

## 日本語基礎能力テストの特性 (1)

### — 国語教育から見た語彙理解力測定項目の内容評価 —

伊藤 博美, \* 佐藤 洋之, \*\* 倉元 直樹 \*\*\*

\* 東北大学大学院文学研究科・秋田県立本荘高等学校

\*\* 東北大学大学院教育情報学教育部

\*\*\* 東北大学アドミッションセンター・東北大学大学院教育情報学教育部

**要旨:** 本研究の目的は、日本語基礎能力テスト(平他、1998)の語彙理解力項目について国語教育の観点からの妥当性を吟味することである。日本語基礎能力テストは、「語彙理解力」、「漢字読み取り能力」の2つの下位尺度から成っている。そのうち、語彙理解力テストの項目プールには小学校1年生用から高校生用まで470項目がある。

まず、項目内容を分類する基準として、「レベルA～C」を設定した。「レベルA」は「通時的に意味や用法が安定しているもの」であり、「レベルB」は「レベルAに準じ、かつ、各種の試験等で頻繁に出題されるもの」である。「レベルC」は「使用機会が少なくなってきた、ないしは、選択肢にやや難があるもの」である。さらに、各レベルに2～4の下位カテゴリーを設けた。

結果的に全体の約71%が「レベルA」に分類され、「レベルC」は5.6%に止まった。現在でも、全体として、十分に語彙理解力の測定に通用する尺度であることが確認された。

**キーワード:** 国語教育、日本語、項目プール、語彙、測定

### 1. 目的

本研究の目的は、平・前川・小野・林部・内田(1998)で作成された日本語基礎能力テスト項目プールのうち、日本語語彙理解力テストの各項目について、主に国語教育の観点から再吟味を行うものである。

日本語基礎能力テストの項目プールは、日本語語彙理解力テスト471項目、漢字読み取り能力テスト977項目から成る。それぞれの下位尺度は、項目反応理論に基づき2パラメタロジスティックモデルで尺度化されている(佐藤・伊藤・倉元、2003)。測定対象とする範囲は、小学校1年生程度から高校生程度まで幅広く、全ての範囲において同一尺度上で日本語の基礎能力が表現できることが特徴となっている。

語彙理解力テストは全て5肢選択形式に統一されており、ほとんどの項目が幹に示された語に最も近い意味を持つ語を選択肢から選ぶ形式を取っている。小学校1～2年生程度を対象に想定して作成された項目の一部には選択肢が絵で示されているものがあ

る。また、高校生程度を対象に想定して作成された項目の中には、幹に示された空所を選択肢に示された語で補充する形式のものもある。

漢字読み取りテストは、問題文の一部として示された漢字にふりがなを振る形式である。

それぞれの下位尺度の項目は、元々11の版(小学校1年生版～6年生版、中学生版、高校生91年版、高校生92年A版～C版)に冊子形式でまとめられて実施されたものである。アンカー項目として複数の版に採録されている項目もある。また、各版には、その後の項目分析の結果項目プールから削除された項目も含まれている。

日本語基礎能力テストの項目は、1980年代後半から1992年にわたる期間に作成された。現時点で、10年以上の時を経ている。また、項目作成の過程において、高校生用の一部には専門家(高校の国語教師)の関与があったものの、それ以外の項目に関しては国語教育の観点から十分な吟味が行われずに作成された。したがって、国語教育的に項目内容を

再吟味し、項目プールの内容的妥当性に裏づけを与えることは、今後の項目の利用に際しての重要な作業である。

本稿では、最初に国語教育の現状をまとめ、その観点から日本語基礎能力テストの教育的利用可能性を論じた後、それぞれの項目の評価を行う。なお、本研究の評価過程の中で、語彙理解力の項目プールから本質的に不適切な項目が1つ見出された。したがって、検討時点で日本語語彙理解力テストの項目プールは470項目となっている。

## 2. 国語教育の現状

学校教育において基礎的な言語能力の育成を多く負っている教科は国語であるが、現在移行期にある新学習指導要領において小学校国語、中学校国語、高等学校国語の教科目標はそれぞれ以下のようになっている。

### (1) 小学校国語

国語を適切に表現し正確に理解する能力を育成し、伝え合う力を高めるとともに、思考力や想像力及び言語感覚を養い、国語に対する関心を深め国語を尊重する態度を育てる。

### (2) 中学校国語

国語を適切に表現し正確に理解する能力を育成し、伝え合う力を高めるとともに、思考力や想像力を養い言語感覚を豊かにし、国語に対する認識を深め国語を尊重する態度を育てる。

