

新しい日本語能力試験のための語彙表作成にむけて

押尾和美・秋元美晴・武田明子・阿部洋子
高梨美穂・柳澤好昭・岩元隆一・石毛順子

〔キーワード〕日本語能力試験、出題基準、データベース、語彙表

〔要旨〕

日本語能力試験の実施機関である国際交流基金と日本国際教育支援協会は「日本語能力試験の改善に関する検討会」を2005年に発足させたが、2009年からの新試験開始を目指して各分科会は日々調査研究を重ねている。本稿は、分科会の一つである出題基準分科会漢字表・語彙表部会が行っている語彙表作成作業の2007年9月現在の中間報告である。

部会の活動は、この2年間で4つの段階を経た。作成方針と選別方針を決定し、データベースに関する調査および整備をする第一段階、語の選別をする第二段階、語の再選別及び記述方法の検討をする第三段階、語の再々選別と初出級の検討をする第四段階である。実際にどのようなデータベースを使い、どのような検討を重ね、どのような選別作業をしたのか。本稿ではそれぞれの段階ごとに具体的な例を挙げながら報告をする。また、それとともに今後考えていかなければならない課題についても言及する。

1. はじめに

日本語能力試験の実施機関である国際交流基金と日本国際教育支援協会は「日本語能力試験の改善に関する検討会」(以下、検討会と略)を2005年に発足させ、2009年からの新試験開始を目指して日々調査・研究を重ねている⁽¹⁾。改定は多岐にわたるが、調査・研究作業は検討会の下に組織された分科会がそれぞれ分担して進めている。本稿は、分科会の一つである出題基準分科会漢字表・語彙表部会が行っている語彙表作成作業の2007年9月現在における中間報告である⁽²⁾。

2. 新しい日本語能力試験について

日本語能力試験に関する課題は、『日本語教育のための試験の改善について』((2001)日本語教育のための試験の改善に関する調査研究協力者会議)に指摘されているとおりである。これらの課題を解決すべく組織された検討会では、次の4点を新しい日本語能力試験の理念とした。

新しい日本語能力試験の構想

- ①課題遂行能力とそのためのコミュニケーション能力を測定する試験とする。
- ②can do statements (ができる) による能力基準の表示を行う。
- ③これまでの4レベルを新たに5レベルで設定し、日本語能力の多様性に対応できるように調整をする。
- ④テストの等化を行う。

新しい日本語能力試験は5つのレベルからなり、出題される問題は、課題遂行能力とそのためのコミュニケーション能力をよりよく測るものになる。そして、測定結果は、受験者の能力が同一であればいつ受けても同じ安定したものとなり、その結果とcan do statements⁽³⁾を照合することによって、受験者の潜在的な日本語能力も技能別に推測が可能になる。これらの実現化に向けて、①～③については能力基準分科会、④については分析評価分科会が調査・研究を行っている。

現行の日本語能力試験では問題作成者用に『出題基準』が作成され、1994年からは一般にも公開されているが、検討会では新しい試験に合わせた出題基準の構築も行っている。出題基準分科会は、そのために組織された分科会である。分科会はさらに2つの部会に分かれ、漢字表と語彙表は漢字表・語彙表部会(以下、部会と略)、文法項目表は機能表・文法項目表部会が担当している。

3. 語彙表作成の作業報告

部会が発足して2年になるが、7名(2006年12月からは8名)のメンバーで作業を進めてきた。以下、順を追って作業内容を報告する。

3.1 第一段階 作成方針と選別方針の決定、データベースに関する調査および整備 (2005年10月～2005年12月)

部会の始動にあたって、「課題遂行能力とそのためのコミュニケーション能力を測定する」という新しい試験の構想を受け、次の3つの方針を立てた。

- (1) 客観的かつ大規模な語彙データベースを組み合わせた複合データベースから語を選別し、初級から中上級までの一覧表を作成する。
- (2) 書き言葉だけでなく話し言葉もこれまで以上に考慮する。
- (3) 描写を豊かにするさまざまな表現も積極的に採用する。

また、選別の方針として以下の4点を定めた。

- (ア) 主に頻度を重視して採否を決める。
- (イ) 機械的に頻度の高いものから採用するのではなく、日本語教育経験者の視点も加える。
- (ウ) 現行試験の『出題基準』語彙表も参考にする。
- (エ) 最終的な語数は、日本人成人の獲得語数などを参考にした上で決定する。

選別作業にあたっては、現行試験の『出題基準』語彙表（以下、現行語彙表と略）を作成した当時と同じく頻度を重視して選別を決め、分野の偏り、話し言葉と書き言葉の違いなども考慮する必要から、日本語教育経験者の視点も加味することを原則とする。そして、現行語彙表が日本語能力試験用に作成された語彙表という枠を超えて多くの教材にも影響を与えている点も考慮し、これも参考にしながら進めることとする。そして、最終的な語数は、選別作業を進める一方、慎重を期して、日本人成人の獲得語数などを参考にした上で決定することとする。

