

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——

宮岡 弥生*・玉岡賀津雄**・酒井 弘***

1. はじめに

第二言語学習者を対象に目標言語に関する質問紙調査や実験を実施する場合、学習者の第二言語能力を測定してデータの分析に利用することがある。このとき、第二言語能力を測定するための言語テストは、主たる目的ではなく質問紙調査や実験に付随して行われるものではあるが、被験者集団の能力別のグループ分けに用いるなど分析結果に大きく影響するため、高い弁別力と信頼性が要求される。しかも、本来の調査や実験などと合わせて実施されることが前提となるため、短時間で完了するものでなくてはならない。つまり、言語能力測定用の付随テストは、少ない設問数で高い信頼性を備えたテストであることが求められる。

言語能力を測定するテストは、テストの目的によって大きく二つに分けられる。一つは、受験者を一般的言語能力や言語熟達度に基づいて直線上に並べ、受験者の間に存在する差を明らかにする「集団基準準拠テスト (norm-referenced test: NRT)」¹⁾、もう一つは、個々の受験者が習得した知識や技能といった言語材料の量を測定する「目的規準準拠テスト (criterion-referenced test: CRT)」である (ブラウン, 1999)¹⁾。

NRT の代表例は、英語の熟達度判定に用いられている TOEFL (Test of English as a Foreign Language) である。アメリカ、イギリス、オー

ストラリア、カナダ、ニュージーランドなど130か国の9,000以上の教育機関が、英語能力を証明するテストとして TOEFL のスコアを利用している。NRT においては、受験者の一般的な言語運用能力を他の受験者と比較する必要があるため、テストの得点が正規分布するように問題が作成してある。

一方、CRT は特定の授業、プログラムといったものに関わる明確で具体的な目標を測定するために問題が作成され、その得点は目標に示されている知識や技能を何パーセント習得したかを示すものであり、得点分布は必ずしも正規分布を描く必要はない (ブラウン, 1999)。これら二つのテストの本質的な違いは、得点の解釈の仕方である。NRT の得点は相対的であると解釈され、ある受験者の成績はパーセンタイル値によって他の受験者の成績と比較される。これに対して CRT の得点は絶対的で、ある受験者の成績は習得した言語材料の量、すなわち、パーセンテージで対比される (ブラウン, 1999)。

NRT と CRT のうち、言語能力測定用の付随テストに適しているのは NRT のほうであろう。これは、「得点を算出して、受験者を能力に基づいて直線上に並べ、受験者の間に存在する差を明らかにする」(ブラウン, 1999) という NRT の目的と、被験者集団を個々の第二言語能力の差に基づいてグループ分けするという付随テストの目的が合致しているからである。したがって、言語能力測定用の付随テストは、NRT の形式で作成されるのが望ましい。

ブラウン (1999) によると、NRT の開発や改

* 広島経済大学経済学部教授

** 名古屋大学大学院国際言語文化研究科教授

*** 広島大学大学院教育学研究科教授

良は、次のような手順で行われる。

1. 当該テストを使って、最終的に評価する予定の学習者群と類似の学習者群に対してかなりの数のテスト項目を予備的にテストする。
2. 項目形式分析と統計技術を使ってテスト項目を分析する。
3. 最良のテスト項目を選定し、より短くより効率的な改訂版を作る。

このブラウン（1999）の手順1では、最終的に評価する予定の学習者群と類似の学習者群に対して予備的なテストを行うことになっている。しかし、現実的な問題として、テスト実施の対象となる相当数の学習者群とは別に予備テストのための学習者群を手配することは容易なことではない。仮に、予備テストを省き、一度のテストのみで十分な信頼性を確保できる方法があるのであれば、そちらに注力するほうがテストの実施および開発のコストが低くなると考えられる。宮岡・玉岡・酒井（2011）では、この考え方に基づいて作成した語彙テストについて述べた。この語彙テストでは、設問の特性を多面的に統制することによって、高い弁別力のみならず多角的な分析が可能となった。本稿では、この語彙テストと並行して作成した文法テストについて、開発の方法とその後の改訂について述べる。

2. 文法テスト項目の構成

一つのまとまった文を理解するのに必要な単語内外の関係の一連の規則を文法知識と定義した。具体的には、図1に示したように、形態素変化（morphological inflections）、局所依存（local dependency）、構造の複雑性（complex structure）の3種類の下位範疇で文法知識を構成した。つまり、一つの単語内での形態素変化、隣り合う単語同士の間局所的な依存関係、さらに一つの文の中で離れて位置する単語同士の関係からなる構造の複雑性である。

