

日本語教育研究のための  
「かんたん日本語テスト」の開発  
— テスト開発経緯と項目分析結果を中心に —

The Development of the ‘Simple Japanese Test’ for the  
use in Japanese Language Education and Research  
— With a focus on the development process and item analysis results —

市江 愛\*・吉田 暁・石黒 圭  
(東京都立大学・城西大学・国立国語研究所)

Ai Ichie, Akira Yoshida and Kei Ishiguro  
(Tokyo Metropolitan University, Josai University and  
National Institute for Japanese Language and Linguistics)

Abstract

In this study, we developed the ‘Simple Japanese Test’ for the use in Japanese language education and research, which can be used stably in various environments, is easy to operate, and can be administered without placing too much burden on learners. It is a computer-based web test and has two types: ‘Vocabulary’ and ‘Grammar’. Each test is in a four-way multiple-choice format, with a total of 60 questions, divided into four levels. The time limit was 30 minutes for each. In developing these tests, a trial study was conducted with 80 and 72 questions in each test. Cronbach’s alpha coefficient, an indicator of internal consistency, was 0.945 for vocabulary and 0.951 for grammar, both of which were found to be highly reliable. Item analysis was also conducted, and the questions were refined based on item difficulty, item discrimination power index, choice rate and G-P analysis chart to prepare the test for practical use.

キーワード：日本語能力、客観テスト、語彙テスト、文法テスト

**Keywords:** Japanese proficiency, Objective test, Vocabulary test, Grammar test

## 1. 問題の所在と本研究の目的

日本語学習者に関する研究が世界中で増えているが、その際、客観的な指標に基づいて日本語能力を統制することが重要である。第二言語習得にはさまざまな要因が関係していると言われており、母語や年齢、学習歴、学習環境など複数の要因が絡み合っている。ある言語的特徴を説明するためには、そのほかの要因の統制が必要であり、日本語能力の統制も然りである。それは、協力者の日本語能力次第で、結果が大きく左右されてしまうからである。客観的指標に基づき、日本語能力を統制することが、日本語学習者を対象とする研究には重要であり、必要不可欠である。一方で、日本語教育研究のために、無料で、世界中どこでも、気軽に利用できるテストはほとんど存在せず、研究者・研究協力者双方にとって不便な環境である。そこで、本研究では、日本語教育研究のために利用しやすいテストを開発し、その信頼性を検証することを目的とする。

## 2. 先行研究からの課題

近年、日本語学習者の増加とその多様化を背景に、さまざまな客観テストが開発されてきた。大規模試験としては、日本語能力試験、日本留学試験、BJT ビジネス日本語能力テストなどが挙げられる。特に、日本語能力試験（Japanese Language Proficiency Test ; JLPT）は、独立行政法人 国際交流基金と財団法人 日本国際教育協会（現 公益財団法人 日本国際教育支援協会）が1984年に開始したテストで、2012年には64か国で実施され、受験者数も約57万人と、世界最大規模の日本語テストである（大隅・谷内 2015）。しかしながら、これらのテストを研究時の日本語レベルの客観的指標として使用するのには、非常に難しい。なぜなら、これらのテストは受験日が決められており、その受験日と研究の調査の実施日が非常に近ければ問題はないが、たいいていの場合にはそうではなく、受験日と調査日が離れてしまっているからである。そうすると、調査時点の日本語能力を測っているとはいえず、客観的に日本語能力を示すものとしては十分ではない。しかしながら、日本語能力試験は多くの学習者が受験するため、その利便性から日本語教育研究の客観的指標として使用している研究も少なくない。

このような大規模試験に対し、各目的や関心に合わせたさまざまな種類のテストも開発されてきた。たとえば、近年では、特定技能の制度化に伴い、主として就労のために来日する外国人が遭遇する生活場面でのコミュニケーションに必要な日本語能力を測定するために、国際交流基金日本語基礎テスト（Japan Foundation Test for Basic Japanese ; JFT-Basic）が登場した。

日本語教育研究でよく用いられる客観テストとしては、J-CAT や SPOT が挙げられる。J-CAT（Japanese Computerized Adaptive Test）は、項目応答理論に基づいた日本語のコンピュータ・アダプティブテストであり、受験者の能力に合致した難易度の問題を選んで出題される（今井 2015）。そのため、受験者の時間を短縮しつつ、テストの精度を高めることが可能になる。現在は一般社団法人 日本語教育支援協会が運営しており、有料となっている。研究費が潤沢にある場合

