項目分析

本研究では、テストを構成する各設問が妥当であるかどうかを検討するために、大友・中村・秋山（2002）によって開発されたテストデータ分析プログラム TDAP(Test Data Analysis Program)Ver.2.0 を用いて項目分析を行った。項目分析では、評価指標として主に用いられるのは、項目困難度 (item difficulty: DIFF)、項目弁別力指数 (item discrimination power index: DISC)、実質選択肢数 (actual equivalent number of options: AENO) の3つである。また、この3つの指標から最適化された項目困難度適切度 (ADIF)、項目弁別力適切度 (ADIS)、実質選択肢数適切度 (AAEN) が導き出され、これらを基に標準化された指標である標準項目困難度適切度 (SADIF)、標準項目弁別力適切度 (SADIS)、標準実質選択肢数適切度 (SAAEN) が計算される。標準化された3つの指標の合計が標準適切度の合計 (standard appropriateness total: SATOT) となり、テストの総合的な適切さを示す指標とされている。

本研究では、項目分析で最も基本的な項目困難度 (DIFF)、項目弁別力指数 (DISC)、実質選択肢数 (AENO) の3つの指標による分析結果を検討し、さらにテスト項目の適切さを決める標準適切度 (SATOT) についても考察を行う。なお、この4つの指標による全体の結果は、付録の資料1に示した。

#### 4.3.1 項目困難度 (DIFF)

項目困難度は、テスト項目がどのくらい難しかったかを示す指標であり、「正答者数／受験者総数」で求められることから正答率とも呼ばれる。項目困難度は、0.000 から 1.000 の間の数値となり、1に近いほど易しく、0に近いほど難しい項目になる。項目困難度の判定基準については、中村（2002）によれば、一般的に正誤の確率は50%であるため0.500が適切だとされるが、多肢選択形式の場合は、主に選択肢の数によって決められる。本語彙テストのように選択肢が4つある場合は、当て推量によって偶然正答が選ばれる確率が4分の1であるため、それを加味して最適困難度は  $0.5 + 0.5$  (1

／選択肢数) で計算すると、0.625 となる。

表 5 は、項目困難度 (DIFF) が最も高い値を示した設問から降順で並べたものである。本語彙テストで最も難しかったのは順位 36 (No.32) の形容詞「ドライ」に関する旧 1 級の設問 (DIFF: 0.146) であった。一方、最も易しかったのは順位 1 (No.11) の動詞「チェック」における高頻度の設問 (DIFF: 0.976) であった。最適困難度に最も近かったのは順位 18 (No.23) の動詞「リフレッシュ」における低頻度の設問 (DIFF: 0.618) であった。また、最適困難度の 0.625 より値が大きく設問として簡単であったのは順位 1~17 までの 17 問で、0.625 より値が小さく設問として難問であったのは順位 19~36 までの 18 問であった。すなわち、36 問のうち、最適困難度に当たる設問を境にして、17 問が最適困難度より易しい設問、18 問が最適困難度より難しい設問であり、ほぼ半分ずつとなっている。このことから、本語彙テストは全体として難易度のバランスが良好であると言えよう。

## 台湾日語教育學報第26号

表 5 項目分析における項目困難度 (DIFF) の結果

順位	問題番号	正解	DIFF	順位	問題番号	正解	DIFF
1	11	チェック	0.976	19	15	アピール	0.582
2	9	テーブル	0.970	20	34	カバー	0.579
3	30	サイズ	0.949	21	5	リストラ	0.579
4	24	コート	0.946	22	35	リード	0.546
5	27	スケジュール	0.934	23	4	ハード	0.540
6	3	コンサート	0.904	24	12	ルーズ	0.499
7	16	マッサージ	0.875	25	19	スムーズ	0.466
8	29	ストップ	0.854	26	25	オーダー	0.415
9	6	スタート	0.827	27	26	ワイド	0.394
10	1	ポスター	0.818	28	22	アタック	0.367
11	2	クール	0.755	29	13	ジャンル	0.352
12	8	モダン	0.740	30	18	コンプレックス	0.352

13	36	トップ	0.693	31	14	デリケート	0.346
14	33	アップ	0.693	32	20	キャリア	0.287
15	21	シンプル	0.690	33	31	ゴージャス	0.284
16	28	スキル	0.672	34	17	ナンセンス	0.242
17	7	サーカス	0.633	35	10	スマート	0.212
18	23*	リフレッシュ	0.618	36	32	ドライ	0.146