日本語は、多様な形態素の組み合わせで様々な文法的な表現を生み出す。特に、動詞に結合する形態素は多様である。そのために、国文法あるいは学校文法では、助動詞という品詞を設けており、これは、時制（テンス）、相（アスペクト）、態（ヴォイス）、法（ムード）などの文法機能を表す品詞と定義されている。たとえば「殴られそうになる」であれば、「殴ら」が「殴る」の五段活用の未然形、「れ」は受け身の助動詞「れる」の連用形、「そうに」は様態の助動詞の「そうだ」の連用形、「なる」は五段活用の動詞の終止形であると分析する。本研究でいう形態素変化のテスト問題は、こうした助動詞を含む動詞や形容詞の変化の理解を問うものである。たとえば、「誤って花びんを壊した私を、父は（ ）。」の括弧の中に入る正しい表現を、「責めないだった」「責めるなかった」「責めな

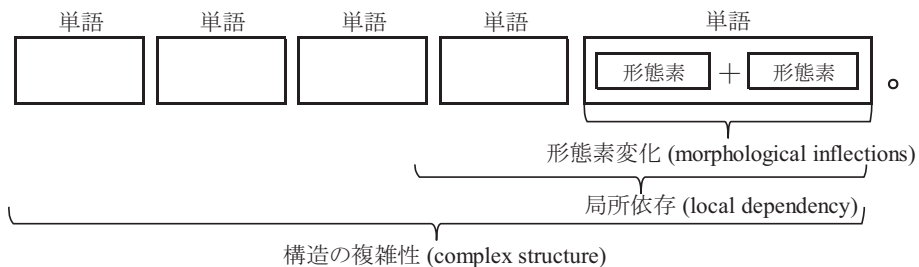


図1 文法知識を構成する3つの下位範疇

かった」「責めなくて良かった」の四つの中から一つ選ぶ問題である。この場合、「責めなかった」が正しい答えであるが、問題となっている部分以外の意味が分からなくても形態素の変化規則を知っているだけで正しく回答できる。

局所依存は、二つの隣り合う単語が正しく結び付けられるかどうかを問うテスト問題で構成した。たとえば、「彼女はいつも卵焼きを()作る。」という文の括弧の中に入る表現として、「上手な」「上手で」「上手の」「上手に」から正しい答えを一つ選ぶ問題である。この問題では、四つの選択肢の表現自体は正しい。しかし、次にくる「作る」という動詞との関係において、「上手な作る」「上手で作る」「上手の作る」は誤りであり、「上手に作る」だけが正しい答えであることが分かる。このように、隣り合う二つの単語の組み合わせが正しいかどうかを問う問題である。

構造の複雑性は、二つの要素が局所依存のように隣り合うのではなく、一文の中で離れたところに位置している場合に、全体的な文の構造を正しく構築できるかどうかを問うものである。たとえば、「どんなに彼女が(), あの大学には合格しないだろう。」であれば、「どんなに」に繋がる構造になるものを「がんばって」「がんばるが」「がんばっても」「がんばるのに」の四つから選ぶ。この場合は、「どんなに」と「～ても」の共起関係が理解できていれば、「がんばっても」を正しく選択できるはずである。

これら3つの下位範疇がそれぞれ同数の12問ずつとなるように問題数を統制した。この文法テストと同様に言語能力測定用の付随テストとして作成した語彙テスト(宮岡・玉岡・酒井, 2011)は、日本語能力試験配当級(旧日本語能力試験用, 国際交流基金, 2002)などによって設問の特性を多面的に統制してあるため、各下位範疇の得点を直接比較することができる。本文法テストも、文法項目は日本語能力試験配当

級が3級および4級のもののみを選定し、できるだけ下位範疇ごとの難易度の偏りがないように努めた。しかしながら、「形態素変化」「局所依存」「構造の複雑性」という文法的特性を優先させたため、日本語能力試験配当級を完全には統制することができなかった。したがって、文法の下位範疇について得点を直接比較することはできない。