### (3) 高等学校国語

国語を適切に表現し的確に理解する能力を育成し、伝え合う力を高めるとともに、思考力を伸ばし心情を豊かにし、言語感覚を磨き、言語文化に対する関心を深め、国語を尊重してその向上を図る態度を育てる。

教科目標は、中学校は小学校、高等学校は小・中学校のものをそれぞれ受けたものである。また、新指導要領では旧要領と比較して多くの点の変更が行われている。

例えば高等学校においては、「国語を適切に表現し的確に理解する能力を育成」という従前のものに加え、「伝え合う力」を高めることが位置づけられている。「伝え合う力」とは端的に言えば、「円滑に

相互伝達、相互理解を進めていく能力」のことでありとされるが、これは新指導要領が社会人として必要とされる言語能力の育成を重視することを目的の1つにしていることとも密接に関連している。

科目構成においては、必修科目として「国語表現Ⅰ」及び「国語総合」を設け、すべての生徒にいずれかを選択履修させることとしている。新たに設けられた「国語表現Ⅰ」では、国語の表現の特色や語句、語彙などに関する理解とともに、それらを用いて適切に表現する能力の育成が求められており、表現の基礎となる語句や語彙の習得が重視されている。「国語総合」においても「話すこと・聞くことを主とする指導」に15単位時間程度、「書くことを主とする指導」に30単位時間程度の配当時間が明示され、具体的な形で表現力の育成を求めるものとなっている。

このように見た場合、今後の国語教育において、豊かな表現力の基礎となる語句や語彙の習得はその目標の根幹と深く関わってくるといえるだろう。教育現場では、特に、語句、語彙に関する習得の程度を測定するための有効なテストの存在は欠かせないものとなり、ますます良質なものが求められる傾向が強くなることは間違いない。

日本語基礎能力テストは、今後想定される国語教育からの要請にも極めて有効な存在となるはずである。それゆえテストの項目に関しては、教育現場での運用の実用性という観点から、適切性および妥当性に関して十分に吟味される必要がある。以下、主として国語教育の観点から、日本語基礎能力テストの各項目の特性を中心に検討した結果を報告する。

## 3. 日本語基礎能力テストの有効性

日本語基礎能力テストは、様々な目的で利用可能なものであるが、学校教育における学習者の漢字および語彙の理解の測定にも適している。これは項目作成段階で、各学年対応の冊子として作成されたためであるが、同時に各項目については、項目反応理論に基づき、識別力、困難度が算出され、異なる項目で得られた結果でも比較可能となっている(詳細は佐藤・伊藤・倉元、2003)。また項目内容も多岐に渡っており、内容的にも多面的な評価が可能である。

項目反応理論に基づいて作成された本テストは、平均点や偏差値の比較を行う従来型の学力の経年変

化測定とは異なり、潜在特性の推定によって日本語の基礎能力の測定を行うものである。テストの運用面から考えた場合、母集団全体の能力分布に依存する偏差値を用いた比較等と異なり、そうした制約から自由な点が特徴である。その点で学校教育における児童、生徒の漢字・語彙力の経年変化を測定する上では極めて有効なものであると言える。

ただし、その運用に際しては留意すべき点もある。それは、言語生活を取り巻く諸要素の通時的変化等により、漢字および語彙項目が使用頻度や機会の変化で変化している可能性があるという点である。

具体的には、コミュニケーション・メディアの多様化や、学校教育現場でのカリキュラムの変更・教材の違いの影響などが考えられるだろう。さらに、通時的変化と関連して、一般に伝統的な語彙とされるものであっても、理解と使用に明確な世代差がみられるものも多くあるはずである。国語教育の観点から国語の力の経年変化を想定した場合、特に漢字、語彙という測定対象にとってそうした影響の大きさは十分に想定されることである。

さらに国語教育全体の底流となり、そのカリキュラムに影響を与える大きなものとして、国語審議会答申、学習指導要領等に加え、一般あるいは様々な専門領域からの「日本語」論からの影響力も無視できない。最近では特に「日本語力」をめぐる論議が盛んであるが、それも今後の国語教育の目的や方法、実践面において、多方面にわたる変化の動因となるものであろう。

以上の点から、日本語基礎能力をみた場合、その「基礎」は、日常の使用実態にもとめられる面が大きく、語彙能力測定という目的に基づく個々の語彙項目に関する評価・判断においては、現状を多面的に考慮しながら行う必要がある。