現行語彙表を参考にしつつ、当時と同じ手法で選別作業をするのであれば、結果的にあまり違いのないものになりそうだが、新しい語彙表の大きな特徴を強調しておきたい。それは、現行語彙表は、上位級と下位級で異なる語彙資料を使用し、下位級は4級 3級、上位級は1級

2級の順で級分けとともに選別を進めたのに対し、新しい語彙表は、新たにこの作業のために作成した複合データベースを使用して語を選び出した後、級分けをする、という手順を取る点である。この方法によれば、どのような基準で級分けをするか、というところが大きなポイントとなるが、完成した語彙表は、上位級と下位級で性質の差のない一貫性を持ったものになる。

これらの方針を決めた上で、国立国語研究所、国立教育政策研究所を訪問するなどして協力を仰ぎ、情報を収集し、複合データベースのもととなるデータベースを決めた。使用したデータベースは表1のとおりである。

なお、今日「コーパス」という用語を耳にすることが多くなったが、部会ではこれに相当するものを「データベース」と呼んでいる。これは、英語の場合は40万～4億語にも及ぶかなり大規模なコーパスが作成されているのに対し、日本語の場合8～9万語と小規模なものしかなく、「コーパス」と呼ぶ規模ではないと判断したためである。

表1 新しい語彙表作成のために使用したデータベース

		データベース名	出版年	出版社	便宜的な通称	
第 三 次 D	第 一 次 D B	『日本語の語彙特性第1期』CD ROM	1999	三省堂	親密度	
		『現代新国語辞典改訂第3版』CD ROM	2000	学研	新国語	
		『日本語の語彙特性第2期』CD ROM	2003	三省堂	頻度	
		『出題基準(改訂版)』	2004	凡人社	JLPT	
		『現代雑誌の語彙調査 1994年発行70誌』	2005	国立国語研究所	雑誌	
	第 二 次 D	外 来 語	『日本語の語彙特性第1期』CD ROM	1999	三省堂	親密度
			『現代新国語辞典』改訂第3版	2002	学研	新国語
			外来語認知に関する調査	2002~2004	国立国語研究所	国研
			『例解新国語辞典』第6版	2002	三省堂	例解
			『出題基準(改訂版)』	2004	凡人社	JLPT
B	オ ノ マ ト ペ	『現代国語例解辞典』第4版	2006	小学館	現国例	
		『擬音語・擬態語辞典』	1974	東京堂	天沼	
		『擬音語・擬態語辞典』	1978	角川書店	角川	
		『ぎおんごぎたいご』	1994	アルク	アルク	
		『現代擬音語擬態語用法辞典』	2002	東京堂	飛田	
B	話 し 言 葉	『出題基準(改訂版)』	2004	凡人社	JLPT	
		『現代国語例解辞典』第4版 擬音語・擬態語集成	2006	小学館	現国例	
		「男はつらいよ」48編シナリオ ⁽⁴⁾ 話し言葉資料	1969~1995 1997~2001、 2004	CASTEL/J研究会 筑波大	寅 筑波	
		『女性のことば 職場編』	1999	ひつじ書房	職場女	
		『男性のことば 職場編』 BTSによる多言語話し言葉コーパス 日本語会話 ⁽¹⁾⁽²⁾ (2003年版)	2002 2003	ひつじ書房 東京外国語大学 宇佐美まゆみ監修 東京外国語大学大学院 地域文化研究科 21世紀COEプロジェクト 「言語運用を基盤とする言語情報学拠点」	職場男 BTS	
慣 用 表 現 な ど		『小学国語学習辞典』	1994	偕成社	偕成社	
		『小学国語新辞典』第3版	2002	旺文社	旺文社	
		『くもんの学習国語辞典』第3版	2002	くもん出版	くもん	
		『例解新国語辞典』第6版	2002	三省堂	例解	
		『現代新国語辞典』改訂第3版	2002	学研	新国語	
		『例解学習国語辞典』第8版ワイド版	2004	小学館	学習国語	
		『例解小学国語辞典』第3版	2005	小学館	例解小学	

3.2 第二段階 語彙の選別（2005年12月～2007年2月）

本来ならば、あらゆる語が網羅された大規模データベースを整備したのち、選別作業を進めるのが理想的であろう。しかしながら、いくつものデータベースを組み合わせた複合データベースを作るには、完成までかなりの時間を要することは初めからわかっていた。そのため、第二段階としては、外来語、オノマトペのような小さなデータベースでも完成したものから順次検討を始め、選別を進めていくことにした。

3.2.1 外来語

外来語は新陳代謝が激しく、使用実態に個人差があるため、使われ方に差が大きい。そのため、部会では、上に挙げた辞書の見出し語として載っているか、国立国語研究所の外来語調査の対象項目として挙げられているか、現行語彙表に記載されているかどうかで出現度を集計し、判断の手掛かりとした。