注：\*は項目困難度（DIFF）の最適値に最も近かったことを表す

#### 4.3.2 項目弁別力指数（DISC）

項目弁別力指数とは、あるテスト項目が言語能力の高い上位者と低い下位者をどの程度正確に識別することができるかを表す指標である。項目弁別力指数は、点双列相関係数<sup>5</sup>によって得られた-1.000から+1.000までの範囲の数値となり、1に近いほど項目弁別力は高いと判断される。すなわち、この指標を用いると、言語能力の高い上位者が言語能力の低い下位者より多く正解したような項目の場合は弁別度が高く1に近い数値が、両者とも同じくらい正解したような項目の場合は弁別度が低く0に近い数値が、下位者が上位者より多く正解したような項目の場合は負の数値が算出される。また、項目弁別力の判定基準については、中村（2002）によれば、0.30以上であれば良いとする考え方が一般的となっている。したがって、本研究では、0.30以上の値は弁別力が良好な設問とみなすことにする。表6は、項目弁別力指数（DISC）が最も高い値を示した設問から降順で並べたものである。

表6を見て分かるように、全体においては、最も指数が高かったのは順位1（No.23）の動詞「リフレッシュ」における低頻度の設問（DISC：0.521）であり、最も指数が低かったのは順位36（No.32）の形容詞「ドライ」における旧1級の設問（DISC：0.120）であった。

<sup>5</sup> 点双列相関係数は、その項目において正解者のテスト得点の平均（ $\bar{X}_p$ ）と不正解者のテスト得点の平均（ $\bar{X}_q$ ）による差をテスト得点の標準偏差（SD）で割り、項目の正答率（p）と誤答率（q）の積の平方根を掛けて求められる数値である（中村2002）。計算式は  $DISC = (\bar{X}_p - \bar{X}_q) / (SD \sqrt{pq})$  となる。



また、36問のうち、項目弁別力指数が0.300を超えていた設問は順位1～27までの27問であり、0.300を満たしていない設問は順位28～36までの9問である。すなわち、全体の8割近くは項目弁別力指数が0.300以上あった設問である。このことから、本語彙テストは弁別力の高い設問が比較的多いテストであると言える。

表6 項目分析における項目弁別力指数(DISC)の結果

順位	問題番号	正解	DISC	順位	問題番号	正解	DISC
1	23*	リフレッシュ	0.521	19	1*	ポスター	0.360
2	36*	トップ	0.519	20	16*	マッサージ	0.357
3	21*	シンプル	0.504	21	31*	ゴージャス	0.348
4	2*	クール	0.504	22	5*	リストラ	0.346
5	34*	カバー	0.495	23	20*	キャリア	0.339
6	15*	アピール	0.494	24	33*	アップ	0.333
7	25*	オーダー	0.467	25	27*	スケジュール	0.320
8	28*	スキル	0.464	26	22*	アタック	0.317
9	4*	ハード	0.443	27	14*	デリケート	0.314
10	7*	サーカス	0.418	28	9	テーブル	0.294
11	19*	スムーズ	0.409	29	3	コンサート	0.284
12	35*	リード	0.404	30	18	コンプレックス	0.280
13	26*	ワイド	0.404	31	10	スマート	0.273
14	13*	ジャンル	0.391	32	11	チェック	0.267
15	24*	コート	0.391	33	6	スタート	0.263
16	30*	サイズ	0.390	34	17	ナンセンス	0.262
17	29*	ストップ	0.381	35	12	ルーズ	0.232
18	8*	モダン	0.370	36	32	ドライ	0.120

注：\*は項目弁別力指数(DISC)において問題の適切さを示す望ましい値を満たしていることを示す

#### 4.3.3 実質選択肢数(AENO)

実質選択肢数とは、多肢選択形式の問題項目で準備した選択肢が

実質的に機能したかを示す指標である。実質選択肢数は、平均情報量（エントロピー<sup>6</sup>） $H$ を求めた上で、実質選択肢数 $=2^H$ の式に当てはめて得られた0.000から選択肢の数までの数値となる。本研究は四者択一の問題形式であるため、実質選択肢数が0.000から4.000までの数値となる。実質選択肢数は、0に近いほど測定した項目に魅力的な錯乱肢がないこと、4に近いほど正答以外の錯乱肢が有効に働いたことを示すが、一般的に2.500を超えていれば概ね適切な錯乱肢であったとされる（中村 2002）。ただし、実質選択肢数は正答率に左右される数値でもある。設問は困難であるほど実質選択肢数が多くなり、容易であるほど実質選択肢数が少なくなる。それは、設問として困難である場合には、正答率が低く、正答以外の錯乱肢が選ばれる可能性が高いため実質選択肢数が多くなるからである。一方、設問として簡単である場合には、正答率が高く、正答以外の錯乱肢が選ばれる可能性が低いため実質選択肢数が少なくなる。表7は、実質選択肢数（AENO）が最も高い値を示した設問から降順で並べたものである。