なら有料で精度の高いテストを用いることが可能であるが、そうではない場合、特に学生が研究で利用しようとするには難しいだろう。

SPOT (Simple Performance-Oriented Test) は、TTBJ という筑波日本語テスト集 (酒井・加納・小林 2015) の一つであり、自然な話速度の読み上げ分を聞きながら空欄となっているひらがな 1 文字を四つの選択肢の中から選ぶテストである。小林 (2015) によると、筑波大学で使用されていたプレースメントテストと SPOT の得点の相関も強く、言語能力を反映したテストとされる。SPOT は web 上で無料で受けられることから、多くの日本語教育機関でプレースメントテストとして用いられ、習得研究時の日本語レベルの客観的指標として活用されている。

J-CAT も SPOT も Computer Based Testing (以下、CBT) であるが、web を介したテストであればどのテストでも課題となってくるのが、安定して接続・動作できるかという点である。特に中国では、広く知られている通り通信制限があり、VPN などの特殊な環境を除けば、基本的には日本のサイトにアクセスすることができない。筆者らが 2021 年 12 月に中国各地で SPOT の接続状況を確認したところ、サイトにアクセスすることはできたが、SPOT が求める受験条件をクリアすることができず、多くの場所で受験が困難であった。学習者人口の多い中国語を L1 とする日本語学習者は、日本語教育研究の対象になることもしばしばであるが、中国国内で調査を行おうとした場合、動作が不安定なテストでは調査遂行に困難をきたしてしまう。

また、既存のテストは所要時間が 30 分以上であったり、多くの問題に回答しなくてはならないなど、調査協力者にある程度の負担がかかってしまう。調査の種類によっては、研究対象の調査自体は 10-15 分程度で終わるものもあるが、そのような場合、研究の主眼である調査よりも、テスト受験にかかる負担の方が大きくなってしまふ。そのため、研究のために日本語レベルを統制するという目的として使いやすいものとは言い難い。

このように、日本語にかかわるテストはさまざまな種類が開発され、外部に開かれたテストも多数ある。しかしながら、無料で、世界中どこであっても、気軽に、安定的に使用できるテストはほとんど存在しない。そのような背景からか、客観テストで調査対象者を統制するのではなく、所属機関の学年や、学習年数などで統制している研究も散見される。日本語教育研究のために、研究者が気軽に利用できるテストがあれば、より研究の幅が広げられるだろう。日本語教育学をさらに活性化させるためにも、無料で、多様な環境で安定的に使用でき、結果が簡単に集計でき、かつ学習者に負担をかけすぎないテストが必要である。

### 3. 「かんたん日本語テスト (Simple Japanese Test)」の開発

本研究では、日本語教育研究のための利便性を備えた学習者レベル判定用簡易日本語テストとして、「かんたん日本語テスト (Simple Japanese Test)」を開発した。テスト名の「かんたん」には、研究者も学習者も気軽に利用してほしいという想いを込めた。

本テストは、世界中の研究者が気軽に使用でき、かつ結果の集計が容易になるよう、CBT にした。

CBT の場合、通信負荷が大きいと、実施可否が国や場所などの実施環境によって左右されてしまいが、本研究ではサーバー環境を整えたことで、そのような問題を回避した。実際に、日本・韓国・中国・タイ・ベトナムで動作確認を行い、順調な作動を確認している。

テストは、「ごい」と「ぶんぼう」の2種類（以下、それぞれ語彙テスト、文法テスト）を作成し、四者択一式とした。また、さまざまなレベルの学生が受験できるように、どちらのテストも4レベルで問題を構成した（表1）。語彙テストでは、「日本語教育語彙表」をもとに、初級前半・後半、中級前半、中級後半、上級の4レベルとした。文法テストでは、国際交流基金（2002）『日本語能力試験 出題基準 [改訂版]』をもとに、旧日本語能力試験の4級、3級、2級、1級の4レベルとし、最終的に各レベル15問ずつ、計60問のテストとした。語彙テストは、語種（和語・漢語・混種語・外来語）と品詞（名詞・動詞・形容詞・副詞/オノマトペ）に分類し、作問した。文法テストは、日本語記述文法研究会編（2003-2010）を参考に、活用・文型、助詞、文末形式、テンス・アスペクト、複文（時間・状況、原因・理由、条件・目的、等位・並列、名詞修飾・引用）に分類し、作問した。なお、本テストの開発の発端は、書き言葉に関する縦断調査で使用できる簡便で実用的なテストがほしいというものであった。そのため、問題文は会話形式ではなく、一文完結の形式で提示した。