これらもふまえ、項目に関する評価基準とその実際を提示する。

#### 4. 語彙項目評価基準

語彙項目テストは、選択肢式であり、提示した表現について「意味の似ているもので一番良いと思ったもの」を5つの選択肢から選ぶ形をとっている。こうした形式の場合、被験者の「意味の似ている」という判断は様々な要因に基づいて行われる可能性が高いと言える。具体的な発話形成以前の個々の語

の持つ「語義」であったり、それに連合した言語外要因も含んだ「内容」である場合もある。設問が基本的に文形式をとっている以上、「内容」に加え、発話意図と関連した「意義」や設問文の表現全体の持つ「含蓄」「ニュアンス」等も判断要因に含まれている可能性が高いと言えるだろう。語彙項目について個々に検討する場合、測定対象の語彙を含む設問文全体と判断に関わる選択肢構成も考慮に入れながら行う必要があるのである。

今回はこれらをふまえながら、観点と基準を設定、各項目に関して評価を行っている。評価基準の作成と実際の判定に関しては、国立国語研究所(1989)、文化庁(1997～2002)等を参考にした。

以下に評価の観点を示す。

##### (1) レベル A

レベル A は、以下の観点と基準にしたがった。

基準 1：意義に加え、使用機会や場面も通時的に安定しており、今後においても安定した継続使用がなされていくと思われるもの。

基準 2：上記「基準 1」により、各種のテストその他でも安定して取り上げられる機会が多いと思われるもの。

「レベル A」は、通時的な安定性を主基準とし、あわせて類義語との弁別の特徴等からも各種テストにおいて理解の適切性が問われる機会が多いと思われるもの、という観点から下位分類したものである。実使用面において、またそれに関する社会的な要求性の高さという点において安定性の高い語彙項目であると言える。

##### (2) レベル B

レベル B は、実使用面を中心とした項目の安定性においてレベル A に準ずるものである。項目自体の評価に加え、設問に関わる選択肢構造の観点からも下位分類を行っている。以下にその基準を示す。

基準 1：慣用句等の固定的な表現であり、社会生活上の使用を想定した就職・公務員試験等の各種テストでは頻出であるが、実際の日常的使用頻度は比較的低いと思われるもの。

基準 2：使用頻度において減少傾向にあると思われるもの、あるいは、意味・使用機会において通時的変化がみられるもの、関連して使用に世代差がある(若年層は用いない、等)と思われるもの。

基準 3：設問の選択肢構成において、選択肢が当該語彙項目の基本義や派生義、または感情的意味等の微差におよぶものとなっており、正解の意味が各選択肢に分散されていると思われるもの。

「基準 1」は各種の試験問題として頻出のものであるが、日常生活では使用頻度が低いと思われるものである。学校教育現場での実際の試験問題の作成手順から推測した場合、過去に実施した問題のデータから重要度が高いものを選択し、それをベースにして問題作成を行っているということが考えられる。それならば問題自体は大きく変化することはない。しかも作成上の「重要度」の基準は、日本語基礎能力という観点からではなく、「試験での出題頻度」に依る場合が多く、それゆえ、問題で設定される語彙項目は、日常的語彙としての使用実態から乖離してくる傾向は否めない。「基準 1」に含まれるものは日常的な語彙でありながらもそうした傾向が見られる項目である。

「基準 2」は、使用頻度や意味等において通時的変化が見られるものである。そうした項目では、理解や使用に世代差が生じていることも考えられる。意味伝達は、話し手と聞き手の共同作業である。被験者のそれまでの人生における、その語が指す事物をめぐる個人的体験の積み重ねの内実といった経験的要素も意味解釈を大きく左右するものである。さらには設問文による場面把握に関しても、何が前景化し、何か背景化するか、といった点に関しても世代間的な違いが発生していることが想定できる。「基準 2」に含まれる項目はそうした通時的変化やそれに関連したコーホート効果が含まれている可能性が高いと思われるものである。

「基準 3」は項目自体に関するものではなく、選択肢の構成に関わるものである。語の意味作用について、明示的意味（denotation）と暗示的意味（connotation）に区別する場合があるが、「基準 3」とした項目は選択肢が明示的意味と暗示的意味の組合せから構成されており、意味の中核である明示的意味の選択が難しくなっているものであると言える。あるいは、個人や文脈等の特定性を捨象した「意義素」的な見方をとるならば、「基準 3」に該当するとした項目は、意義素として記述される、文法的特徴（品詞的特徴、統語的特徴）や語義的特徴（前提的

