

第一言語・第二言語としての日本語語彙量と漢字変換力の測定

松下達彦 (東京大学)

佐藤尚子 (千葉大学)

笹尾洋介 (豊橋技術科学大学)

田島ますみ (中央学院大学)

橋本美香 (川崎医科大学)

1. はじめに

日本の大学の留学生の日本語レベルが多様化する一方、日本人学生の日本語能力も低下していると言われる (日本リメディアル教育学会 2011 等)。家庭環境や在住歴も多様化し、日本国内の大学で「留学生」とひとくくりにして日本語必修とする正当性が疑われる状況さえある。田島・佐藤・橋本・笹尾・松下 (2015) は日本語が第一言語 (L1) か第二言語 (L2) かの区別なく語彙量を測定するために「日本語を読むための語彙量テスト (50K 版)」(VSTRJ-50K) を開発し、日本語 L1 学生に試行した。本研究では語彙力と漢字力との関係、L1 と L2 による語彙発達の違い等を調べるため、新たに「漢字変換力テスト」(KCT) も開発し、VSTRJ-50K と合わせて日本語 L2 学生を対象に実施した。

研究課題は以下の 5 点である。1) 日本の大学生の一般的な日本語語彙量と留学生 (L2 学習者) の日本語語彙量には差があるかないか、あるとすればどの程度か。2) 日本の大学生の一般的な漢字変換力と留学生 (L2 学習者) の漢字変換力には差があるかないか、あるとすればどの程度か。3) L2 学習者の場合、語彙量や漢字変換力は学習期間とどのような関係があるか。4) 一般的に語彙量と漢字変換力はどの程度相関があるのか。5) 語彙量と漢字変換力の両テストにつき、L1 によって得点のパターンに違いがあるか。ある場合、どのような項目に違いがあり、どのようなパターンがあるか。

2. 研究方法

2.1 日本語を読むための語彙量テスト (50K 版) (VSTRJ-50K)

VSTRJ-50K は、「日本語を読むための語彙データベース」(松下 2011) の頻度上位 5 万語から抽出した 125 語を対象とし、意味記述を選ぶ四肢選択式である (図 1)。上位 5 万語を 400 語ずつ 125 のゾーンに区切り、各ゾーンから 1 語ずつ無作為に選んでいる。他の問題と同一漢字が使われる語が選ばれた場合には頻度の近い別語に変更するようにした。順序効果や疲労効果による影響を軽減するため、問題配列の異なる 2 バージョンを作成し、受験者に無作為に配布した。

意味記述はできるだけ初級語彙で記述し、それが難しい場合にも当該項目の頻度よりもやさしい語彙のみで平易に記述するようにした。錯乱肢 (不正解の選択肢) の記述では意味が部分的に重複しないようにすることとし、それぞれが全く違う意味になるようにした。錯乱肢を含む選択肢は、すべてターゲット

[6000 語レベル]

礼儀 ひと あの人には礼儀 し を知らない。

- 1) おおよその内容をまとめたもの
- 2) 人間関係を守るための行動様式
- 3) ものを大切に思う気持ち
- 4) 理論とは異なる現場の仕事

* 実際の問題とは異なる

図 1 VSTRJ-50K 問題例

ト文に当てはめても通じるものとし、ターゲット文は意味の違いに手がかりを与えない非定義文とした。ターゲット文、選択肢ともに、ターゲット語以外の漢字にはすべて振り仮名をつけた。

2.2 漢字力変換テスト (KCT)

「漢字力変換テスト」(KCT)は、常用漢字から抽出した80字を対象とし、短文中の下線部の漢字を選ぶ四肢選択式である。「現代日本語文字データベース」(松下 2013, 2014)の頻度上位2000字を25字ずつ80のゾーンに区切り、基本的には各ゾーンの中心の字を選ぶようにした。ただし、旧日本語教育能力試験(旧 JLPT)の級別、および小学校の学年配当のそれぞれのレベルの割合に応じて問題数を割り振り、頻度順位の低位から上位にかけて、旧 JLPT、学年配当のそれぞれで順にレベルが上がるようにした(表1)。かなで分かち書き表記した短文に、正解の字を使用した語をすべてカタカナで表記して埋め込み、ターゲット漢字の部分に下線を引いた(図2)。正解以外の三つの錯乱肢は、正解の字と、意味・字形・発音のそれぞれが類似する、最も頻度の近い字を選んだ。