表7 項目分析における実質選択肢数（AENO）の結果

順位	問題番号	正解	AENO	順位	問題番号	正解	AENO
1	17*	ナンセンス	3.976	19	7*	サーカス	2.858
2	13*	ジャンル	3.880	20	23*	リフレッシュ	2.722
3	31*	ゴージャス	3.870	21	28*	スキル	2.603
4	14*	デリケート	3.841	22	21*	シンプル	2.595
5	26*	ワイド	3.652	23	36*	トップ	2.558
6	22*	アタック	3.582	24	33*	アップ	2.530
7	18*	コンプレックス	3.551	25	8	モダン	2.266
8	19*	スムーズ	3.501	26	2	クール	2.209

<sup>6</sup> エントロピー（entropy）は情報量の尺度の一つであり、情報の曖昧さや乱雑度の増減を示す指標として用いられる（大友 1996）。テストでは、選ばれた選択肢の多様性を表すことができる。エントロピーは  $H = -\sum p_j \log_2 p_j$  の式で求められる。 $p_j$  はある項目における選択肢  $j$  を選んだ受験者の相対度数である。

9	10*	スマート	3.479	27	1	ポスター	1.764
10	20*	キャリア	3.394	28	29	ストップ	1.716
11	35*	リード	3.233	29	6	スタート	1.704
12	4*	ハード	3.118	30	16	マッサージ	1.626
13	15*	アピール	3.078	31	3	コンサート	1.502
14	34*	カバー	3.028	32	27	スケジュール	1.368
15	5*	リストラ	3.015	33	24	コート	1.306
16	12*	ルーズ	3.013	34	30	サイズ	1.291
17	32*	ドライ	2.940	35	9	テーブル	1.175
18	25*	オーダー	2.861	36	11	チェック	1.135

注：\*は実質選択肢数（AENO）において問題の適切さを示す望ましい値を満たしていることを示す

表 7 に示したように、全体において最も実質選択肢数が高かったのは順位 1（No. 17）の形容詞「ナンセンス」における旧 1 級の設問（AENO：3.976）であり、最も低かったのは順位 36（No. 11）の動詞「チェック」における高頻度の設問（AENO：1.135）であった。また、実質選択肢数が 2.5 を超えていた設問は順位 1～24 までの 24 問で、全体の 7 割近くを占めている。このことから、本語彙テストにおいて、選択肢は概ね適切であったと考えられる。一方、実質選択肢数が 2.0 以下の設問は順位 27～36 の 10 問である。これらを前掲した表 5 の項目困難度（DIFF）と照らし合わせてみると、全て項目困難度（DIFF）が 0.818 以上を超えていた簡単な設問と重なっていたことが分かった。つまり、これらの 10 問は正答率の高い簡単な設問であるため、実質選択肢数が低くなったと考えられる。

#### 4.3.4 項目困難度、項目弁別力、実質選択肢数の標準適切度の合計（SATOT）

項目困難度、項目弁別力、実質選択肢数の標準適切度の合計（SATOT）は、前述のように、これが項目困難度（DIFF）、項目弁別力指数（DISC）、実質選択肢数（AENO）の指標から最適化された上で、さらに標準化されたものの合計であるため、テスト項目の総合的な良さを決める

際の一つの指標とされている。ただし、SATOT の数値は包括的な値であるため、適切な数値となる場合であっても、それがある指標の低い数値と別の指標の高い数値との合計によるものである可能性があることから、個々の指標について詳細に検討する必要がある。表 8 は、SATOT の数値が高い順に並べたものである。

表 8 項目分析における標準適切度の合計 (SATOT) の結果

順位	問題番号	正解	SATOT	順位	問題番号	正解	SATOT
1	36*	トップ	1.880	19	29	ストップ	1.476
2	21*	シンプル	1.868	20	16	マッサージ	1.448
3	15*	アピール	1.854	21	27	スケジュール	1.447
4	34*	カバー	1.819	22	14	デリケート	1.443
5	23*	リフレッシュ	1.817	23	31	ゴージャス	1.403
6	7*	サーカス	1.771	24	3	コンサート	1.403
7	2*	クール	1.762	25	22	アタック	1.371
8	28*	スキル	1.758	26	9	テーブル	1.357
9	35*	リード	1.693	27	17	ナンセンス	1.341
10	4*	ハード	1.669	28	1	ポスター	1.335
11	19*	スムーズ	1.620	29	25	オーダー	1.320
12	5*	リストラ	1.614	30	12	ルーズ	1.316
13	33*	アップ	1.602	31	18	コンプレックス	1.308
14	8*	モダン	1.570	32	11	チェック	1.263
15	13	ジャンル	1.543	33	6	スタート	1.219
16	26	ワイド	1.533	34	20	キャリア	1.209
17	30	サイズ	1.505	35	10	スマート	1.133
18	24	コート	1.504	36	32	ドライ	0.828