3. 文法テストの実施

——2006年西安調査——

3.1 調査協力者

中国の西安外国語大学で日本語を専攻する1年終了時の学生146名(女性124名, 男性22名)、2年終了時の学生135名(女性116名, 男性19名)の合計281名に対して、2006年5月に文法テストを実施した。この文法テストは、総合的な日本語能力テスト(語彙, 文法, 談話, 読解, 聴解)の一つとして行った。調査協力者の学生たちは、日本への留学経験はない。年齢は、最も低かった学生が16歳4ヶ月、最も高かった学生が26歳8ヶ月であった。1年終了時の学生の平均年齢は19歳8ヶ月で標準偏差が12ヶ月、2年終了時の学生は20歳9ヶ月で標準偏差が1年2ヶ月であった。2年終了時の学生の方が1年終了時の学生よりも年齢が有意に高かった [$t(279) = 8.217, p < .001$]。

3.2 文法テストの得点と信頼性係数

文法テストの問題形式は、四者択一の空所補充である。文法テストは36問で構成され、各問1点の36点満点である。表2に示したように、独立したサンプルの t 検定の結果、2年終了生の方が1年終了生よりも合計の得点が有意に高かった。この文法テストの得点は全体的に高かったが、特に局所依存は1年終了生でも12点満点の平均で10.14点で、2年終了生の平均の10.28点と比べて差はなかった。このことから、

表2 文法テストの平均, 標準偏差および t 検定の結果

文法テストの 下位カテゴリ	満点	合計 ($n=281$)		1年終了生 ($n=146$)		2年終了生 ($n=135$)		t 検定の結果
		平均	標準偏差	平均	標準偏差	平均	標準偏差	
形態素変化	12	8.16	2.35	6.66	1.83	9.77	1.70	$t(279) = 14.694, p < .001$
局所依存	12	10.21	1.40	10.14	1.52	10.28	1.25	<i>n.s.</i>
構造の複雑性	12	7.57	2.40	6.09	1.76	9.17	1.94	$t(279) = 13.964, p < .001$
合計	36	25.94	5.06	22.90	3.88	29.22	4.03	$t(279) = 13.411, p < .001$

この局所依存の問題は1年生の時点で習得できるものであったことがうかがえる。これに対して、形態素変化と構造の複雑性については、どちらも2年終了生の方が1年終了生よりも有意に得点が高かった。

文法テスト全体の得点について、テストデータ分析プログラムの「TDAP (Test Data Analysis Program) Ver. 2.0」(大友・中村・秋山, 2002)を用いて信頼性係数(クロンバックの α)を算出したところ、0.795であった。信頼性とは、そのテストを用いて同一の受験者を測定した場合、どれだけ安定した結果が得られるかを検討するものである(中村, 2002)。信頼性係数は0.000から1.000の数値となり、1.000に近ければ近いほど信頼性の高いテストであると解釈できる。表3に問題数が増えた場合の信頼性の

表3 問題数が増えた場合の信頼性係数の変化

信頼性係数の予測値	問題数の予測値
0.800	37
0.820	42
0.840	49
0.860	57
0.880	68
0.900	83
0.920	107
0.940	145
0.960	223
0.980	454

変化を示した。問題数が増えれば増えるほど信頼性は高くなるが、言語能力測定用の付随テストの場合、所要時間は短いことが望ましいため、テスト実施に関わる現実的な問題も考慮に入れながら最適な問題数を探ることになる。どのくらいの信頼性係数があればよいのかについては諸説あるが、中村(2002)は0.800が1つの目安になるとしている。本文法テストの信頼性係数は0.795であったことから、比較的信頼性の高いテストであると言えるだろう。

3.3 文法テストの項目分析

テストデータ分析プログラムの「TDAP (Test Data Analysis Program) Ver. 2.0」(大友・中村・秋山, 2002)を用いて、文法テストの問題について項目分析を行った。算出された指標のうち、「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つを表1に示した。

3.3.1 項目困難度

項目困難度 (DIFF) 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、0.000から1.000の間の値をとる。項目困難度は実は正答率であるため、1.000に近ければ近いほど易しい項目、0.000に近ければ近いほど難しい項目となる(中村, 2002)。中村(2002)によると、この文法テストのように選択肢が四つある場合、偶然

の正当確率が4分の1であるため、それを加味して最適困難度は0.625 ($0.500 + 0.500 \times 1/4 = 0.625$)となる。今回の文法テストの項目困難度を見ると(表1参照)、局所依存のカテゴリに数値の高い項目、つまり1.000に近く易しい項目が多数あることがわかる。例えば、ID23の疑問詞に関する問題である「あなたは(どんな)映画が好きですか」(括弧内は正答)は、項目困難度が0.996と非常に高く、ほとんどの調査協力者が正答している問題であった。このほか、局所依存には項目困難度が0.625を超える問題が12問中10問あり、そのうち7問が0.900を超える非常に易しい問題となっていた。これらの問題の難易度を上げることによって、文法テスト全体の信頼性係数が上がると考えられる。