また、ターゲットの字を含む語の頻度も、上位2万語を250語ずつに区切った80のゾーンの頻度順位が中央に最も近い語とするようにした。すなわち、頻度、旧 JLPT、学年配当、語彙レベルのすべてにおいて、問題が易から難と並べられるようにした。ただし、問題を無作為に配列した2バージョンを作成し、受験者に無作為に配布した。

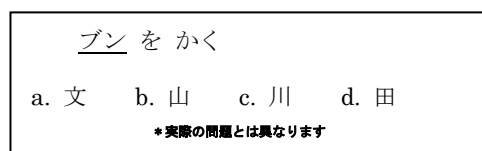


図2 漢字変換力テストの問題例

表1 漢字変換力テストの項目選定基準

漢字頻度ランク	1-250	251-500	501-750	751-1000	1001-1250	1250-1500	1501-1750	1751-2000
問題数	10	10	10	10	10	10	10	10
標準サンプリング間隔(ランク)	25	25	25	25	25	25	25	25
総語彙数目安	2500	5000	7500	10000	12500	15000	17500	20000
語彙頻度ランク(基準)	1-2500	2501-5000	5001-7500	7501-10000	10001-12500	12501-15000	15001-17500	17501-20000
旧日本語能力検定試験レベル	4級/3級	3級/2級	2級		1級			1級/級外
学年配当	1年~2年	3年	3年~5年	5年~中学	中学以上			
レベル	初級		中級		上級			上級/超級

2.3 テストの実施時期、場所、および対象

調査は2015年4月~2016年6月にかけて、日本、韓国、中国、インド、スウェーデンで日本語をL1またはL2として使う大学生等851名(うちVSTRJ-50K受験者846名、KCT受験者452名)を対象に、各学生の所属校(日本国内3大学、国外5大学を含む計11機関)で実施した(表2)。二つのテストと合わせて、学習者の使用言語や学習期間などを回答してもらう質問紙調査を実施した。制限時間はVSTRJ-50Kが40分で、KCTが20分であった。

表2 テスト参加者の属性・受験バージョン

テスト	男女別(人)				L1タイプ別(人)				年齢(歳)		バージョン	
	合計	男	女	不明	日	中	韓	その他	M	SD	A	B
VSTRJ-50K	846	490	352	4	407	201	100	138	20.0	3.5	427	419
KCT	452	196	253	3	34	192	100	126	21.2	4.3	230	222

VSTRJ-50K: 日本語を読むための語彙量テスト(50K版)、KCT: 漢字変換力テスト

日: 日本語、中: 中国語、韓: 韓国語、M: 平均、SD: 標準偏差

3. 結果と考察

VSTRJ-50K (n=846) の平均は 80.5 点、標準偏差は 30.9 点であった。KCT (n=452) の平均は 55.6 点、標準偏差は 18.8 点であった。テストの信頼性 (内的一貫性) は高く、VSTRJ-50K および KCT のクロンバック α の値はそれぞれ.98 と.97 であった。

- 1) 日本の大学生の一般的な日本語語彙量と留学生 (L2 学習者) の日本語語彙量には差があるかないか、あるとすればどの程度か。
- 2) 日本の大学生の一般的な漢字変換力と留学生 (L2 学習者) の漢字変換力には差があるかないか、あるとすればどの程度か。