表記は、語彙力・文法力の測定を妨げないようにすることを基本とし、漢字表記については、語彙テスト・文法テストどちらにおいても、国際交流基金（2002）『日本語能力試験 出題基準 [改訂版]』の漢字級を基準とした。語彙テストでは、初級問題はひらがなと分かち書きとし、それより高いレベルの問題では分かち書きはなしとした。漢字表記については、中級と中上級の問題では2級、上級の問題では1級までの漢字はルビなしで出題した。なお、これらの基準を基本としつつも、漢語の表記や選択肢の表記の統一性を鑑みて作問した。文法テストでは、4級と3級の問題は分かち書きを用いた。漢字表記は当該級よりも高い漢字にはルビを付与し、当該漢字級に応じて、「学こう」のような混ぜ書きも使用した。

また、制限時間は各30分とした。問題数が各60問であるため、1問当たり30秒で答える必要がある。このように、よどみなく答えられなければ最後まで終わらないようにすることで、能力の差が反映できると考えた。前述したとおり、日本語教育研究のためのテストとしては、各30分という時間は長すぎるかもしれないが、開発の最初の段階では、テストの精度を高めることが最優先であると判断し、今回は各30分のテストとした。なお、出題順は各レベル順とし、解き進めるにつれ、やさしい問題から難しい問題となるよう設定した。

表1 かんたん日本語テストの構成と問題数

語彙テスト		文法テスト	
初級前半・後半	15問	4級	15問
中級前半	15問	3級	15問
中級後半	15問	2級	15問
上級	15問	1級	15問

## 4. 試行調査の手続き

語彙テスト・文法テストともに、最終的に各 60 問のテストにするために、それぞれ 80 問・72 問の問題を作成し、試行調査を実施した。その結果をもとに、各テストの信頼性の検証と、項目分析を踏まえた問題の精査を行う。なお、語彙テストのほうが作成問題数が多いのは、特に漢字語彙は母語による差が出やすく、余裕を持って検討できるようにするためである。

調査協力者数は語彙テスト 133 名、文法テスト 134 名、実施地域はベトナム・韓国・台湾で、実施時期は 2020 年 10 月から 11 月であった。いずれも国外の大学で日本語を学ぶ JFL 環境の学生で、平均学習歴は語彙テストで 37.56 月、文法テストで 37.23 月であった。各国の協力者の情報を表 2 に示す。

表 2 実施地域別の調査協力者数と平均学習歴

実施地域	語彙テスト		文法テスト	
	人数 ( <i>n</i> )	平均学習歴 (月)	人数 ( <i>n</i> )	平均学習歴 (月)
ベトナム	17	37.13	17	37.13
韓国	37	40.67	39	39.63
台湾	79	36.24	78	36.09
計	133	37.56	134	37.23

## 5. 結果

内的整合性の指標であるクロンバックの  $\alpha$  係数は、語彙テスト 0.945、文法テスト 0.951 であり、いずれも高い信頼性が確認された。正答率の平均値は語彙テスト 0.68、文法テスト 0.67 で、標準偏差はそれぞれ 0.22 と 0.19 であった。各レベルの平均正答率と標準偏差について、語彙テストは表 3 に、文法テストは表 4 に示す。

各レベル間の正答率に差があるか検討するために、有意水準を 5% に設定し、一元配置分散分析を行った。その結果、語彙テスト ( $F(3,76) = 29.70, p < .05, \eta^2 = .54$ )、文法テスト ( $F(3,68) = 29.57, p < .05, \eta^2 = .57$ ) とともに主効果が有意であった。holm の方法による多重比較の結果、語彙テストは初級と中級でのみ有意でなく ( $t(76) = .37, n.s., r = .04$ )、文法テストは 4 級と 3 級でのみ有意でなかった ( $t(68) = .47, n.s., r = .06$ )。下位レベルが有意でなかったのは協力者のレベルが比較的高かったためだと考えられる。