3.3.2 項目弁別力指数

「項目弁別力指数(DISC)」とは、あるテスト項目が日本語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指標である。-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)。中村(2002)によると、本文法テストのように、受験者グループ全員のデータを用いて算出する「点双列相関係数による項目弁別力指数」を用いた場合には、0.300以上あればよいとする考え方が一般的だということである。表1に示したように、本文法テストでは、項目弁別力指数が0.300に満たなかった項目は局所依存のカテゴリに多かった。12問中7問が項目弁別力の低い問題で、そのうち6問は項目困難度が低い問題と重なっていた。難易度が低すぎる問題は得点に天井効果をもたらし、問題の弁別力が低くなっているのだと考えられる。

3.3.3 実質選択肢数

実質選択肢数(AENO)とは、多肢選択形式(multiple choice format)の問題項目で、準備した選択肢が偏りなく選択されているかどうか、

準備した選択肢が「実質的」にはいくつ分の選択肢として働いたかを検討するものである(中村, 2002)。実質選択肢数の数値は0.000から選択肢の数までとなるため、選択肢が四つある本文法テストでは、0.000から4.000までとなる。したがって、実質選択肢数が4.000に近ければ近いほど正答以外の錯乱肢が有効に働いたと考えられるが、2.500を超えていれば概ね適切な錯乱肢であったと言えるであろう(中村, 2002)。表1に示したように、本文法テストで実質選択肢数が2.500よりも低かった問題は局所依存のカテゴリに集中しており、12問中11問にも上った。さらに、そのうち10問が、項目困難度が低かった問題と重なっていた。

4. 文法テストの改訂①

——2008年西安調査——

2006年に西安で実施した文法テストの結果をふまえて問題を改訂し、2008年に西安で調査を実施した。問題の改訂は、2006年西安調査の際に主に項目困難度の数値が高かったもの、つまり難易度の低かった問題を中心に、項目弁別力および実質選択肢数の値も考慮に入れながら、難易度を上げる方向で行った(表1および表4参照)。項目困難度の数値が高かった問題は局所依存のカテゴリに集中していたため、このカテゴリについては実に12問中10問を改訂することとなった。改訂は、ターゲットとする文法項目や品詞は変えず、それらの文法事項や品詞を具体化した際の語彙の難易度を上げることで行った。つまり、問題文中の空所に入れる語彙の難易度を上げたのである。また、「私は～と思う」という主述の一致に関する問題であるID27「私は彼が腹を立てたのは(当然だと思う)。」は実質選択肢数が低く、錯乱肢の作り方に問題があると思われたため、錯乱肢を一部差し替えた。

文法テスト改訂版を、2008年9月、中国西安

表 4 2008年西安調査の際に改訂した問題

カテゴリ	ID	品詞／文法項目	級	問題文 (正解)	錯乱肢
形態素変化 (morphological inflections)	2	活用 動詞	4	私の母はちょっと(太っています)。	太っています 太りています 太るています 太っています
	14	格助詞	4	私は毎朝7時発の(バスに)乗って大学に行きます。	バスを バスで バスの
局所依存 (local dependency)	15	格助詞	4	私の弟は来年(大学を)卒業する。	大学から 大学に 大学で
	17	形容詞 + 名詞	4	私は(かっこいい)人が好きだ。	かっこいいの かっこよく かっこいいな
	18	形容詞 + 動詞	4	テレビの音を(小さく)してください。	小さいに 小さい 小さくて
	19	形容動詞 + 名詞	4	山田先生は、とても教育に(熱心な)先生です。	熱心の 熱心に 熱心い
	20	形容動詞 + 動詞	4	友達と一緒に教室を(きれいに)掃除しました。	きれいく きれいな きれい
	21	数詞	4	先週、パソコンを(2台)買いました。	2個 2本 2冊
	22	こそあど	4	(この)かばんはとても重い。	ここ これ こちら
	23	疑問詞	4	あなたのお父さんは(何)才ですか。	いくつ いくら いつ
	24	動詞の連体修飾	3	これは私の兄が(とった)写真です。	とって とるの とったの
	構造の複雑性 (complex structure)	27	主述の一致	3	私は彼が腹を立てたのは(当然だと思う)。