L1, L2 の受験者がともに存在した日本国内の大学生だけに絞ってみると、表 3 の結果になる。

表 3 日本国内の正規大学生の日本語 L1、L2 の大学別、テスト別の得点 (平均・標準偏差) と平均の差の検定結果

	VSTRJ-50K 「日本語を読むための語彙量テスト(50K版)」								KCT 「漢字変換力テスト」																			
	日本語L1			日本語L2			t 検定結果				効果量				日本語L1			日本語L2			t 検定結果				効果量			
	n	M	SD	n	M	SD	t	df	p	r	n	M	SD	n	M	SD	t	df	p	r	n	M	SD	t	df	p	r	
A大学	109	112.7	5.8	14	69.4	24.3	6.41	13.18	.000	0.87	34	79.1	1.1	14	48.9	18.1	5.99	18.04	.000	0.82								
B大学	289	100.2	15.6	2	45.0	1.0	5.00	289.0	.000	0.28	0	--	--	2	52.0	8.0	--	--	--	--								
C大学	9	99.7	9.9	36	50.9	24.3	9.0	31.4	.000	0.85	0	--	--	17	59.9	14.4	--	--	--	--								
3大学	407	103.6	14.6	52	55.6	25.3	13.3	55.4	.000	0.87	34	--	--	33	54.8	16.7	--	--	--	--								

日本語 L2 のテスト参加者の少なかったケースを除き、いずれも有意差があり、B 大学以外の効果量の値も 0.5 以上と大きく、日本語 L1 の学生と L2 の学生には大きな差があるといえる。平均理解語彙量は 4 万語以上と推計されるのに対し、日本国内の日本語 L2 の大学生では、中国語 L1 学生 (n=22) の平均理解語彙量が 3 万 1 千語程度なのに対し、非漢字圏出身学生 (n=33) は 1 万 7 千語程度であった。ただし、日本語 L1 の学生にも、VSTRJ-50K の得点が日本語 L2 の平均 55.6 点より低い学生が 407 人中 8 人 (2.0%) もおり、L1 なら常に十分な語彙量があるとは言えない。

- 3) L2 学習者の場合、語彙量や漢字変換力は学習期間とどのような関係があるか。

表 4 L2 日本語レベル別推定既知語数と学習時間

合計点	推定平均既知語数*	推定平均増加語数 (A)	人数	平均学習時間	平均学習時間差 (B)	1語平均学習時間 (分) (B/A)
0-20	5123	5123	26	580.6	580.6	6.8
21-40	12683	7560	65	844.0	263.4	2.1
41-60	20213	7529	64	1014.3	170.2	1.4
61-80	28454	8241	67	1643.0	628.7	4.6
81-100	36602	8149	85	2349.8	706.8	5.2
101-125	41891	5289	11	2414.5	64.8	0.7

*該当する受験者の得点の平均に400をかけた数値

L2 日本語学習者が L1 学習者並みになるには 2000 時間程度は必要だと考えられる (表 4) が、韓国語母語学習者は L1 並みになるのが速い (表 5)。また初級では語も漢字も比較的学习に時間がかかるが、中級にかけて効率的に学習できるようになる。しかし、中級から上級に至ると再び時間がかかるようになる。この点は語も漢字も共通している。ただし、L1 による違いを考慮する必要はある。

表 5 L2 日本語レベル別推定「可変換漢字」数と学習時間

合計点	推定可変換字数*	推定平均増加可変換字数 (A)	人数	平均学習時間	平均学習時間差 (B)	1漢字平均学習時間 (分) (B/A)	L1中国語		L1韓国語		L1その他	
							人数	平均学習時間	人数	平均学習時間	人数	平均学習時間
0-20	359	359	21	547.8	547.8	91.5	1	576	0		20	546
21-40	765	406	50	705.7	157.9	23.3	6	844	12	461	30	788
41-60	1259	493	90	1003.3	297.6	36.2	20	1575	45	466	24	1554
61-80	1778	519	146	2129.3	1126.0	130.1	105	2470	33	1392	6	708

*該当する受験者の得点の平均に25をかけた数値

4) 一般的に語彙量と漢字変換力はどの程度相関があるのか。

VSTRJ-50K と KCT の総得点の相関係数（ピアソン）は $r=.82$ で、ラッシュ分析の能力値で見ると $r=.85$ である。一般的に語彙量と漢字変換力にはかなり高い相関があるといえる。