つぎに、古典テスト理論を用い、項目分析を行い、問題を精査した。具体的には、項目困難度、項目弁別力指数、選択肢率についてみていく。

項目困難度 (Item difficulty) は、テスト項目がどのくらい難しかったかを検討するもので、正答率、あるいは通過率とも呼ばれる (中村 2002)。語彙テスト・文法テストの各レベルにおける問題の正答率分布を図 1、図 2 に示す (ラベルには出題 ID を示し、その 1 文字目が A のものは語

表3 語彙テストの問題レベル別平均正答率と標準偏差

問題レベル	問題数 (n)	平均正答率	標準偏差
初級	20	0.84	0.14
中級	20	0.82	0.11
中上級	20	0.59	0.17
上級	20	0.46	0.17

表4 文法テストの問題レベル別平均正答率と標準偏差

問題レベル	問題数 (n)	平均正答率	標準偏差
4級	18	0.80	0.15
3級	18	0.78	0.11
2級	18	0.63	0.14
1級	18	0.45	0.10

彙テスト、Bのものは文法テストを指す)。ボックスプロットの上下レンジから逸脱している7問は、外れ値として除外した(語彙テストではA\_1\_04, A\_1\_12、文法テストではB\_1\_07、B\_1\_16、B\_1\_17、B\_3\_03、B\_3\_18)。正答率の外れ値の中でB\_1\_04だけは、点双列関数が0.64と高く、錯乱肢の選択率もよいため、弁別力の高い問題として残した。

項目弁別力指数 (Item discrimination power index) は、あるテスト項目が、能力の高い受験者と低い受験者を弁別、あるいは識別することができたかどうかを検討するもので、-1.000 から +1.000 の範囲で示され、+1.000 に近くなればなるほど項目弁別力は高いと判断される(中村2002)。本研究では、点双列相関係数を用いて項目弁別力を算出した。語彙テストの点双列相関係数の最小値は0.055、最大値は0.644であった。文法テストでは、最小値0.229、最大値0.699であった。中村(2002)では、点双列相関係数を用いた場合は0.300以上であればよいとする考え方が一般的とされていることから、本研究においても、0.300を下回る項目を除外した。

選択肢率については、錯乱肢が選ばれた割合が極端に偏っていないかなどを確認し、0%のものを修正対象とし、選択肢の修正を行った。

最後に、G-P分析図を描き、除外項目の精査を行った。G-P分析図は、25%ずつ分類した合計得点ランクの順に、選択肢が選ばれた割合を示すものである。合計得点ランクが高まるとともに、正答選択肢の割合が高まる、つまり右肩上がりになっていることが望ましく、勾配が最も急な得点ランク帯は、その項目が最も感度を発揮するものと考えられる。他方、勾配が右肩下がりになってしまう項目は、日本語能力を正確に測定していないか、あるいは魅力的な錯乱肢の影響が考えられる。語彙テストの初級問題を例に、G-P分析図をみていきたい(図3;ラベルは出題IDを示す)。正答率が外れ値であったA\_1\_12は、得点ランクが高くなっても正答率が低いままであり、他の初級問題の傾向とも異なっていることが見てとれる。同じく正答率が外れ値であったA\_1\_04は得点ランクが50%-70%と75%-100%のところ最も勾配が急になっているが、これは初級の問題であ