特徴、本来的特徴）さらには含蓄的特徴（文体的特徴、喚情的特徴、文化的特徴）が、単独であるいは複合形として各選択肢に分散されているもの、といった見方も可能であろう。語彙テストは、「一番良いと思ったもの 1 つだけ」を選択する形式で、その語の第一義的な意味作用である明示的意味の選択を要求するものであるが、「基準 3」の項目では、正解である明示的意味の選択肢に加え、それと必然的に関係する概念や連想される観念等を含んだ暗示的・副次的意味の選択肢が正解の周囲を取り巻くように構成されている。単純な誤答と正答の組合せからなる選択肢ではなく、語の多面的な意味作用が複数の選択肢に分散されているものであり、それゆえ被験者にとって判断が難しい面を含んでいると思われる項目である。

以上がレベル B に関する評価基準である。

### (3) レベル C

レベル C の項目は、以下の基準に基づいて決定している。

基準 1：使用面において、ほぼ就職・公務員試験等での出題や学校で扱う教科書作品に出現するケースのみに留まっていると思われるもの。あるいはまた、選択肢構成から正解の類推可能性が高いと思われるもの。

基準 2：選択肢構成上の問題等により、正解に至ることが難しいもの。

「基準 1」は、日常生活での使用頻度や機会の少なさ、選択肢構成上の問題から、語彙能力以外の判断要因の働きも大きいと考えられるものである。選択肢間のタイプ（品詞等）の異なりや選択制限（共起制限）等の情報から正解とその他の選択肢の選別が可能な場合、などが挙げられる。

また「基準 2」は、「レベル B：基準 3」で提示した選択肢構成が難化したものであり、語彙理解力だけでは正解に至ることが困難であると思われるものである。「レベル C」は、正解の判断にあたって、語彙理解力以外の様々な要素も含むものであると言える。

以下、各項目について、上記のレベル、基準に基づく評価を行う。

## 5. 項目評価結果

### 5.1. 項目プール全体の評価結果

はじめに項目プール全体の評価結果を提示する。

項目プール全体の比率としては、「レベル A」に該当する項目が 71.2%、「レベル B」が 23.2%、残りが「レベル C」の項目となった。日常的使用を想定した「レベル A」の比率が圧倒的に高く、日常的な使用頻度は小さいか、世代差や選択肢構成といった観点を含む「レベル B」に該当する項目は四分の一程度である。また、語彙能力以外の判断要素が介入する「レベル C」に該当する項目は極めて少ない。項目プール全体の妥当性とバランスの良さが確認できる。なお、次節で詳述するが、表によると、版の違いにより項目比率は異なっている。そこには各版の対象とする被験者の特性、要求される語彙の質を考慮した点や、複数の版に含まれるアンカー項目の比率等が影響している。その意味においても各版とも目的とする測定能力にうまく合致したものになっていることが確認できる。