表2 外来語選別のために作成した表（一部）

	見出し語	JLPT	例解	現国例	学研	親密度	国研	出現度
5881	フローリング							3
5882	プログラマー							4
5883	プログラミング							4
5884	プログラム							6
5885	プログラム言語							1
5886	プロジェクター							2
5887	プロジェクト							5

は見出し語にあったことを表す。

選別を終えた結果は以下の通りである。

表3 外来語の選別結果

	総数	採用数	具 体 例
出現度 6	14語	14語	アプローチ、テーマ、ストレス
出現度 5	503語	472語	アイデンティティー、コンピューター、シンポジウム
出現度 4	1,775語	823語	アニメ、クレジット、シャンプー
出現度 3	1,597語	170語	アンダーライン、コイン、パートタイム
出現度 2	1,731語	78語	アメリカ、エアコン、ガールフレンド
出現度 1	2,119語	45語	アフリカ、ハンドアウト、マークシート
合計	7,739語	1,602語	

3.2.2 オノマトペ

オノマトペは、外界で発せられる声と音を写した言葉である擬音語と、ある動きや状態などを音によって象徴的に表す言葉である擬態語、また、擬情語と呼ばれる人の心の状態を表す言葉を総称した語である。ここではオノマトペの辞典3冊、『現国例』の巻末付録、アルクの学習者用参考書、現行語彙表を活用することにし、これらの記載状況をまとめることで出現度を出した。

表4 オノマトペ選別のために作成した表(一部)

通し番号	見出し語	飛田	角川	天沼	現国例	アルク	JLPT	出現度
4	あたふた							4
203	ずらずら							4
414	ごくんごくん							1
993	にこにこ							5

出現度5のものはあまりなじみがないものを除いてかなりの数が採用されたが、出現度4以下のものは、どのくらい一般的で、なじみがあるかなどを基準として部会で議論を重ね、選別を行った。結果は次の通りである。

表5 オノマトペの選別結果

	総数	採用数	具体例
出現度5	118語	107語	すやすや、にこにこ
出現度4	392語	230語	あたふた、さらさら
出現度3	319語	7語	きっかり、ちやほや
出現度2	315語	2語	あやふや ずらりと
出現度1	576語	0語	
合計	1,720語	346語	

3.2.3 話し言葉

話し言葉のデータベースとしては、国立国語研究所の「日本語話し言葉コーパス」が有名である。部会でもこのコーパスの利用を検討したが、学会講演、模擬講演、インタビューといった比較的改まった場面でのモノログが多いこと、語よりも音韻や音声の分析を目的にデータが収集されていることを理由に、今回は話し言葉データベースには含めなかった。それに代わり、映画のシナリオと実際の話し言葉を録音して文字化したデータを利用することにした。データベース作成にあたっては、当初は文書解析システム「茶碗」を利用したが、パソコンと

の相性が悪く期待した結果が得られなかったため、その代わりに「word excalibur」を用いて頻度集計を試みた。

同ソフトも、「茶碗」と同じく形態素解析システム「茶筌」に対応しているが、「茶筌」には話し言葉特有の言い回しや表現は登録されていないため、精度の高い頻度集計は難しかった。そのため、最終的には比較的長い単位で発話を解析して集計し、その中から話し言葉特有と思われる語を部会でピックアップし、採否を決めるという手段を取ることにした。話し言葉のデータベースの場合、集めた語や表現は通し番号にして57,000弱になんなんとしたが、この中には書き言葉もかなり含まれていたため、人間の目による抽出が不可欠であった。このような方法で選別を行った場合、採用語、不採用語の数を結果として挙げて意味があるとは思えない。そのため、ここでは話し言葉選別のために作成した表を表6で、そして、データベースから抽出された話し言葉特有と思われる語を表7で報告する。

表6 話し言葉選別のために作成した表(一部)

通し番号	見出し語	頻度
6301	うん	7,967
44679	はい	2,723
40670	なんか	2,605
5	あー	2,148
1214	あの	1,675
5596	うーん	1,429
36162	でも	1,306
37744	とか	1,266
6534	えー	1,216
1233	あのー	1,185

表7 話し言葉特有として抽出された語(一部)

見出し語	頻度
あのね	124
あっち	40
バイバイ	25
そのまんま	16
でっかい	13
よいしょ	12
いまいち	5
しょっぱい	4
かったるい	2
わあ	2
っぼっち	2

表7を見ると、話し言葉特有と判断した語は、話す際にしか使われないもの、書く、話すどちらにも使われるが、話す際に音に変化が生じるものに大別されることがわかる。話し言葉は日本語学習者が使い方を誤ると悪い印象を与えてしまう恐れがあるため、書き言葉よりも使用に慎重を期す必要がある。話し言葉特有の語を集める過程はここまでの段階で終了したので、次の段階で日本語学習者にとって必要か否かを考慮しながら検討を進めていくことにしたい。