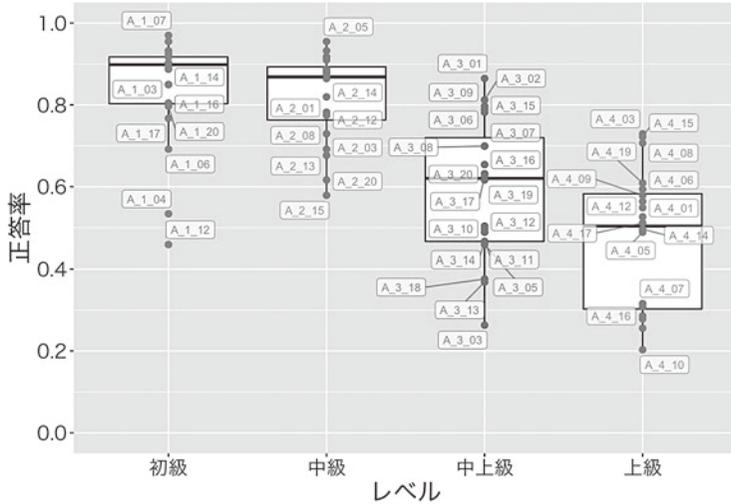


図1 語彙テストにおけるレベル別の正答率分布

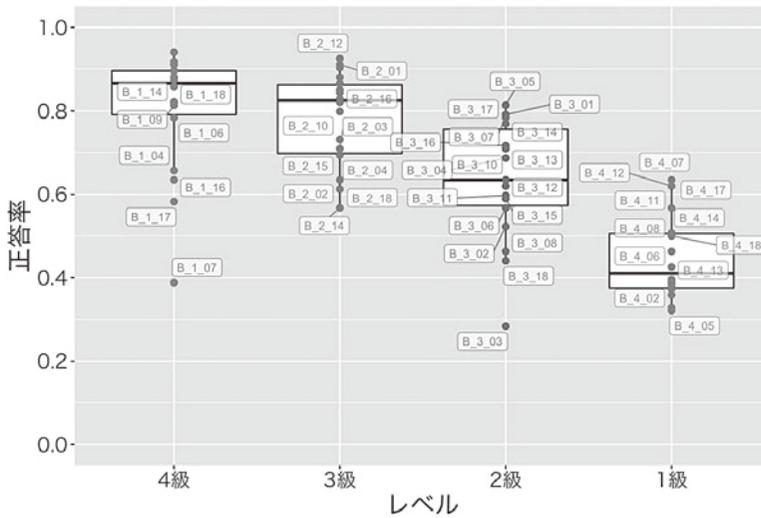


図2 文法テストにおけるレベル別の正答率分布

るため、適切な難易度の問題とは言い難い。一方、正答率が0.69と他の初級問題よりも低かったA\_1\_06は、G-P分析図で見ると、得点ランクが上がるにつれて、正答の選択率が右肩上がりになっていることがわかる。A\_1\_06は初級の問題としては確かに難易度の高いものではあるかもしれないが、合計得点が高いグループではしっかりと答えられており、不適切な問題とも言い難く、今回は許容範囲であると判断した。このように、語彙テスト・文法テストともに、一つずつ傾向を分析し、除外項目を検討した。