### 5.2. 各版ごとの評価結果

次に、各版ごとに評価した結果を提示する。最初に小学生 1 年生版から 4 つの高校生版まで、各基準による項目数およびその比率を示す。

各評価レベル・基準による項目比率の版比較から以下の事が確認できる。

表 各レベル、各基準に分類された項目の数

		レベル A		レベル B			レベル C		削除済項目	計
		基準 1	基準 2	基準 1	基準 2	基準 3	基準 1	基準 2		
プール全体	項目数	160	175	55	38	16	19	7	-	470
	比率	34.0%	37.2%	11.7%	8.1%	3.4%	4.0%	1.5%	-	100.0%
小 1 版	項目数	19	7	0	1	3	0	0	0	30
	比率	63.3%	23.3%	0.0%	3.3%	10.0%	0.0%	0.0%	0.0%	100.0%
小 2 版	項目数	25	18	1	1	5	0	0	0	50
	比率	50.0%	36.0%	2.0%	2.0%	10.0%	0.0%	0.0%	0.0%	100.0%
小 3 版	項目数	28	25	1	1	4	0	1	0	60
	比率	46.7%	41.7%	1.7%	1.7%	6.7%	0.0%	1.7%	0.0%	100.0%
小 4 版	項目数	31	25	0	1	3	0	0	0	60
	比率	51.7%	41.7%	0.0%	1.7%	5.0%	0.0%	0.0%	0.0%	100.0%
小 5 版	項目数	33	18	0	3	5	0	0	1	60
	比率	55.0%	30.0%	0.0%	5.0%	8.3%	0.0%	0.0%	1.7%	100.0%
小 6 版	項目数	24	17	1	8	5	0	2	3	60
	比率	40.0%	28.3%	1.7%	13.3%	8.3%	0.0%	3.3%	5.0%	100.0%
中学生版	項目数	31	27	2	6	4	1	4	5	80
	比率	38.8%	33.8%	2.5%	7.5%	5.0%	1.3%	5.0%	6.3%	100.0%
高校生 91 年度版	項目数	14	24	14	7	0	2	1	11	73
	比率	19.2%	32.9%	19.2%	9.6%	0.0%	2.7%	1.4%	15.1%	100.0%
高校生 92 年 A 版	項目数	7	35	16	8	0	7	0	2	75
	比率	9.3%	46.7%	21.3%	10.7%	0.0%	9.3%	0.0%	2.7%	100.0%
高校生 92 年 B 版	項目数	11	29	21	6	0	6	0	2	75
	比率	14.7%	38.7%	28.0%	8.0%	0.0%	8.0%	0.0%	2.7%	100.0%
高校生 92 年 C 版	項目数	19	36	8	7	0	4	1	0	75
	比率	25.3%	48.0%	10.7%	9.3%	0.0%	5.3%	1.3%	0.0%	100.0%

小学生版では、小 6 版を除いて「レベル A」に含まれる項目の比率が、80%の後半から 90%の前半に収まっており、項目比率の安定性が確認できる。小 6 版では「レベル A」の比率が 68.3%、「レベル B：基準 2」の比率が 13.3%と、他学年の版とは異なった特性を示しているが、これは小 6 版の項目が中学生版との重なりが多いことと関連している。項目数と比率の概観からは、小学生版では各項目がレベルごとにバランスのとれた形で配分されていることが確認できるはずである。

次に中学生版であるが、項目バランスは小 6 版と似通った傾向を示している。中学生版も高校生版との共通項目が多く含まれており、必然的にレベル B に該当する項目が多くなっている。

最後に高校生版であるが、高校生版は 91 年版と 92 年 A ~ C 版の計 4 版作成されている。91 年版は、中学生版同様、高校生全体に対して 1 種類であるが、92 年版は難度の高いものから A、B、C の順に 3 種類存在する。ここでは、92 年版の各版の項目レベルを確認する。

A 版は、主として進学校での使用を想定したものであるが、このことは各レベルの項目比率にも反映している。「レベル A：基準 2」の比率が C 版と並ん

で高いが、A版は、使用面において通時的にも安定し、なおかつ質の高い言語使用能力の要求という面も満たした語彙項目を中心に構成されている。

B版は、A版と比較して「レベルA：基準2」の項目比率が低いが、「基準1」の比率はA版と比較して高い。「レベルA全体」での両版の項目比率は、A版、B版それぞれ55.0%、53.4%であり、ほぼ同じであると言っていいだろう。B版は、使用頻度が高く、通時的に安定していると思われる項目（「レベルA：基準1」に該当）に加え、将来の社会生活の使用を想定した就職・公務員試験等の試験で出題される機会が多い「レベルB：基準1」の項目比率が高い。B版は、総じて日常性の高い項目が中心となっていると言える。

さらにC版であるが、C版はA版、B版と比較して「レベルA全体」の項目比率が73.3%と圧倒的に高く、使用に世代差がみられる場合がある「レベルB」に該当する項目の比率は低い。C版は使用においてB版以上に極めて日常性の高い項目を中心に構成されており、日常的・基本的言語使用能力の測定という目的に最も合致したものとなっていると言える。

以上が、高校生用92年版の特徴である。なおここで特にふれなかった91年版に関しては、「レベルA

およびB」の項目比率からみると、B版とC版の間に近いものになっていることが確認できるはずである。

### 5.3. 項目パラメタ値による評価結果

ここでは、語彙項目の識別力パラメタ  $a$ 、困難度パラメタ  $b$  に関して、各評価基準との関連性を探り、各版比較を行う。

項目パラメタ値が版によって著しく異なるため、項目パラメタは項目プール作成過程においてそれぞれの版を受験した解答者の特性値の分布で標準化した値を用いる。すなわち、 $x$ 版において標準化された項目パラメタの値は、