注：\*は項目困難度 (DIFF)、項目弁別力指数 (DISC)、実質選択肢数 (AENO) の各指標において、問題の適切さを示す望ましい値を満たしていることを示す

表 8 に示したように、SATOT の順位で上位 14 位までのものは、項目困難度 (DIFF)、項目弁別力指数 (DISC)、実質選択肢数 (AENO) の各指標において、問題の適切さを示す望ましい値を満たしている

設問である。具体的に、これらは項目困難度の指標において全ての数値が最適困難度 0.625 に近い値で、項目弁別力指数においては、全ての数値が良好項目とされる 0.300 を超えており、かつ実質選択肢数においても、それぞれの数値が適切とされる 2.5 を超えた値とそれに近い値となっている<sup>7</sup>。このことから、上位 14 位までの 14 問はバランスが良い設問であることが明らかになった。

一方、SATOT の順位で 15 位以下のものでは、順位 24 (No. 3) の「コンサート」、順位 26 (No. 9) の「テーブル」、順位 32 (No. 11) の「チェック」、順位 33 (No. 6) の「スタート」の 4 問が項目困難度、項目弁別力指数、実質選択肢数の各指標<sup>8</sup>において、項目の適切さを示す望ましい値を 3 つとも満たしていなかった。これらの 4 問は、いずれも項目困難度が 0.80 を超えており、ほとんどの調査協力者が正答している非常に簡単な設問であった。難易度が低すぎる問題は、得点に天井効果をもたらし、実質選択肢数だけでなく、問題の弁別力も低くなっているのだと考えられる。これらの問題については、今後難易度を上げるよう、テストの改訂を行う必要がある。

#### 4.4 日本語学習歴によるテスト得点の検証

本研究で作成した外来語の語彙テストで日本語学習期間の異なる学習者の差を弁別することができるかどうかを検証するために、調査協力者の 3 年生 125 名と 4 年生 210 名のテスト得点について独立したサンプルの  $t$  検定で分析を行った。分析には、IBM SPSS Statistics Version 21.0 を使用した。その結果は、以下の表 9 に示した通りである。

本語彙テストは 36 問で構成され、各問 1 点の 36 点満点である。表 9 のように、テスト全体の平均は、4 年生が 22.67 点で、3 年生が

<sup>7</sup> 各設問に関する各指標の詳細は付録の資料 1 を参照。

<sup>8</sup> 各指標における値について、「コンサート」は DIFF:0.904、DISC:0.284、AENO:1.502 であり、「テーブル」は DIFF:0.970、DISC:0.294、AENO:1.175 であり、「チェック」は DIFF:0.976、DISC:0.267、AENO:1.135 であり、「スタート」は DIFF:0.827、DISC:0.263、AENO:1.704 である。

20.17 点であった。両者の平均得点を  $t$  検定で比較したところ、4 年生の方が 3 年生よりも有意に得点が高かったことが分かった [ $t(333)=4.01, p<.001$ ]。このことから、本語彙テストは調査協力者の外来語の語彙力が適切に測定できていることを確認した。

表 9 本語彙テストの平均 (M) と標準偏差 (SD) および  $t$  検定の結果

分類	配点	全体 (n=335)		3 年生 (n=125)		4 年生 (n=210)		$t$ 値	$p$
		M	SD	M	SD	M	SD		
全体	36	21.73	5.63	20.17	5.47	22.67	5.54	4.01	***
名詞	12	8.51	1.99	7.88	1.99	8.89	1.90	4.61	***
形容詞	12	5.31	2.31	4.99	2.32	5.50	2.30	1.97	<i>n. s.</i>
動詞	12	7.91	2.33	7.30	2.26	8.28	2.30	3.80	***
旧 1 級	9	4.59	1.59	4.27	1.53	4.78	1.61	2.83	**
旧 2 級	9	6.12	1.75	5.72	1.72	6.36	1.73	3.30	**
高頻度	9	6.80	1.65	6.41	1.61	7.03	1.63	3.39	**
低頻度	9	4.23	2.06	3.77	1.95	4.50	2.08	3.19	**