3.2.4 慣用表現など

慣用表現、連語、諺などは現行語彙表には記載されていないが、日本語独特の豊かな発想が

見られ、初級レベルから使ってもよい基本的なものも数多くある。そこで、新しい語彙表ではこれらをまとめた別表を作成することとし、そのためのデータベースとして、小学生用の国語辞書5冊、一般用の小型国語辞書2冊から該当する表現を集めたデータベースを作成した。

表8 慣用表現など選別のために作成した表(一部)

通し番号	キーワード		表 現	学 習 国 語	例 解 小 学	旺文社	偕成社	くもん	例 解	新国語	出現度
501405	き	気	気がいい								2
501406	き	気	気が多い								3
501407	き	気	気が大きい								1
501408	き	気	気が置けない								7
501409	き	気	気が重い								4

慣用表現に関する作業は、一度始めたが、現在は中断している。これは、あれこれ手を伸ばすよりも、まずは語彙表本表を完成させる方がよい、という判断に至ったことによる。慣用表現などを記載した表は、本表完成後作業を再開し、別表としてまとめる予定である。

3.2.5 第一次データベース

第一次データベースは、主に書き言葉をまとめたデータベースである。収集したものは表1に挙げた5冊で、「頻度」は、1985年から1998年までの14年間に発行された朝日新聞の記事を形態素分析システム「すもも」を用いて解析し、集計した頻度、「雑誌」は1994年に発行された雑誌70誌に、どのような語がどのくらい出現したかを単位で分析・集計した頻度、「親密度」は32名の被験者がある単語を見たり聞いたりした際、その語をどれほど親密に感じるかを7段階の尺度で評定した単語親密度が示されている。「新国語」は辞書掲載の有無を確認し、同音異義語の見出し語数を見るため、「JLPT」は現行出題基準の初出級を見るために欄を設けた。さらに検討を進めるにあたっては、品詞情報もあった方が、判断が容易となるため、日本語形態素分析システム「茶釜」を用いて品詞情報も付け加えた。

第一次データベースは合計約12万語あり、最も頻度が高い語は「の」で、頻度15,162,857だったが、今回も、現行語彙表の選定基準に則り、「原則として付属語(助詞、助動詞)を除く語」を検討の対象とした。作業は、メンバー全員で頻度、親密度などを確認しながら、一語一語採否を決める方法で進めたが、複合語、特定の分野でのみ使われる語、頻度1という語もかなりあったため約12万語すべてを見るまでもないと判断し、3万語(降順で頻度174まで)を超えたところで第一次データベースを用いた第二段階の作業は終了とした。

新しい日本語能力試験のための語彙表作成にむけて

表9 第一次データベース(一部)

通し番号	見出し語	表記	茶釜の切り方	品詞(茶釜のまま)	頻度	雑誌	親密度	新国語	JLPT
9407	テンゴク	天国	天国	名詞	1,376	8	6,219	1	1
9408	オクソク	憶測	憶測	名詞	1,376		5,281	1	
9409	サソイ	誘い	誘い	名詞	1,375	10	5,906	1	
9410	ユライ	由来	由来	名詞	1,375	13	5,625	1	
9411	ツチカウ	培う	培う	動詞	1,375	6	4,656	1	
9412	ショウアク	掌握	掌握	名詞	1,375		4,906	1	
9413	アマサ	甘さ	甘+さ	名詞	1,375	18	6,219	0	

新国語の欄は見出し語の有無とその数、JLPTの数字は現行語彙表の級を表す。

表10 第一次データベースの選別結果

	採用語	不採用語	合計
第一次データベース	16,769語	13,252語	30,021語

3.3 第三段階 語彙の再選別及び記述方法の検討(2007年2月~2007年7月)

第三段階として、第一次データベースに、オノマトペ、外来語、頻度173以下でも現行語彙表に記載されている語、現行試験の作題を担当している試験小委員会からの要望があった語を加え、第二次データベースを完成させた。

表11 第二次データベース(一部)

	見出し語	表記	茶釜の切り方	品詞	採=1 否=0	頻度	雑誌	親密度	新国語	JLPT
19422	サックス	サックス	サックス	名詞	1	411	31	5,938	1	
302254	サックス	サックス		名詞	0				1	
27716	ザックバラ	ざっくばらん		ナ形容詞	1	218		3,188	1	
200752	ざっくばらん	ざっくばらん		ナ形容詞	0				1	
18823	サッサト	さっさと	さっさと	副詞	1	435	8	5,844	1	2
200124	さっさと	さっさと		副詞	1				1	
11321	サッシ	冊子	冊子	名詞	1	1,024	11	4,938	1	
1522	ザッシ	雑誌	雑誌	名詞	1	14,515	96	6,25	1	4

1~30050は第一次DB、30051~は頻度173以下でも必要と判断した語、200000~はオノマトペ、300000~は外来語、400000~は略号から来た語であることを表す。