$$Z_{a(x)} = a / a_{g(x)} \quad (1)$$

$$Z_{b(x)} = (b - \bar{\theta}_{(x)}) / S_{\theta(x)} \quad (2)$$

ただし、 $Z_{a(x)}$  は  $x$ 版で標準化された項目パラメタ  $a$  の値、 $a_{g(x)}$  は  $x$ 版に含まれる項目パラメタ  $a$  の幾何平均、 $Z_{b(x)}$  は  $x$ 版の被験者パラメタの推定値の平均、分散によって標準化された項目パラメタ  $b$  の値、 $\bar{\theta}_{(x)}$  は、 $x$ 版を受験した被験者の  $\theta$  推定値の平均、 $S_{\theta(x)}$

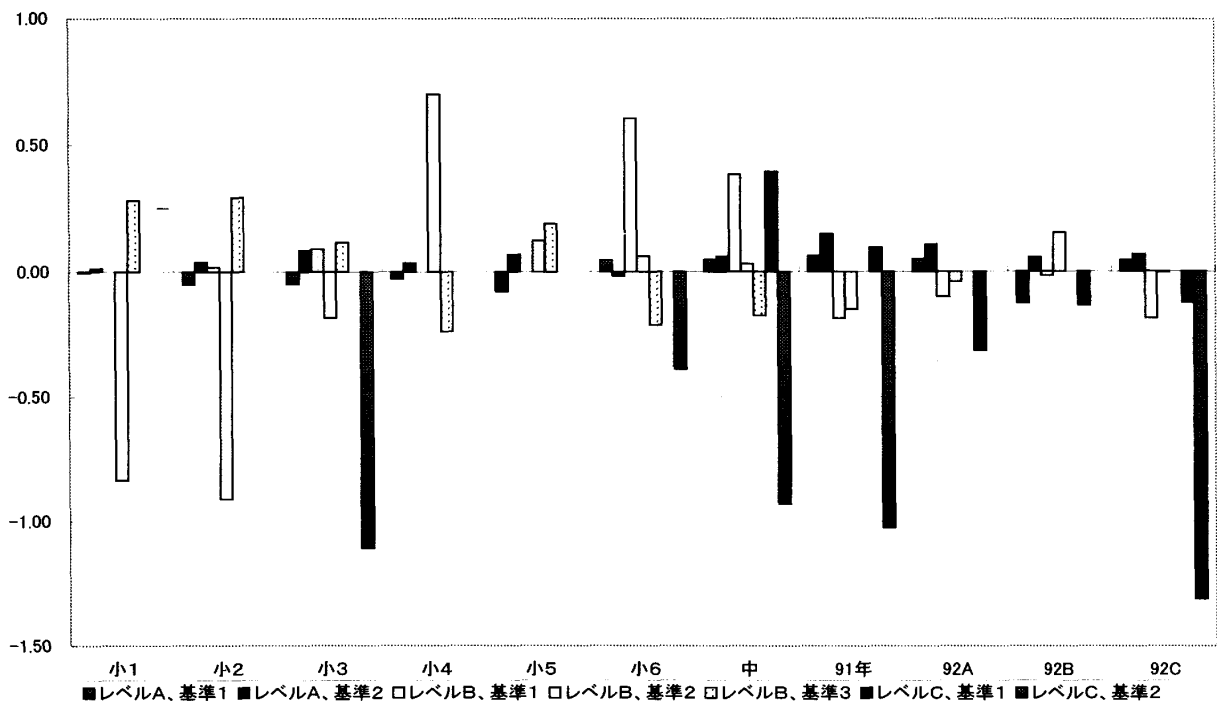


図1. 項目パラメタ  $a$  の幾何平均値に対する比の値（自然対数の平均値）

は、その標準偏差を表す。

なお図1は、 $Z_{a(x)}$ の自然対数を取ってその平均値を表示したものである。

まず顕著な特徴として、各版を通じて「レベルC：基準2」の値が極めて低いことが確認できる。「レベルC：基準2」は、選択肢構成上の問題等により、語彙理解能力だけでは正解に至ることが困難かもしれないと推測されるものである。被験者の判断には、語彙能力以外の要素も働いていることが考えられ、項目反応理論のモデルに対するフィットが良くないと推測できる。それゆえ、識別力において他のレベルおよび基準と比較して極端に低くなっているのだと言えよう。また、小1版、小2版では「レベルB：基準2」の値が低く、「レベルB：基準3」の値は高い。「基準2」は日常的使用に世代差や通時の変化がみられるものであり、「基準3」は選択肢構成から正解を導きにくいものである。小学校入学後、新たな語彙を学習する子どもたちにとって、使用に世代差等がある項目は、能力差を敏感に反映しないことが考えられ、結果的に識別力が低くなっていると推測できる。また、各選択肢に正解の意味が分散されている「レベルB：基準3」の値の高さは、語彙理解の正確