注:  $p<.05^*$ ,  $p<.01^{**}$ ,  $p<.001^{***}$ , *n. s.* not significant

一方、下位分類の観点からみると、品詞別に比べてみた場合、名詞と動詞においては、いずれも 4 年生の平均点が 3 年生より有意に高かったが [名詞:  $t(333)=4.61, p<.001$ 、動詞:  $t(333)=3.80, p<.001$ ]、形容詞においては、両者の平均点に差が見られたものの、有意な違いではなかったことが分かった [ $t(333)=1.97, p=0.05, n. s.$ ]。難易度別にみると、日本語能力試験の旧 1 級と旧 2 級の全てにおいて、4 年生の平均点が 3 年生より有意に高かった [旧 1 級:  $t(333)=2.83, p<.01$ 、旧 2 級:  $t(333)=3.30, p<.01$ ]。使用頻度別では、高頻度の問題も低頻度の問題のいずれも、3 年生よりも 4 年生のほうが有意に平均点が高かった [高頻度:  $t(333)=3.39, p<.01$ 、低頻度:  $t(333)=3.19, p<.01$ ]。

このように、 $t$  検定の結果から、4 年生と 3 年生の間に有意差が見

られなかったのは、形容詞のみであることが分かった。外来語では、数の上で圧倒的に多いのが名詞であり、比較的少ないのが形容詞であることが一般的に知られている。実際、これまでに語彙学習の目安にされることが多かった『日本語能力試験出題基準』（国際交流基金 2002）を調べてみると、1 級から 4 級までの外来語の語彙 533 語の内、名詞が 449 語で最も多く、動詞が 67 語で、形容詞が 17 語で最も少ない。したがって、上述した形容詞に関する結果から、学習者が 4 年生になっても既習の外来語の形容詞が少なかったことがうかがえる。

以上の検証から、本研究で作成した外来語の語彙テストは、日本語学習期間が異なる学習者の差を有効に弁別できていると言える。ただし、下位分類では、形容詞に関する設問は学習者のレベル差に対する弁別力が他の分類より弱かったため、さらなる改訂を行う必要がある。

## 5. まとめと今後の課題

本研究では、日本語学習者における外来語の語彙力を簡便で的確に測定できるよう、設問を難易度（旧日本語能力試験 1・2 級）、使用頻度（高・低）、品詞（名詞・形容詞・動詞）の観点から統制した外来語の語彙テストを開発した。作成した語彙テストの信頼性と妥当性を検証するため、台湾人日本語学習者に対して実施したテストの結果については、信頼性の検定、項目分析、 $t$  検定によって分析し、評価を行った。

まず、信頼性の検定においては、算出したクロンバックの  $\alpha$  信頼性係数から、本語彙テストは少ない項目数でありながら高い信頼性を有していることを確認した。次に、本語彙テストについて項目分析を行ったところ、項目困難度（DIFF）においては最適困難度の値より高い項目と低い項目がほぼ半々となっていることから、本語彙テストは全体として難易度のバランスが良好であることが分かった。項目弁別力指数（DISC）においては、8 割近くの項目が良好項目と

される値を満たしていることから、本語彙テストは比較的弁別力の高いテストであることが認められた。実質選択肢数（AENO）においては、全設問数の7割近くが適切とされる値を超えたことから、設問の選択肢は概ね適切であったと言える。また、前述した項目困難度（DIFF）、項目弁別力指数（DISC）、実質選択肢数（AENO）から導き出された標準適切度の合計（SATOT）においては、全設問数のうち、上位14位までのものは均衡の取れた設問であることが分かった。これらの結果を踏まえ、本研究では、日本語学習期間の異なった3年生と4年生のテスト全体の平均点についてt検定で分析したところ、両者の平均点による差が有意であることから、本語彙テストは、日本語学習期間による外来語の語彙力の差を有効に弁別できていることが確認できた。

以上のように、本研究で作成した外来語の語彙テストは、信頼性と妥当性を兼ね備えており、かつ測定力を有していることが検証によって実証できた。すなわち、本語彙テストは、日本語学習者の外来語習熟度を客観的に把握するための測定ツールとして妥当であることが示された<sup>9</sup>。また、本語彙テストは設問に使う語彙を多面的に統制したため、単なるレベル測定の用途のみならず、品詞や難易度や使用頻度による角度からの検証が可能であり、それにより日本語学習者の最も弱い部分がどこにあるかを詳細に把握できるといった点が最大の特徴である。今後は、台湾の日本語教育現場において外来語習熟度の測定への活用を実現するために、この語彙テストについてさらなる検証を繰り返しながら、適切でない項目の調整および改訂に取り組んでいきたい。

---

<sup>9</sup> 本語彙テストの得点でレベル分けを行う場合、次のような基準が考えられる。得点が「平均点+（2\*標準偏差）」より大きい値の場合は上位群の上、得点が「平均点+（2\*標準偏差）」～「平均点+標準偏差」の間にある値の場合は上位群の下、得点が「平均点+標準偏差」～平均点の間にある値の場合は中位群の上、得点が平均点～「平均点-標準偏差」の間にある値の場合は中位群の下、得点が「平均点-標準偏差」～「平均点-（2\*標準偏差）」の間にある値の場合は下位群の上、得点が「平均点-（2\*標準偏差）」より低い値の場合は下位群の下にあたる。