第二次データベースは、出所が異なるデータベースが一つにまとまっているため、番号で出

所の区別をした。例えば、表11の 19422と 302254に「サックス」があるが、前者は第一次データベース、後者は外来語のデータベースから来たことを意味する。「サックス」は、採否欄を見ると、第一次データベースでは採用、外来語では不採用となっており、2回の検討結果が異なっている。第三段階の作業としては、このように重複した語の再検討をし、最終的に語を絞り込む作業を行った。その結果、採用語は約1,000語削減され、次のようになった。

表12 第二次データベースの選別結果

	採用語	不採用語	合計
第一次データベース	16,769語	13,252語	30,021語
第二次データベース	15,703語	24,669語	40,372語

この作業とあわせて、新しい語彙表が完成形に近づくよう、記述方法についての検討も進めた。具体的には、「接頭語/接尾語」「造語成分」の区別がわかるよう各語の前や後に「～」をつける、同音異義語はそれぞれを別見出しにする、省略語は意味を明確にするため、省略前の語を〔 〕内に表示する、などである。

3.4 第四段階 初出級の検討、語の再選別(2007年8月～9月現在)

第三段階で更新された第二次データベースに、別途選別作業を進めていた話し言葉のデータベースからの採用語を追加し、第三次データベースを作成した。

表13 第三次データベース(一部)

	見出し語	表記	JLPT	品詞	語種	頻度	雑誌	親密度
5578	サッキウ/ソウキュウ	早急	1	名詞	漢	2,783		4,625
5579	サッキョク	作曲	2	名詞	漢	3,168	37	6,25
5580	サッキン	殺菌		名詞	漢	568	12	5,969
5581	ザックバラン	ざっくばらん		ナ形容詞	和	218		3,188
5582	サッコン	昨今		名詞	漢	2,069	8	5,375
5583	サッサト	さっさと	2	副詞	和	435	8	5,844
5584	サッシ	冊子		名詞	漢	1,024	11	4,938
5585	ザッシ	雑誌	4	名詞	漢	14,515	96	6,25

現在はこの第三次データベースを用いて、選別結果の再々検討と級分けの作業を進めている。作業はまだ継続中のため最終的な語数は確定していないが、ア行は以下ようになった。

表14 級別ア行の採用語と具体例（ア行のみ）

級 ⁽⁵⁾	採用語数	具 体 例
1 級	761	アイデンティティー、一時、右往左往、栄光、大わらわ
2 級	507	愛、癒やす、うたた寝、エアロビクス、お前
2.5級	308	あっという間、いじめ、うらやましい、駅ビル、オープン
3 級	202	アンケート、一時停止、売れる、ATM、重たい
4 級	152	朝ごはん、いろいろ、うどん、エアコン、大雨

4 . 今後の検討事項

4.1 最終的な語数の決定

4.1.1 先行研究

現行語彙表には、1 級8,009語、2 級5,035語、3 級1,409語、4 級728語の語彙が記載されているが、新しい語彙表には何語載せるのが適当だろうか。部会では過去の研究を調査してみた。

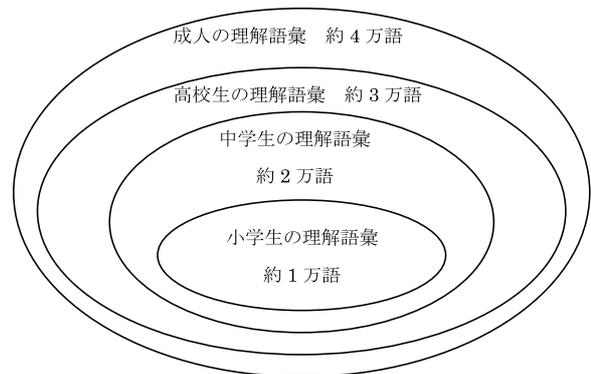
日本人の獲得語数に関する調査は難しく、近年新しい報告はなされていないようである。そのため、判断を過去の調査に頼るしかないが、阪本（1984）の調査では、小学校から中

学校にかけての義務教育の9年間に理解させるべき単語は、小学校6年間で約10,000語、中学校3年間で約10,000語、合計で20,000語という結果が出ている。

また、森岡（1951）は義務教育終了者（高校1年生）がどれくらい理解語彙を持っているかを調査したが、それによると、最高36,000語、最低23,000語、平均30,000語、被験者全員が知っていた語は12,000語ということである。更に日本人の一般成人の理解語彙について見ると、林（1974：149）は、「阪本氏や森岡氏の調査から、日本人の成人の理解語彙量は大体四万語程度であろうと推測される」と述べている。これらをまとめると図1のようになる。

