

多肢選択テストで会話能力を予測する際に生じる誤差の分析

齊藤, 信浩
九州大学留学生センター : 准教授

<https://doi.org/10.15017/4783557>

出版情報 : 九州大学留学生センター紀要. 29, pp.13-19, 2021-03. 九州大学留学生センター
バージョン :
権利関係 :

多肢選択テストで会話能力を予測する際に生じる 誤差の分析

Analysis of gaps in predicting conversational ability with multiple-choice tests

齊 藤 信 浩*

〈要旨〉

本研究は、四肢選択式の文法・聴解・読解の各テストの得点と、対面方式のインタビューテストの結果がどれくらい差があるのかを調査した。そして、文法・聴解・読解の各テストの得点を、語圏別にインタビューテストの結果のSレベルを予測する重回帰分析によって検証した結果、非漢字圏と準漢字圏の両群は全ての得点がSレベルを予測していたが、漢字圏は文法と聴解は有意な因果関係が見られたが、読解の得点だけはSレベルを予測していなかった。この結果は、漢字圏の学習者の読解の得点か他の文法や聴解の得点と異なった出方をしていることを示唆している。そして、インタビューテスト後のレベル変動から、語圏に関係なく、四肢選択式の文法・聴解・読解の各テストの得点は実際の会話能力よりも高い評価を与える傾向があることがわかった。

キーワード：会話能力、漢字圏、語圏別分析、得点分析、重回帰分析

1. はじめに

文法・聴解・読解の能力は四肢選択式のテストでも測定することは可能だが、会話能力を測定することは困難である。しかしながら、大規模な測定テストを実施する場合においては、一定レベルの経済性が求められる(尾崎2008)。それでは、会話能力をインタビューテストなしで測定することは可能であろうか。恐らく、それはかなり困難な作業であると思われる。私達日本人の英語学習の経験からも、高度な文法知識や大量の語彙知識を持っているにも関わらず、

会話は不得意であるということがあり、言語知識と言語運用の間にギャップがあることはよく知られている。しかし、基本的な文法・読解・聴解の能力を測定することができれば、一定程度の会話力を予測することはできないのであろうか、もしできないのであるとすれば、そのギャップはどれくらい存在するのであろうか、これが本研究の着眼点である。九州大学留学生センター(および基幹教育院)で開講されている日本語科目(Japanese Academic Courses; JACs)では、総合日本語(Integrated Japanese)、会話(Speaking)、漢字(Kanji)、作文(Writing)

*九州大学留学生センター准教授

の4つのコースが開講され、学生は日本語の習熟度ごとに8レベルに分けられる。その習熟度を測定するために、オンライン・プレースメント・テスト (Online Placement Tests ; OPT) の形式で、文法テスト、聴解テスト、読解テストが与えられている。まず、このOPTにより、Jレベルを8レベルに分ける。そして、そのJレベルを参考にして、会話 (Speaking) クラスのレベル分けのために、教員と学生の1対1の対面方式でインタビューテストが行われ、Sレベルが決定される。しかしながら、インタビューテストの結果で分けられたSレベルとOPTによって分けられたJレベルとの間には、一定の関連性はあるつつも、かなり大きな差がみられる。本研究では、このJレベルとSレベルの差がどれくらいあるのかを数値的に示すことを目的とし、その上で、語圏別にSレベルを文法・読解・聴解の各テストの得点が予測しうるかどうかを重回帰分析で検討していく。

2. 調査方法

2. 1. 受験者と受験方法

調査期間は2011年から2017年までの6年間で、九州大学留学生センターのオンラインテス

トで収集された多国籍の日本語学習者484名 (男 =208, 女 =276) のデータを基にする。受験者の内訳は、語圏別^{注1)}に、漢字圏 (中国・台湾・香港) が148名、準漢字圏 (韓国) が91名、非漢字圏^{注2)} が245名である。これらの受験者は、旧日本語能力試験 (JLPT) に準拠した^{注3)} 四肢選択式の文法・聴解・読解の3種類のテストを受ける (国際交流基金2007)。文法テストは、60問 (30分)、聴解テストは18問 (50分)、読解テストは38問 (45分) で構成されている。これらの得点を基に、Jレベルが設定される。得点の基準は表1に示した。得点の根拠については小森 (2011)、斉藤 (2012) を参照された。

J-1からJ-4までは文法の得点によってレベル分けを行い、J-5とJ-7を読解の得点を中心にレベル分けをし、最上級のJ-7とJ-8のレベルを聴解の得点で分けている^{注4)}。このJレベルが学生に通知されると、会話 (Speaking) クラスの受講希望者は、一旦は、Jレベルに従って、Sのインタビューテスト^{注5)}を受ける。J-3と判定された者はS-3のインタビューテストを受け、レベルに合致しない場合は、下、あるいは上のレベルのインタビューテストを受け、Sレベルが決定される仕組みである。以下の3. 1. 節では、

表1 日本語レベルの判定基準

J level		文法 (60点)	読解 (38点)	聴解 (18点)	S level
J-1	Beginner	0 ~ 12			—
J-2	Elementary 1	13 ~ 24			S-2
J-3	Elementary 2	25 ~ 36			S-3
J-4	Pre-Intermediate	37 ~ 44			S-4
J-5	Intermediate 1	45 ~ 48			S-5
J-6	Intermediate 2	49 ~ 60	20 ~ 25	9 ~ 10	S-6
J-7	Pre-Advanced		26 ~ 31	11 ~ 13	S-7
J-8	Advanced		32 ~ 38	14 ~ 18	S-8