外国語大学の日本語専攻学生224名（2年終了時と3年終了時の各112名）に対して、日本語力の総合的なテスト（「のだ」「のか」の使用／非使用72問、文法知識36問、語彙知識48問、読解12問の4種類、所要時間約120分）の一つとして実施した（趙・玉岡・木山，2009）。分析の結果、文法テストの信頼性係数（クロンバックの α ）は0.733と比較的高かったが、改訂前の0.795よりも若干下がった。これは、改訂前の2006年西安調査の調査協力者が1年終了時と2年終了時の学習者であったのに対して、2008年西安調査では2年終了時および3年終了時の学習者であったためであると考えられる。2008年の西安調査では同時に、2006年西安調査の際の語彙テストと同じものを使用して調査を行ったが、この語彙テストの信頼性係数は0.859と高かった。日本語能力試験において文法に分類される項目の多くは初級の段階で学習するもので、日本語能力試験配当級の3級もしくは4級に属している。2008年西安調査に際して、文法項目を具体化する語彙の難易度を上げることによって改訂を行ったが、信頼性係数は上がらなかった。このことから、日本語能力を測定するテストで文法に分類される問題の難易度を上げるためには、『日本語能力試験出題基準』（国際交流基金，2002）で「機能語」に分類されている語句に関する問題を設定するのが良いと思われる。

5. 文法テストの改訂②

——2010—2012年日本国内外調査——

2010年から2012年にかけて日本国内外で実施した学習者作文コーパスデータ収集（李・宮岡・林，2013）の際に文法テストを実施する必要性が生じたことから、2008年西安調査で使用した文法テストを改訂して用いることにした。2008年西安調査では、2006年西安調査の際の文法テストを改訂したにもかかわらず、2006年のときと比べて文法テストの信頼性係数が上がら

なかった。そこで、この日本国内外調査においては改訂の方向性を修正して再改訂を行った。

主な改訂点は、『日本語能力試験出題基準』（国際交流基金，2002）で「機能語」に分類されている語句に関する問題を設定したことである。表5に示したように、この改訂版では文法項目の下位範疇と問題数を見直し、形態素変化の問題を8問、局所依存を8問、構造の複雑性を8問、機能語を12問設定した。この文法テストは他の調査や実験に付随して日本語能力を測定するために使用するものであるため、テストにかかる時間ができるだけ短くなるようにする必要がある。そのため、機能語をテストに含めるにあたって、これまでの2回のテストでは12問ずつあった形態素変化、局所依存、構造の複雑性の問題数を4問ずつ削って各8問とし、テスト全体の問題数が増えないよう配慮した。

日本語能力試験配当級は、形態素変化、局所依存、構造の複雑性の3つのカテゴリ全体で3級と4級が同数の12問ずつとした。機能語は、1級と2級がそれぞれ6問ずつである。この機能語の問題は、宮岡・玉岡・酒井（2011）で述べた語彙テストに含まれていたものである。機能語を語彙と文法のどちらの範疇に属するとみなすかは、意見が分かれるところであろう。しかし、純粋な文法項目は3級もしくは4級の段階で学習するものが多く、これらの上に絞ったテストでは得点の天井効果が出て高い信頼性係数が得られない可能性があることが2008年西安調査で明らかになった。そこで、この再改訂版では1級および2級配当となる機能語を入れることによって、テスト全体として天井効果が出るのを防ぐことにした。

文法テストは、日本語学習者の作文コーパスデータ収集の一環として、中国語母語話者86名（日本語学習歴2年未満が4名、2年以上5年未満が80名、5年以上が2名）、および韓国語母語話者92名（日本語学習歴2年未満が20名、2年