次に、雑誌やテレビといった特定のメディアを対象とした語彙調査についても見てみる。国立国語研究所（1984）の『現代雑誌90種の用語用字Ⅰ』によると、90種全体でその順位までの見出し語が延べ語数のどれくらいの割合を占めるかを調べたところ、上位10,000語までで、91.7%をカバーしたという。この結果を受けて、玉村（2002：51）は、日本語の基本語は「12,000語ぐらいは必要であろう」と述べている。テレビについても、同じく国立国語研究所（1999）の調査だが、『テレビ放送の語彙調査Ⅲ』によると、話し言葉（テレビ音声）の場合

図1 年代別日本人の理解語彙



は上位17,422語、書き言葉(雑誌90種)の場合は上位39,995語を知っていれば、対象となった媒体に現れた語を100%カバーするという結果が出ている。

最後に、辞書の見出し語に注目してみる。甲斐(1986:3)は「小学生用の各国語辞典は平均して約25,000語の見出しを持っている」と述べている。また、一般的に、小型国語辞典の見出し語数は60,000~80,000語と言われている。これらの数字は、対象となっている使用者が持つ上限に近い語数と考えられる。

4.1.2 部会の結論

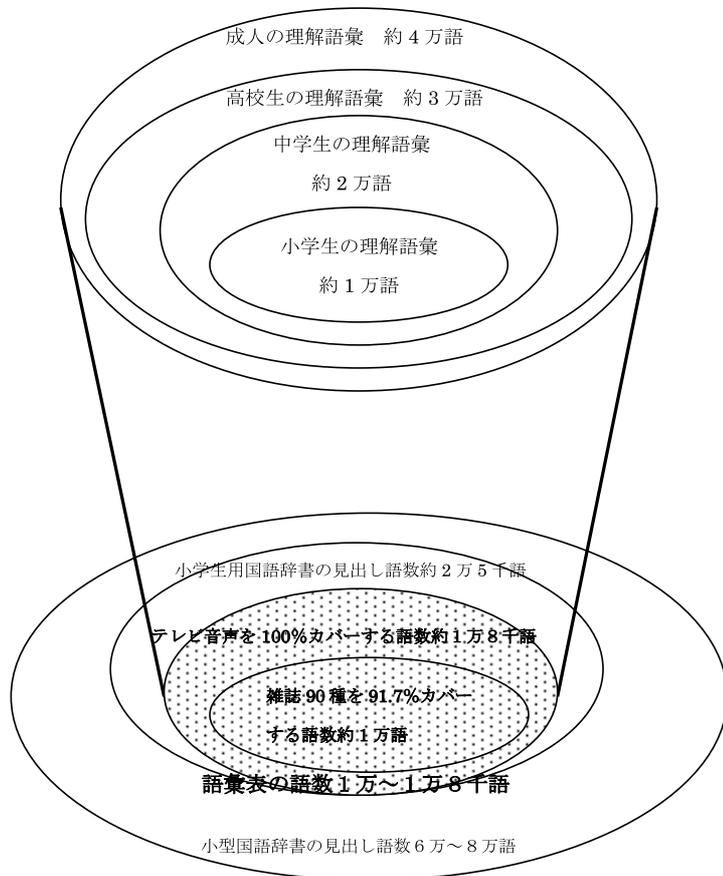
以上の先行研究の成果を考え合わせた結果、部会では次のような結論を出した。

日本語学習者が、成人日本人の理解語彙約40,000語を覚える必要はない。しかし、日本語で書かれたものを読み、話されたことを聞いて理解するには、約10,000~18,000の語が必要である。そこで、選定の基軸は、日本語で書かれた書物を読み、話された言葉を聞いて理解するに足る語数、つまり、約10,000語~18,000語をおおよその目安とする。

語が選ばれる範囲および語彙表が最終的にカバーする範囲を図示すると、図2のようになる。

部会ではこの数字を目安とし、作業を進めているが、この数字は最大で現行試験の語彙表の約2倍にもなるため、受験者の負担が増えるということは容易に想像できる。しかしその一方で、加藤(1990:107)は「日本の小学校の教科書には約1万語があらわれ、高校生が理解しうる語彙量は平均で約3万語であるところをみても、外国人に対する日本語教育においては、ほぼこの程度の語彙の習得を考えれば十分であることがわかる。」と述べており、また、実際に留学生を指導している教師たちも、日本で大学生生活を送るには1

図2 新しい語彙表の語数(あみかけ部分)



級の語数約1万では到底たりない、ということを実感している。日本語を母語としない人々に、どれほどの語数、どれほどの深い理解を求めるのか。日本語能力試験の場合、最も上のレベルをどの程度に設定するのかわからずと決まってくるが、最終的な語数は、広く日本語の基本語彙と基礎語彙、使用語彙と理解語彙を考える上でも大きな影響力を持つ数字となるであろう。

4.2 級別の語数

4.1では、総語数についての先行研究および部会の結論を述べたが、それでは、各級の語数はどのように考えればよいのだろうか。あらかじめ各級の語数を決め、その数に達するまで選別を続ければよいのか、それとも選別した全ての語に初出の級を設定し、その結果を各級の語数とすればよいのか。