性という意味において能力差を強く反映したものになっていると言える。

その他、小6版および中学校版では「レベルB：基準1」の値が高いが、小6から中学校にかけては日常的使用におさまらない新たな語彙が学習内容として多く入ってくる。そのため、そうした語彙の理解は被験者の能力を如実に反映したものとなるのだと言えよう。

次に項目パラメタ b (困難度) に関し、標準化した値の平均値のグラフを提示する。

まず、「レベルC：基準2」の値が低いことが、図1で確認した結果とあわせると、これにあたる項目は識別力、困難度ともに低い項目であると言える。つまり、項目自体は易しいが、選択肢構成上の問題から被験者の能力の識別性能も低く、実際の運用にあたっては注意が必要な項目ということができる。また、小1版、小2版では「レベルB：基準2」の項目の値が高く、「基準3」の項目は低い。「基準2」の項目は被験者にとって日常的な語彙ではないため、困難度が高い項目であり、他方で選択肢上の難しさを含む「基準3」の項目は、困難度自体は低いものとなっている。項目自体はさほど難しくはないが、選択肢の判断にあたって語の中心的意味を正確に把握できる

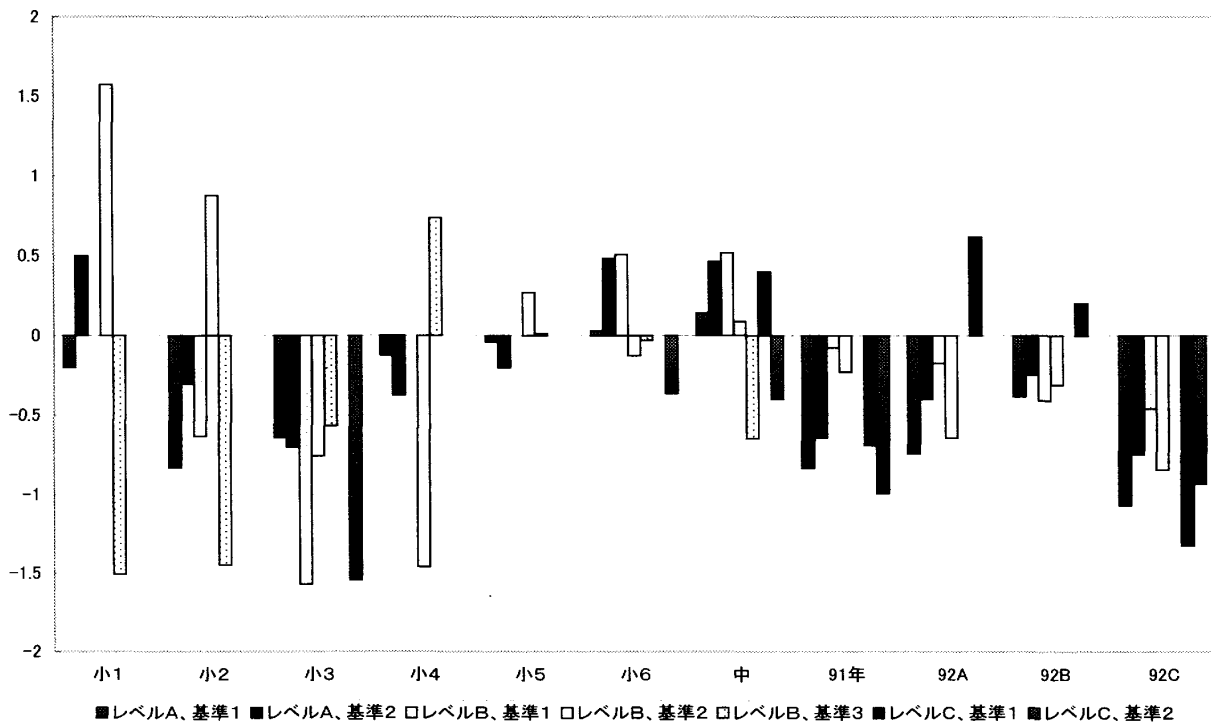


図2. 項目パラメタ b の標準化した平均値

かどうかがカギとなっているという解釈が可能であろう。

その他、小5版から中学生版においては、全体的に高い傾向を示すが、小学校高学年版では中学校版との、中学校版では高校版と重なる項目が多く含まれており、そのことが要因として大きく働いているものと思われる。なお、高校の91年版であるが、項目比率で確認した結果と同様、困難度においても92年B版とC版の中間的な位置にあることが確認できる。