上記すべての結果を踏まえ、除外項目を選定し、最終的に語彙テスト・文法テストともに各60

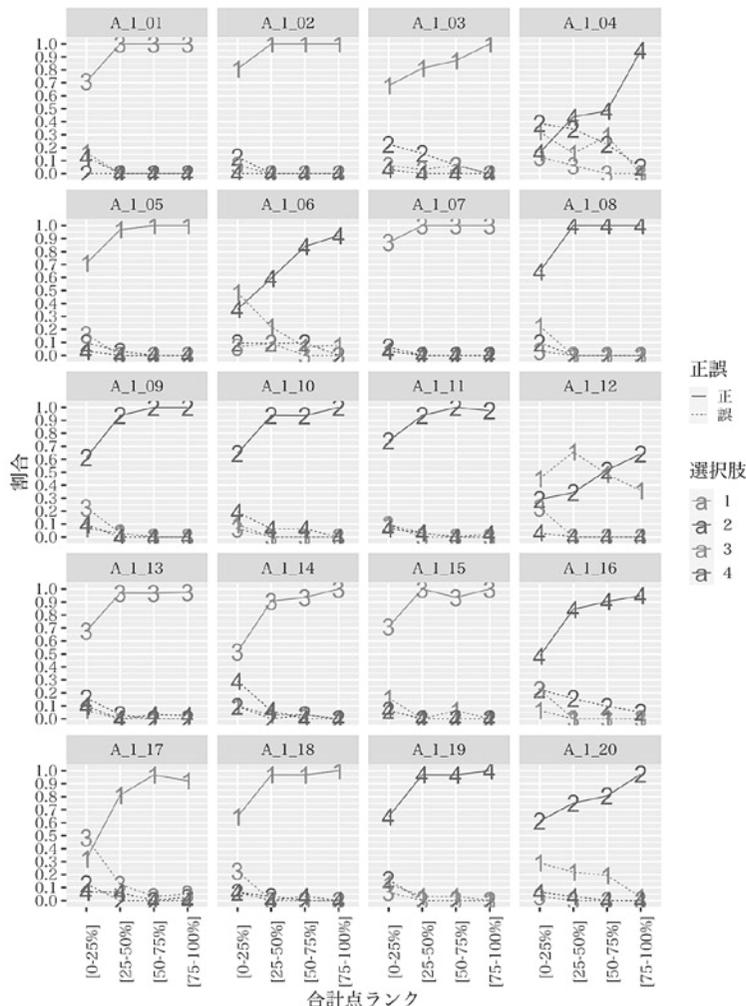


図3 語彙テストにおける初級の問題項目のG-P分析図

問のテストを作成した。

## 6. まとめと今後の課題

本研究では、日本語教育研究での利用を目的とした「かんたん日本語テスト (Simple Japanese Test)」ごい・ぶんぼうを開発した。本テストの一番の特徴は、日本語教育研究で利用しやすくなるために、多様な環境で安定的に使用でき、簡便に運用でき、かつ学習者に負担をかけすぎずに実施できることである。本テストを作成するにあたり、試行調査を実施し、信頼性の検討と項目分析を行った。項目分析では、項目困難度、項目弁別指数、選択肢率、G-P分析図をもとに問題を精査し、実用に耐えうるテストに整えた。本テストを日本語教育研究に用いることで、簡易的に語彙

や文法の日本語能力を測定し、日本語レベルの統制を行うことができるようになる。

また、本テストの開発の発端は、書き言葉に関する縦断調査で使用できる簡便で実用的なテストがほしいというものであった。作文など書き言葉の産出を評価することは難しく、労力を伴うものである。今後、本テストと書き言葉の産出との関係を見ていくことで、書き言葉の能力を客観テストにより測定できるようにしていきたい。さらに、本テストは各30分としたが、解答率を観察しながら、より適切な時間も検討していきたい。このように分析を進め、より使いやすいテストに整えた上で、最終的には研究者に公開し、広く日本語教育関係者に使用してもらえるテストにしていきたい。

#### 〈付記／謝辞〉

本研究は、科研費 JP21H04417 および国立国語研究所機関拠点型基幹研究プロジェクト「多様な言語資源に基づく日本語非母語話者の言語運用の応用的研究」の研究成果です。テスト問題の作成にご協力をいただきました金井勇人先生、佐野彩子先生、本多由美子先生、調査にご協力いただきました東國大・東國大・東國大・東國大の皆様に心より感謝申し上げます。

#### 〈参考文献〉

- 今井新悟（2015）「第4章 J-CAT（Japanese Computerized Adaptive Test）」李在鎬（編）『日本語教育のための言語テストガイドブック』くろしお出版。
- 大隅敦子・谷内美智子（2015）「第2章 日本語能力試験」李在鎬（編）『日本語教育のための言語テストガイドブック』くろしお出版。
- 国際交流基金・日本国際教育支援協会 著・編（2002）『日本語能力試験 出題基準 改訂版』凡人社。
- 小林典子（2015）「第6章 SPOT（Simple Performance-Oriented Test）」李在鎬（編）『日本語教育のための言語テストガイドブック』くろしお出版。
- 酒井たか子・加納千恵子・小林典子（2015）「第5章 TTB（Tsukuba Test-Battery of Japanese）」李在鎬（編）『日本語教育のための言語テストガイドブック』くろしお出版。
- 中村洋一（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。
- 日本語記述文法研究会編（2003-2010）『現代日本語文法 全7巻』くろしお出版。
- 国際交流基金日本語基礎テスト〈<https://www.jpf.go.jp/jft-basic/>〉最終閲覧日 2022年11月30日。
- 日本語教育語彙表〈<http://jhlee.sakura.ne.jp/JEV/>〉最終閲覧日 2022年11月30日。
- 日本語能力試験〈<https://www.jlpt.jp/about/message.html>〉最終閲覧日 2022年11月30日。