## 付記

本研究は 103 年度科技部專題研究「台灣日語學習者的日語外來語學習觀與其語彙理解之研究」(MOST 103-2410-H-216-002) による研究成果の一部である。また、調査を実施するにあたり、多大なるご協力をいただいた調査実施校の先生方ならびに学生の皆様に重ねて御礼を申し上げたい。

## 参考文献

- 石田敏子 (1992) 『入門 日本語テスト法』東京：大修館書店、pp. 1-233
- 岩崎典子 (2002) 「日本語能力簡易試験 (SPOT) の得点と ACTFL 口頭能力測定 (OPI) のレベルの関係について」『日本語教育』114、東京：日本語教育学会、pp. 100-105
- 大友賢二 (1996) 『項目応答理論入門』東京：大修館書店、pp. 1-313
- 魏志珍 (2010a) 「台湾人日本語学習者の事態描写における視点の表し方—日本語の熟達度との関連性—」『日本語教育』144、東京：日本語教育学会、pp. 133-144
- 魏志珍 (2010b) 「事態描写における台湾人日本語学習者と日本語母語話者の視点の比較—視座の置き方に注目して—」『言葉と文化』11、名古屋：名古屋大学大学院 国際言語文化研究科 日本言語文化専攻、pp. 255-270
- 邱學瑾 (2012) 「漢字圏日本語学習者における日本語単語の意味処理に及ぼす母語の影響—聴覚呈示の翻訳判断課題による検討—」『教育心理学研究』60 (1)、東京：日本教育心理学会、pp. 82-91
- 国際交流基金 (2002) 『日本語能力試験出題基準 改訂版』東京：凡人社、pp. 1-234
- 国立国語研究所 (1964) 『現代雑誌九十類の用語用例 第3分冊 分析』東京：秀英出版、pp. 1-337
- 国立国語研究所 (2005) 『現代雑誌の語彙調査—1994年発行70誌—』国立国語研究報告 121、東京：国立国語研究所、pp. 1-465
- 斉藤信浩・玉岡賀津雄・母育新 (2012) 「中国人日本語学習者の文章

- および文レベルの理解における語彙と文法能力の影響」『ことばの科学』25、名古屋：名古屋大学言語文化研究会、pp. 5-20
- 関根健一（2003）「新聞記事の中のカタカナ語」『日本語学』22（8）  
東京：明治書院、pp. 30-39
- 中村洋一（2002）大友賢二監修『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』、東京：桐原書店、pp. 1-204
- 中山恵利子（2001）「日本語教科書の外来語と新聞の外来語」『日本語教育』109、東京：日本語教育学会、pp. 90-99
- 初相娟・玉岡賀津雄（2013）「中国人日本語学習者による語彙・文法知識から述部構造の理解を仲介とした読解への因果関係モデルの検証」『ことばの科学』26、名古屋：名古屋大学言語文化研究会、pp. 5-24
- 早川杏子・玉岡賀津雄（2015）「改訂版・構造分類による日本語文法知識テストの開発—中国人日本語学習者のデータによるテスト評価—」『ことばの科学』29、名古屋：名古屋大学言語文化研究会、pp. 5-24
- フォード丹羽順子・小林典子・山本啓史（1995）「日本語能力簡易試験（SPOT）は何を測定しているか—音声テープ要因の解析」『日本語教育』86、東京：日本語教育学会、pp. 93-105
- 宮岡弥生・玉岡賀津雄・酒井弘（2011）「日本語語彙テストの開発と信頼性—中国語を母語とする日本語学習者のデータによるテスト評価—」『広島経済大学研究論集』34（1）、広島：広島経済大学経済学会、pp. 1-18
- 宮岡弥生・玉岡賀津雄・酒井弘（2014）「日本語の文法能力テストの開発と信頼性—日本語学習者のデータによるテスト評価」『広島経済大学研究論集』36（4）、広島：広島経済大学経済学会、pp. 33-46
- 大和祐子・玉岡賀津雄（2013）「中国人日本語学習者による外来語処理への英語レキシコンの影響」『レキシコンフォーラム』6、東京：ひつじ書房、pp. 229-267

## オンライン資料

日本語学習辞書支援グループ (2015) 「日本語教育語彙表 Ver2.62」

<http://jreadability.net/jev/> (2015年11月25日アクセス)

## テスト分析ソフト

大友賢二・中村洋一・秋山實 (2002) Test Data Analysis Program

(TDAP) Ver.2.0 [Windows 版] (中村洋一著・大友賢二監修『テ

ストで言語能力は測れるか』桐原書店、に添付された解析ソフト)