以上、識別力、困難度について、各版、各レベルごとに確認した。

## 6. まとめ

以上、日本語基礎能力テストについて、主として国語教育の観点から語彙テスト項目に関する評価を中心に論じてきた。あらかじめ項目評価基準を設定し、それに基づき各項目の評価を行ったが、ほとんどの項目が「レベルA」および「レベルB」に含まれることが確認された。実使用面を考慮した場合、極めてバランスのとれたものとなっていると言える。また、「レベルC」に含まれる項目に関しては、その特殊性ゆえに使用には注意を要するが、各レベルの項目を組み合わせることにより、より多様なテスト作成が可能になることも確かである。その意味で言えば、「レベルC」の項目も語彙理解の内実を探る上で有効な場合があると思われる。

日本語基礎能力テストの持つ被験者の能力測定上の有効性と、データから得られる様々な解釈上の視

点は、得点、偏差値といった従来のテストの結果分析では見えてこない多くのものを提供してくれることは疑いない。ただ、現時点では、項目評価基準について同一次元で並べるのが困難なものや、基準自体の揺れ等の問題点もある。項目プールの充実とあわせて今後さらに検討が必要である。社会人として必要とされる言語能力の育成という近年の学校教育への要請もふまえると、日本語基礎能力テストの運用とその充実は、学校教育における学力評価に関して大きな影響を与える可能性が高いと言えるだろう。

## 文献

- 足立正治 (1996). 言語学 東京法令出版  
文化庁文化語国語課編 (1997～2002). 国語に関する世論調査 大蔵省・財務省印刷局  
国立国語研究所 (1989). 高校・中学校教科書の語彙調査 秀英出版  
国広哲弥 (1982). 意味論の方法 大修館  
文部省 (1999). 高等学校学習指導要領解説国語編 東洋館出版社  
佐藤洋之・伊藤博美・倉元直樹 (2003). 日本語基礎能力テストの特性 (2) -項目反応理論に基づく語彙理解力測定項目の分析一、教育情報学研究、1、(印刷中).  
平直樹・前川眞一・小野博・林部英雄・内田照久 (1998). 日本語基礎能力テストの項目プールの作成、大学入試センター研究紀要、28、1-12.



## Properties of the Broad-Range Japanese Fundamental Language Skills Test I: Item Contents Review of Vocabulary Subscale from the Viewpoints of First Language Education

Hiromi Ito,<sup>\*</sup> Hiroyuki Sato,<sup>\*\*</sup> Naoki T. Kuramoto<sup>\*\*\*</sup>

<sup>\*</sup> *Tohoku University Graduate School of Letters; Honjo Akita Prefectural High School*

<sup>\*\*</sup> *Education Division, Tohoku University Graduate School of Educational Informatics*

<sup>\*\*\*</sup> *Admission Research Center of Tohoku University; Education Division,  
Tohoku University Graduate School of Educational Informatics*

The purpose of this study is to reinvestigate the vocabulary items of the Broad-Range Japanese Fundamental Language Skills Test (BRJFLST) developed by Taira et al. (1998). BRJFLST consists of two parts: Vocabulary Subscale and Chinese Character Pronunciation Subscale. The present paper reviews vocabulary parts consisting of 470 items.

Items were reviewed from the viewpoints of Japanese language education as first language. Three levels of category were set for the purpose of evaluation: Level A as the ‘considerably stable in the sense of meaning and social significance over time’, Level B as ‘fairly stable in the sense of meaning and social significance over time, frequently quoted as test items’ and Level C as ‘limited use in daily life or confusing distracters’. Each level was divided into two to four categories according to other criteria mainly set from the viewpoints of usage and exposure in the testing situations.

The result showed that 71% of items were classified into level A, while only 5.6% were evaluated as level C. It verifies that the item pool is still functioning well as a measurement tool for basic skills of the Japanese language as a whole, even a decade passed since the items were created.

**Key words:** first language education, Japanese, item pool, vocabulary, measurement