表5 文法テスト2010年改訂版の項目と問題文

カテゴリー	ID	品詞/文法項目	受け身	級	問題文 (正解)	錯	乱	脱
形態論 (morphological inflections)	1	接続 助動詞	受け身	3	道でだれかに肩を(たたか)れて、振り向いたら友達だった。	たたかれて	たたかれて	
	2	接続 助動詞	う・よう	3	このお金を銀行に(預け)ようと思う。	預けよう	預けよう	
	3	接続 助動詞	やすい・にくい	3	あの人なら、仕事に(頼み)やすいと思う。	頼むやすい	頼んでやすい	
	4	接続 助動詞	使役受け身	3	もう少しで留学を(あきら)めさせられそうになった。	あきらめられさせうになった	あきらめさせうになった	
	5	活用 動詞	使役受け身	4	私の母はちよっと(大)っています。	大っています	大っています	
	6	活用 形容詞	て形	4	食事を作るのが(面倒)くさくて、ほとんど外食ですませています。	面倒くさいで	面倒くさくて	
	7	活用 形容動詞	現在・否定	4	私はあなたのわがままに(つき)きあっているほど(ひま)じゃやない。	ひまじゃない	ひまくない	
	8	活用 形容動詞	過去・肯定	4	彼の答えは(あいまい)だった)ので、彼の本心はわからない。	あいまいだった	あいまいかった	
局所依存 (local dependency)	9	自他動詞		3	やかんでお湯を(わか)して、ポットに入れておいてください。	わいて	わく	
	10	動詞の連体修飾	～た	3	これは私の母が(と)った写真です。	とって	とったの	
	11	格助詞	に	4	私は毎朝7時発の(バス)に乗って大学に行きます。	バスで	バスの	
	12	形容詞 + 動詞		4	テレビの音を(小さ)くしてください。	小さいに	小さくて	
	13	形容動詞 + 動詞		4	友達と一緒に教室を(きれ)いに掃除しました。	きれいに	きれいな	
	14	数詞		4	去年、車を(2台)買いました。	2個	2冊	
	15	こそあど		4	(この)かばんはととても重い。	ここ	こちら	
	16	疑問詞	いくら	4	この自販車の値段は(いくら)ですか。	いくつ	いつ	
複雑構造 (complex structure)	17	条件	もし～たら	3	もし私が5時までここにいてもどって(来)なかったら、先に行ってください。	来なくて	来なかった	
	18	主述の一致	私は～と思う	3	私は彼が腹を立てたのは(当然)だと思つた。	当然です	当然ではないでしょうか	
	19	「～がる」と主語の一致		3	私の妹は有名人を見るに必ず(と)りたがる。	とりたいたです	とりたがるです	
	20	名詞句の中の「が」	タカシが買った本を～	3	(森山さんが)買った本を岡田さんは先生に見せた。	森山さんを	森山さんで	
	21	疑問詞 + ～ても		3	どんなに彼女が(が)んばつても)、あの大学には合格しないだろう。	がんばつて	がんばるが	
	22	アスペクト	これから～ところだ	3	私はこれから大学に(行)くところです。	行つたところです	行っている	
	23	時制	過去	4	去年、私は祖母のセーターを(編)んだ。	編む	編もう	
	24	副詞	まだ～ていない	4	今朝はまだ朝ごはんを(食)べていない。	食べて	食べなくて	
機能語 (function words)	25	ないことはない		2	講演会に行きたくはないが、ぜひにと頼まれれば(行)かないことはない。	行かない	行つたことはない	
	26	からして		2	個性的な彼女も、着ているもの(から)して)ふつうの人は少し違う。	からしては	から言う	
	27	だけあって		2	日本に10年ものいる(だけ)あって)、彼はとても日本語が上手だ。	だけでなく、	からといって	
	28	～から～にかけて		2	昨夜から今朝(に)かけて)、日本各地で大雪が降った。	とかけて	にして	
	29	～ざるをえない		2	こんなひどい雨では、試合は(中)止せざるをえない。	中止するに	中止するにすぎない	
	30	～から～といって		2	(きらい)だからといって)野菜を食べないのは体に良くない。	きらいなくせに	きらいだといえ	
	31	～ではあるまいし		1	ことども(ではある)まいし)、大學生のあなた(が)ひとり旅行でいきいわけないでしょう。	にもかかわらず	きらいなだけあって	
	32	～が早いか		1	「火事だ」と(さ)げおが早いか)、彼は外へ飛び出して行った。	さげお最中	さげおうちに	
	33	～を余儀なくされる		1	材料費が値上がりしたため、我が社は商品の値上げを(余)儀なくされた。	余儀なくさせた	余儀なくできた	
	34	～たところだ		1	今から(勉)強したところだ、成績の悪い私が大学に合格するのは無理だろう。	勉強したところ	勉強したところが	
	35	～に至るまで		1	あのスーパーでは、食料品から洋服(に至る)まで)何でも売っている。	勉強するところ	勉強するところが	
	36	～を皮切りに		1	その選手は、今日の大会でのコンサートを(皮)切りに)、いつものように全国ツアーに出かける。	きつかけに	にわたって	