林(1982)は、日本人の頭の中の語彙を(1)文法機能語(助詞、接続詞、助動詞、感動詞)、(2)準文法機能語(形式名詞、5つの動詞)、(3)機能的接辞(接頭語、接尾語)、(4)思考基本語(概念を表す名詞、疑問詞、判断を表す動詞・形容詞・形容動詞・副詞など数百語)、(5)叙事基本語(生活場面で用いる語数千語)、(6)方面別基本語(例えば中学高校の教科書別に現れる基本語、メディアから雑然と入ってくる語数千語)、(7)方面別発展語(普通の人が生活するには知らなくても困らない専門用語)の7つのグループに分けているが、これらを級別にどこまでどうカバーすれば日本語学習者にとって十分な語彙表になるのか。日本語教育の世界に大きな影響力を持つ語彙表となるだけに、can do statementsなどを参考にしつつ、広く意見を集め、反映させていきたいと考えている。

図3 林(1982)による語彙の分類



「日常語・専門語および表現語」
『講座日本語学Ⅰ総論、41、明治書院』

4.3 完成した語彙表のイメージ

完成した語彙表は、表15のようにイメージされる。

現行語彙表は、級ごとに用意されており、見出し語がひらがな50音順で、縦に並び、見出し語の横にその意味を表す漢字が添えられた一覧表の形式を取っている。また、漢字については、一文字ずつ50音順で、縦に並んだ一覧表が別に用意されている。この方式は、その級で出題される範囲をざっと見渡すには使い勝手がいいと言える。しかし、問題作成のために使うには、ある語とそれを書き表す漢字は必ずしも同じ級で出題できるとは限らないため、問題作成者は、語を選んだら今度はその漢字級を調べるといった二度手間を取らざるを得なかった。このような実情を改めるべく、新しい語彙表は、見出し語を50音順に縦に並べ、横に、表記・品詞情報・

表15 完成した語彙表のイメージ

見出し語	表記	品詞情報	語種 (読み)	初出級	漢字級	備考 ()内数字は級を表す
あまい	甘い	イ形	和	4	3	~チョコレート(4)
いらいら	いらいら	副	和	2		
かえる	変える	動Ⅱ	和	4	3	
スマート	スマート	ナ形	外	2		痩せて、~になった(2)
しょくば	職場	名	(重)	3	2 3	
でんしメール	電子メール	名	混	3	3 4	
とても	とても	副	和	4		~ + 肯定(4)
どりよく	努力	名	漢	2	2 3	
どりよく	努力	動Ⅲ	漢	2	2 3	

語種・初出級(テストに出題できる最も下の級)・漢字級・問題作成時に必要な補足的な情報を記載する備考欄を設けることとした。この形式であれば情報が一か所にまとめられているため、問題作成に費やす時間が短縮できるはずである。新しい試験実施を2年後に控え、完成作業は時間との戦いだが、精度を犠牲にすることはできない。21世紀に誇れる語彙表の完成に向け、部会では手を休めることなく連日作業を進めている。

〔注〕

- (1) 詳しくは「日本語能力試験の改定中間報告」2007年度日本語教育学会春季大会参照のこと。
- (2) 本稿は常に進んでいる作業の中間報告である。よって、今後語数や級分けに変更が生じる可能性を含むことをご了承いただきたい。
- (3) can do statementsの試行版については、2006年に、第5回日本語OPIシンポジウム(ベルリン)にて野口裕之、熊谷龍一、大隅敦子、石毛順子、長沼君主が「日本語能力試験can do statements(試行版)のIRT尺度化と日本語能力試験の得点段階との対応付けの試み」というタイトルでポスター発表を、大隅敦子、野口裕之、熊谷龍一、石毛順子、長沼君主、和田晃子、伊東祐郎が「日本語能力試験can do statementsとCEFR Dialangとの対応付けの試み」というタイトルで口頭発表をしている。また、長沼君主、大隅敦子、和田晃子、伊東祐郎、熊谷龍一、野口裕之(2007)も参照のこと。
- (4) シナリオ48編の内訳は次の通り。本シナリオは国立教育政策研究所日本語教育支援システム研究会(CAS-TEL/J)が開発した日本語教育用データベースCD-ROM(1998)に収録されている。
山田洋次・森崎東(1968)『男はつらいよ』松竹株式会社
山田洋次・小林俊一・宮崎晃(1969)『続・男はつらいよ』、同(1970)『男はつらいよ・フーテンの寅』松竹株式会社
山田洋次・宮崎晃(1970)『新・男はつらいよ』、同(1970)『男はつらいよ・望郷篇』、同(1971)『男はつらいよ・純情篇』松竹株式会社