## 付録

資料1 テストについての項目分析の結果 (SATOTによる順位)

順位	標準適切 度の合計 (SATOT)	問題 番号	品詞	難易度 ／使用 頻度	正解	項目困 難度 (DIFF)	項目弁別 力指数 (DISC)	実質選 択肢数 (AENO)
1	1.880	36	名詞	旧2級	トップ	0.693	0.519	2.558
2	1.868	21	形容詞	高頻度	シンプル	0.690	0.504	2.595
3	1.854	15	動詞	高頻度	アピール	0.582	0.494	3.078
4	1.819	34	動詞	旧2級	カバー	0.579	0.495	3.028
5	1.817	23	動詞	低頻度	リフレッシュ	0.618	0.521	2.722
6	1.771	7	名詞	低頻度	サーカス	0.633	0.418	2.858
7	1.762	2	形容詞	高頻度	クール	0.755	0.504	2.209
8	1.758	28	名詞	低頻度	スキル	0.672	0.464	2.603
9	1.693	35	動詞	旧1級	リード	0.546	0.404	3.233
10	1.669	4	形容詞	低頻度	ハード	0.540	0.443	3.118
11	1.620	19	形容詞	旧2級	スムーズ	0.466	0.409	3.501
12	1.614	5	動詞	高頻度	リストラ	0.579	0.346	3.015
13	1.602	33	動詞	旧1級	アップ	0.693	0.333	2.530
14	1.570	8	形容詞	旧2級	モダン	0.740	0.370	2.266
15	1.543	13	名詞	旧1級	ジャンル	0.352	0.391	3.880
16	1.533	26	形容詞	高頻度	ワイド	0.394	0.404	3.652
17	1.505	30	名詞	旧1級	サイズ	0.949	0.390	1.291
18	1.504	24	名詞	高頻度	コート	0.946	0.391	1.306
19	1.476	29	動詞	旧2級	ストップ	0.854	0.381	1.716

20	1.448	16	動詞	旧1級	マッサージ	0.875	0.357	1.626
21	1.447	27	名詞	旧2級	スケジュール	0.934	0.320	1.368
22	1.443	14	形容詞	低頻度	デリケート	0.346	0.314	3.841
23	1.403	31	形容詞	低頻度	ゴージャス	0.284	0.348	3.870
24	1.403	3	名詞	高頻度	コンサート	0.904	0.284	1.502
25	1.371	22	動詞	低頻度	アタック	0.367	0.317	3.582
26	1.357	9	名詞	高頻度	テーブル	0.970	0.294	1.175
27	1.341	17	形容詞	旧1級	ナンセンス	0.242	0.262	3.976
28	1.335	1	名詞	旧2級	ポスター	0.818	0.360	1.764
29	1.320	25	動詞	低頻度	オーダー	0.415	0.467	2.861
30	1.316	12	形容詞	旧1級	ルーズ	0.499	0.232	3.013
31	1.308	18	名詞	低頻度	コンプレックス	0.352	0.280	3.551
32	1.263	11	動詞	高頻度	チェック	0.976	0.267	1.135
33	1.219	6	動詞	旧2級	スタート	0.827	0.263	1.704
34	1.209	20	名詞	旧1級	キャリア	0.287	0.339	3.394
35	1.133	10	形容詞	旧2級	スマート	0.212	0.273	3.479
36	0.828	32	形容詞	旧1級	ドライ	0.146	0.120	2.940

台湾日語教育學報第26号

資料2 外来語の語彙テスト

次の文の \_\_\_\_\_ に入れるのに最もよいものを、①・②・③・④から一つ選んで、( ) に番号を書いてください。

1. ( ) 図書館の掲示板に、館内飲食禁止を呼びかける \_\_\_\_\_ が貼ってある。  
①ポスト            ②ラベル            ③ライター            ④ポスター
2. ( ) \_\_\_\_\_ な彼は普段からあまり笑わないし、感情も顔に出さない。  
①フル            ②ドラマチック    ③クール            ④クリーン
3. ( ) 大学時代の親友はピアノの先生で、\_\_\_\_\_ を開くそうだ。  
①コンサート    ②カーテン            ③ページ            ④セレモニー
4. ( ) 勉強もアルバイトも忙しくて、今週は結構 \_\_\_\_\_ だった。  
①ハード            ②ダブル            ③ハードル            ④スリム
5. ( ) 父は32年間勤めていた会社を \_\_\_\_\_ されてしまった。  
①ストライキ    ②リストラ            ③ボイコット            ④ロック
6. ( ) 来週の水曜日から新しい新聞小説が \_\_\_\_\_ した。  
①スター            ②オープン            ③オーケー            ④スタート