以上5年未満が60名、5年以上が12名)の計178名に対して実施した。文法テストの所要時間は、これと同時に実施した語彙テストと合わせて50分である。

得点を集計した結果、文法テストの平均値は23.26(1問1点で36点満点)、標準偏差は8.82であった。文法テスト全体の得点について信頼性係数(クロンバックの α)を算出したところ、0.80と高かった。これは、改訂前のどの文法テストよりも高い数値であった。中村(2002)でも、信頼性係数は0.800が1つの目安になるとされており、この文法テストは高い信頼性を有していると言えるであろう。

6. おわりに

本稿では、質問紙調査や実験に付随して日本語能力を測定するための文法テストの作成と改訂の過程について、信頼性係数を中心に述べた。本文法テストの開発に際しては、語彙テスト(宮岡・玉岡・酒井, 2011)の場合と同様に、第1回目のテストの段階から設問項目の特性を出来る限り統制して作成することによって、「集団基準準拠テスト(norm-referenced test: NRT)」の開発において通常実施される予備テストをなくし、初めから高い信頼性を持つテストとなることを目指した。その結果、第1回目の段階から文法テストの信頼性係数は比較的高かったが、項目分析をもとにして問題を差し替えるなどの改訂を行ったところ、さらに信頼性の高いテストとなった。信頼性係数とテストの項目数には一定した関係が認められ、一般的に項目数が多くなれば信頼性は高くなる(中村, 2002)。しかしながら、言語能力測定用の付随テストの場合には実施時間に制約があるため、項目数をできるだけ少なくせざるを得ない。少ない項目数で高い信頼性を第1回目のテストから実現するためには、項目の特性をできるだけ統制する必要がある。今後も、これまでのテスト

で統制し切れていなかった要素について統制を施し、さらに高い信頼性を少ない項目数で得られることを目指す。ただ注意が必要なのは、中村(2002)が指摘しているように、受験者が異なれば、信頼性係数の値が高くなる、あるいは低くなる可能性があり、ある受験者グループのデータを基に算出した信頼性係数は、そのテストの絶対的な特性として一般化できるデータではないということである。テストを実施したときには常に信頼性係数を算出し、実験の被験者や調査協力者の日本語能力が適切に測定できているかをチェックすることが肝要であろう。

本稿は、首都大学東京の萩原裕子教授が研究代表者となっていたプロジェクト「言語の発達・脳の成長・言語教育に関する統合的研究」(平成16年～平成21年度、独立行政法人科学技術振興機構(JST)・社会技術研究システム推進室・社会技術研究事業研究領域「脳科学と教育タイプII」)、および筑波大学の李在鎬准教授が研究代表者となっていた科研費・基盤研究(C)「自然言語処理の技術を利用したタグ付き学習者作文コーパスの開発」(課題番号:22520537,平成22～24年度)の助成を受けている。

謝辞

本研究に関連して行った調査に際してご協力いただいた皆様に、厚く御礼申し上げます。

注

- 1) 中村(2002)では、NRTは「集団規準準拠テスト」、CRTは「目標基準準拠テスト」の名称が用いられている。

引用文献

- J. D. ブラウン 著・和田 稔 訳(1999)『言語テストの基礎知識』大修館書店
- 国際交流基金(2002)『日本語能力試験出題基準[改訂版]』凡人社
- 趙 萍・玉岡賀津雄・木山幸子(2009)『『のだ』と『のか』の使用・非使用に関する文法及び語彙知識の影響』『言語学会第11回年次国際大会予稿集』57-60.
- 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
- 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), 1-18. 広島経済大学

- 李 在鎬・宮岡弥生・林 炫情 (2013) 「学習者コーパスと言語テスト——言語テストの得点と作文のテキスト情報量の関連性」『言語教育評価研究(AELE)』3, 22-31. 言語教育評価共同研究所

Test Data Analysis (TDAP) の履歴

1996 Test Data Analysis Program (TDAP) Ver. 1.0
[MS-DOS, N88-BASIC 版]

©大友賢二・中村洋一 (大友賢二1996『項目応答理論入門』大修館書店に添付)

2002 Test Data Analysis Program (TDAP) Ver. 2.0
[Windows 版]

©大友賢二・中村洋一・秋山 實 (大友賢二監修/中村洋一著『テストで言語能力は測れるか』桐原書店に添付)