新しい日本語能力試験のための語彙表作成にむけて

山田洋次・朝間義隆(1971)『男はつらいよ・奮闘篇』、同(1971)『男はつらいよ・寅次郎恋歌』、同(1972)『男はつらいよ・柴又慕情』松竹株式会社
山田洋次(1972)『男はつらいよ・寅次郎夢枕』松竹株式会社
山田洋次・朝間義隆・宮崎晃(1973)『男はつらいよ・寅次郎忘れな草』松竹株式会社
山田洋次・朝間義隆(1973)『男はつらいよ・私の寅さん』、同(1974)『男はつらいよ・寅次郎恋やつれ』、同(1974)『男はつらいよ・寅次郎子守唄』、同(1975)『男はつらいよ・寅次郎相合い傘』、同(1975)『男はつらいよ・葛飾立志篇』、同(1976)『男はつらいよ・寅次郎夕焼け小焼け』、同(1976)『男はつらいよ・寅次郎純情詩集』、同(1977)『男はつらいよ・寅次郎と殿様』、同(1977)『男はつらいよ・寅次郎頑張り!』、同(1978)『男はつらいよ・寅次郎わが道をゆく』、同(1978)『男はつらいよ・噂の寅次郎』、同(1979)『男はつらいよ・翔んでる寅次郎』、同(1979)『男はつらいよ・寅次郎春の夢』、同(1980)『男はつらいよ・寅次郎ハイビスカスの花』、同(1980)『男はつらいよ・寅次郎かもめ歌』、同(1981)『男はつらいよ・浪花の恋の寅次郎』、同(1981)『男はつらいよ・寅次郎紙風船』、同(1982)『男はつらいよ・寅次郎あじさいの恋』、同(1982)『男はつらいよ・花も嵐も寅次郎』、同(1983)『男はつらいよ・旅と女と寅次郎』、同(1983)『男はつらいよ・口笛を吹く寅次郎』、同(1984)『男はつらいよ・夜霧にむせぶ寅次郎』、同(1984)『男はつらいよ・寅次郎真実一路』、同(1985)『男はつらいよ・寅次郎恋愛塾』、同(1985)『男はつらいよ・柴又より愛をこめて』、同(1986)『男はつらいよ・幸福の青い鳥』、同(1987)『男はつらいよ・知床慕情』、同(1987)『男はつらいよ・寅次郎物語』、同(1988)『男はつらいよ・寅次郎サラダ記念日』、同(1989)『男はつらいよ・寅次郎心の旅路』、同(1989)『男はつらいよ・ぼくの伯父さん』、同(1990)『男はつらいよ・寅次郎の休日』、同(1991)『男はつらいよ・寅次郎の告白』、同(1992)『男はつらいよ・寅次郎の青春』、同(1993)『男はつらいよ・寅次郎の縁談』、同(1994)『男はつらいよ・拝啓 車寅次郎様』、同(1995)『男はつらいよ・寅次郎紅の花』松竹株式会社

⁽⁵⁾ 級の名称は部会で暫定的に呼んでいるものであり、正式に決定した名称ではない。

〔参考文献〕

- 秋元美晴、阿部洋子、押尾和美、高梨美穂(2006)「日本語能力試験『出題基準』語彙リスト作成中間報告」『清華大学日本語文化国際フォーラム 予稿集』、137-138、清華大学
- 押尾和美、秋元美晴、武田明子、阿部洋子、高梨美穂、柳澤好昭、岩元隆一(2007)「日本語能力試験のための語彙表作成 中間報告」『語彙・辞書研究会第31回研究発表会』資料集、17-25
- 甲斐睦朗(1986)『語彙指導の方法(語彙表編)』光村図書
- 加藤彰彦(1990)「教育基本語」『講座日本語と日本語教育第7巻 日本語の語彙・意味(下)』、106-120、明治書院
- 国際交流基金・日本国際教育支援協会(2004)『出題基準(改訂版)』凡人社
- 国立国語研究所(1984)『現代雑誌九十種の用語用字I』秀英出版
- 国立国語研究所(1999)「高頻度語彙から見たテレビ放送語彙の特徴」『テレビ放送の語彙調査Ⅲ』、12-25、大日本図書
- 阪本一郎(1984)『新教育基本語彙』学芸図書
- 玉村文郎(2002)『NAFL Institute日本語教師養成通信講座8 日本語の語彙・意味』アルク
- 長沼君主、大隅敦子、和田晃子、伊東祐郎、熊谷龍一、野口裕之(2007)『JLPT日本語能力記述文作成の試み』日本語能力試験(JLPT) Can Do Statements試行版の分析から『2007年度日本語教育学会秋季大会予稿集』、215-218、日本語教育学会

日本語教育のための試験の改善に関する調査研究協力者会議(2001)『日本語教育のための試験の改善について』

林四郎(1974)『言語表現の構造』明治書院

林四郎(1982)「日常語・専門語および表現語」『講座日本語学1 総論』、40-58、明治書院

森岡健二(1951)「義務教育終了者に対する語彙調査の試み」『国立国語研究所年報』2、95-107、国立国語研究所