VSTRJ-50K では、漢字を含む書字形態（文字表記）から意味へアクセスできるかどうかをテストしているのに対し、KCT ではカナから（音韻への変換を経て）書字形態（漢字）への変換をテストしている。一見異なる能力のようだが、カナから正しく漢字に変換するためには（文脈に合致する）語を漢字で想起しなければならず、多くの場合、意味へのアクセスを必要とするため、結果的に重複する能力を測定しているものと考えられる。その重複の程度は相当に高いと言える。

5) 語彙量と漢字変換力の両テストにつき、L1 によって得点のパターンに違いがあるか。ある場合、どのような項目に違いがあり、どのようなパターンがあるか。

VSTRJ-50K と KCT の両テストについて、L1 を区別せずにラッシュ分析を行うと、非常に多くの項目がミスフィットとして検出された。そこで、日、中、韓、その他の四つの言語タイプ別に DIF (Differential Item Functioning) 分析を行ったところ、各グループ内ではラッシュモデル適合する項目数が増大した。L1 別、語種別に VSTRJ-50K の項目難度を見ると、L1 中国語学習者にとっては相対的に外来語が難しく、漢語は易しいが、英語やその他の L1 の学習者は全くその逆である（図 3）。

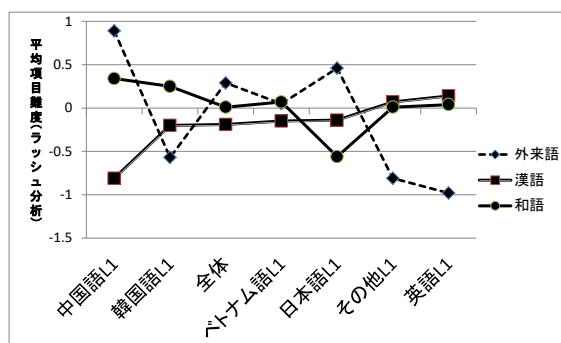


図 3 L1 別、語種別の VSTRJ-50K の平均項目難度

4. まとめと今後の課題

本調査の対象となった複数の大学では、日本語 L1 と L2 の学生の間には少なく見ても 1 万語以上の既知語数の差があり、「可変換漢字」数にも 700 字程度の差がある。学習時間数にして少なくとも数百時間以上の差がある。語彙量と漢字変換力にはかなり高い相関がある。L1 中国語の学習者は漢語に強いが、L1 韓国語学習者の得点パターンは日本語 L1 の学生に近い。これらの結果は、L1/L2 を包括的にみて語彙力・漢字力を高めるカリキュラム開発を行う必要性を示唆している。

今後は、母語別、語のタイプ別により詳しく調査する必要があるだろう。学生生活に十分な語彙力や漢字力を身に着けるのに必要な時間を、カリキュラムにどう反映するかも検討課題である。

【参考文献】

田島ますみ・佐藤尚子・橋本美香・笹尾洋介・松下達彦（2015）「日本人大学生の日本語語彙量測定の試み」『中央学院大学 人間・自然論叢』41, pp.3-20

日本リメディアル教育学会（2011）「日本語リメディアル教育とは」『リメディアル教育研究』6(1), pp.3-46

松下達彦（2011）「日本語を読むための語彙データベース」(The Vocabulary Database for Reading Japanese) Version 1.1

URL: <http://tatsuma2010.web.fc2.com/>（最終確認日：2016年8月11日）

松下達彦（2013）「現代日本語文字データベースの開発と検証」The 18th Biennial Conference of the Japanese Studies Association of Australia, Australian National University, AUSTRALIA.（2013年7月10日）

松下達彦（2014）「現代日本語文字データベース」(Character Database of Modern Japanese) Version 2.0

URL: <http://www17408ui.sakura.ne.jp/tatsum/CDJ-top.html>（最終確認日：2016年8月11日）