補記：2010年度版文法テスト

文法能力

次の文の()に入れることばとして正しいものはどれですか。
4つの中から1つだけえらんで、□に✓を書いて下さい。

- 1 ()かばんはとても重い。
 これ この
 こちら ここ
- 2 講演会に行きたくはないが、ぜひにと頼まれば()。
 行くことはない 行かないことはない
 行かない 行ったことはない
- 3 どんなに彼女が()、あの大学には合格しないだろう。
 がんばっても がんばって
 がんばるのに がんばるが
- 4 このお金を銀行に()と思う。
 預けるよう 預けろう
 預けよう 預けう
- 5 やかんでお湯を()、ポットに入れておいてください。
 わかす わかして
 わいて わく
- 6 こんなにひどい雨では、試合は()。
 中止しない 中止するには及ばない
 中止せざるをえない 中止するにすぎない
- 7 私は彼が腹を立てたのは()。
 当然だ 当然です
 当然ではないでしょうか 当然だと思う
- 8 私の母はちょっと()。
 太っています 太りています
 太っています 太るっています
- 9 テレビの音を()してください。
 小さく 小さいに
 小さい 小さくて
- 10 材料費が値上がりしたため、我が社は商品の値上げを()。
 余儀なくさせた 余儀なくされた
 余儀なくした 余儀なくできた
- 11 友達と一緒に教室を()掃除しました。
 きれいく きれいな
 きれいに きれい
- 12 こども()、大学生のあなたがひとりで旅行できないわけはないでしょう。
 ではあるまいし にもかかわらず
 とはいえ ながらに

- 13 道でだれかに肩を()、振り向いたら友達だった。
 たたかられて たたかれて
 たたかれて たたきられて
- 14 私は毎朝7時発の()乗って大学に行きます。
 バスを バスで
 バスの バスに
- 15 ()買った本を岡田さんは先生に見せた。
 森山さんが 森山さんは
 森山さんを 森山さんで
- 16 昨夜から今朝()、日本各地で大雪が降った。
 とかけて にあって
 にかけて にして
- 17 その歌手は、今日の大阪でのコンサートを()、いつものように全国ツアーに出かける。
 皮切りに きっかけに
 もとに 契機に
- 18 去年、私は祖母のセーターを()。
 編む 編んだ
 編むだろう 編もう
- 19 私はあなたのわがままにつきあっているほど()。
 ひまにない ひまじゃない
 ひまくない ひまじゃない
- 20 もし私が5時までここにまで()、先に行ってください。
 来ない 来なくて
 来なかったら 来なかった
- 21 個性的な彼女は、着ているもの()ふつうの人とは少し違う。
 からして にしては
 のわりに から言うと
- 22 去年、車を()買いました。
 2本 2台
 2冊 2個
- 23 食事を作るのが()、ほとんど外食ですませています。
 面倒くさくて 面倒くさいで
 面倒くさかって 面倒くさって
- 24 ()野菜を食べないのは体に良くない。
 きらいなくせに きらいだといえば
 きらいだからといって きらいなだけあって
- 25 あの人なら、仕事が()。
 頼めやすい 頼みやすい
 頼むやすい 頼んでやすい
- 26 これは私の母が()写真です。
 とった とって
 とるの とったの
- 27 日本に10年もいる()、彼はとても日本語が上手だ。
 からには だけでなく
 だけあって からといって

- 28 あのスーパーでは、食料品から洋服()何でも売っている。
 に至るまで に至っては
 にわたる にわたって
- 29 今朝はまだ朝ごはんを()。
 食べた 食べる
 食べなくて 食べていない
- 30 この自転車の値段は()ですか。
 いつ 何
 いくら いくつ
- 31 彼の答えは()ので、彼の本心はわからない。
 あいまいだった あいまいなだった
 あいまかった あいまいかった
- 32 私はこれから大学に()。
 行ったところです 行っているところです
 行くところです 行っている
- 33 「火事だ」と()、彼は外へ飛び出して行った。
 さけぶが早いか さけび次第
 さけぶうちに さけぶ最中に
- 34 もう少しで留学を()。
 あきらめさせそうられになった あきらめられさせそうになった
 あきらめそうさせられになった あきらめさせられそうになった
- 35 今から()、成績の悪い私が大学に合格するのは無理だろう。
 勉強するところを 勉強したところが
 勉強したところで 勉強するところで
- 36 私の妹は有名人を見ると必ずいっしょに写真を()。
 とりたがる とりたい
 とりたいです とりたがるです