7. ( ) 子供の頃、両親に連れられて\_\_\_を見に行ったことがある。  
①サーカス ②オンライン ③プライド ④ブーム
8. ( ) 駅前にあるあの建物はとてもおしゃれで\_\_\_だ。  
①モーター ②ハンサム ③モダン ④バランス
9. ( ) \_\_\_の上に置いてあったケーキを半分食べた。  
①スペース ②テーブル ③クラブ ④オフィス
10. ( ) 太っている私と違って、妹はとても\_\_\_だから、体にぴったりした服がよく似合う。  
①スタイル ②モニター ③スマート ④モデル
11. ( ) 記入した内容に間違いがないか、もう一度\_\_\_してください。  
①マッチ ②アクセス ③チェック ④スキャン
12. ( ) 田中さんはいつも時間に\_\_\_だから、今日も待ち合わせの時間に遅れて来るだろう。  
①ルール ②ユニーク ③フリー ④ルーズ
13. ( ) 本棚の本を\_\_\_ごとに分けて整理した。  
①フィルター ②ジャンル ③コンパクト ④ポジション
14. ( ) これは\_\_\_な問題で、慎重に扱わなくてはならない。  
①シャープ ②エリート ③スピーディー ④デリケート
15. ( ) 昨日の面接は時間が短くて、自分をうまく\_\_\_できなかった。  
①アプローチ ②デビュー ③アピール ④クリア
16. ( ) ひさしぶりに山登りしたら、後で足がだるくなったので、自分で足を\_\_\_した。  
①マッサージ ②チャレンジ ③コーチ ④マスター
17. ( ) この物語は\_\_\_なところが面白い。  
①プロセス ②アナウンス ③ナンセンス ④ニュアンス
18. ( ) 中学生の頃は、背が低いことに\_\_\_を持っていた。  
①パニック ②テンション ③コンプレックス ④プレッシャー
19. ( ) 問題の\_\_\_な解決を目指し、関係者が集まって協議した。  
①スムーズ ②ストレッチ ③スタミナ ④ワイルド
20. ( ) 父は医者として40年の\_\_\_がある。  
①キャプテン ②ベテラン ③プロ ④キャリア
21. ( ) 母に機能が\_\_\_で使いやすい携帯電話を買ってあげた。  
①リアル ②シンプル ③ダイナミック ④レギュラー

22. ( ) 好きな人には勇気を出して \_\_\_ してみたほうがいい。  
①アタック ②リンク ③プレー ④サポート
23. ( ) 先週末は友達と一緒に運動して、気分を \_\_\_ させた。  
①リクエスト ②セット ③アレンジ ④リフレッシュ
24. ( ) 外はとても寒いので、 \_\_\_ を着て出かけたほうがいい。  
①スカート ②ワンピース ③コート ④スーツ
25. ( ) 自分の体型に合わせて、デパートでスーツを \_\_\_ した。  
①タッチ ②オーダー ③リスト ④ショッピング
26. ( ) 新製品の特徴は、画面が \_\_\_ になったことだ。  
①グローバル ②ワイド ③コンパス ④デザイン
27. ( ) 今週の \_\_\_ はもういっぱい、他の予定は入れられない。  
①シーズン ②スケジュール ③ダイヤ ④シリーズ
28. ( ) 英会話の \_\_\_ を上げるために、大学の集中講義に参加した。  
①レッスン ②ニーズ ③ガイダンス ④スキル
29. ( ) 台風のため、電力の供給が一時的に \_\_\_ した。  
①ストップ ②コントロール ③コピー ④ドライブ
30. ( ) この靴は \_\_\_ が小さすぎて、私には履けない。  
①サイクル ②サイト ③サイズ ④カラー
31. ( ) この店なら、 \_\_\_ な雰囲気です。食事が楽しめるよ。  
①ダイレクト ②ゴージャス ③ストレート ④カラフル
32. ( ) 姉は \_\_\_ な性格だから、感情に左右されず、いつも合理的に物事を判断する。  
①ショック ②オートマチック ③シック ④ドライ
33. ( ) 最近、うちの会社は儲かっているから、今月から給料がわずかに \_\_\_ した。  
①キャッチ ②アップ ③カット ④オーバー
34. ( ) 職場の優しい同僚は、いつも私の失敗を \_\_\_ してくれる。  
①パス ②ガード ③カバー ④プラス
35. ( ) 監督にはうまく選手を \_\_\_ して欲しい。  
①ダウン ②トライ ③バック ④リード
36. ( ) 水泳で2位だった選手が1位の選手を追い抜いて、 \_\_\_ に立った。  
①トップ ②ナンバー ③レベル ④ワット