このJレベルとSレベルのギャップを分析する。

2. 2. 語圏別に見た OPT の得点

3章の分析に入る前に、語圏（漢字圏・準漢字圏・非漢字圏）ごとに、文法・聴解・読解の3種類のテストの得点に差がないかを、3（語圏；非漢字圏・準漢字圏・漢字圏）×3（テスト；文法・聴解・読解）の反復のない一元配置の分散分析（ANOVA）で検討する。分析にはPASW (SPSS) ver.18を用いた (Arbuckle 2009)。以下、表2に結果を展開した。

まず、文法の得点は語圏間で主効果が有意ではなく、差がなかった [$F(2,481)=3.014, p=.050, ns$]。得点を見ても非漢字圏 ($M=42.49, SD=12.81$)、準漢字圏 ($M=45.15, SD=14.17$)、漢字圏 ($M=45.47, SD=12.05$) で差がないことが観察される。一方、聴解は [$F(2,481)=38.944, p<.001$]、読解は [$F(2,481)=110.494, p<.001$] で、主効果が有意であった。主効果が有意であったため、シェフェ (scheffe) の多重比較の方法によって語圏間の得点差の比較を行った結果、聴解では、非漢字圏 ($M=8.02, SD=3.77$) < 準漢字圏 ($M=11.30, SD=5.41$) = 漢字圏 ($M=11.49, SD=4.15$) で、準漢字圏と漢字圏は同レベルの得点だったが、非漢字圏は低かった。読解でも、非漢字圏 ($M=14.16, SD=6.74$) < 準漢字圏 ($M=22.32, SD=7.68$) = 漢字圏 ($M=24.31, SD=6.93$) で、準漢字圏と漢字圏は同レベルの

得点だったが、非漢字圏は格段に低かった。この結果、語圏ごとに見ると、文法の得点は等しいが、聴解と読解の得点は非漢字圏は低いことが観察された。

3. 調査結果

3. 1. OPT の得点とインタビューテストの結果との差

前節2. 2. では、非漢字圏の受験者は、聴解と読解の得点が漢字圏や準漢字圏の受験者よりも劣るという結果を得た。本節では、インタビューテストによって得られた会話能力の判定が、JレベルとSレベルの間で一致するかどうかを報告する。まず、表3はSレベルの結果が記載されている。「下から移動」は、S-5 (=J-5) と判定されてインタビューテストを受けた結果、5のレベルの基準よりも会話が優れていたため、上のレベルへ繰り上げた者を意味している。「上から移動」はその逆にS-5 (=J-5) と判定されたが会話力が低く、下へ落とされた者である。

全数354名のうち55名が「下から移動 (OPT判定よりも会話が上手だった)」し、全体の比率は15.5%である。「上から移動 (OPT判定よりも会話が下手だった)」受験者は354名のうち118名で、33.3%だった。それでは、語圏別に見ると、どうなるのであろうか。表3を語圏別に分

表2 語圏別に見た文法・聴解・読解テストの得点差の検証 (ANOVA)

	非漢字圏		準漢字圏		漢字圏		主効果	多重比較
	n=245		n=91		n=148			
	M	SD	M	SD	M	SD		
文法	42.49	12.81	45.15	14.17	45.47	12.05	ns	—
聴解	8.02	3.77	11.30	5.41	11.49	4.15	***	非<準=漢
読解	14.16	6.74	22.32	7.68	24.31	6.93	***	非<準=漢

注) M=平均, SD=標準偏差. $p<.05$ *, $p<.01$ **, $p<.001$ ***, ns=not significant.

表3 OPT 結果とインタビューの差

	全数	下から移動	比率 (%)	上から移動	比率 (%)
S2	34	1	2.9	18	52.9
S3	64	15	23.4	18	28.1
S4	89	9	10.1	30	33.7
S5	89	13	14.6	37	41.6
S6	78	17	21.8	15	19.2
全体平均	354	55	15.5	118	33.3

表4 語圏別にみた OPT の結果とインタビューの差

		下から移動	比率 (%)		上から移動	比率 (%)
S2	非漢字圏	1	0.4	非漢字圏	9	3.7
	準漢字圏			準漢字圏		
	漢字圏			漢字圏	9	6.1
S3	非漢字圏	9	3.7	非漢字圏	13	5.3
	準漢字圏	1	1.1	準漢字圏	1	1.1
	漢字圏	5	3.4	漢字圏	4	2.7
S4	非漢字圏	8	3.3	非漢字圏	14	5.7
	準漢字圏			準漢字圏	7	7.7
	漢字圏	1	0.7	漢字圏	9	6.1
S5	非漢字圏	6	2.4	非漢字圏	22	9.0
	準漢字圏	4	4.4	準漢字圏	8	8.8
	漢字圏	3	2.0	漢字圏	7	4.7
S6	非漢字圏	4	1.6	非漢字圏	4	1.6
	準漢字圏	6	6.6	準漢字圏	3	3.3
	漢字圏	7	4.7	漢字圏	8	5.4

表5 OPT とインタビューの差 (語圏別合計)

	人数	下から移動 (%)	上から移動 (%)
非漢字圏	245	11.4	25.3
準漢字圏	91	12.1	20.9
漢字圏	148	10.8	25.0

解し、表4のように展開し、語圏別の合計は表5に集計をした。

表4を見ると、全体に分散しているため傾向が読み取れないが、表5のように集計すると、全体の傾向が見えて来る。「下から移動 (OPT

よりも会話が上手だった)」受験者は、非漢字圏は11.4%、準漢字圏は12.1%、漢字圏は10.8%であり、語圏間に特筆するような差がなかったと言える。一方、「上から移動 (OPTよりも会話が下手だった)」のは、非漢字圏は25.3%、準漢

字圏は20.9%、漢字圏は25.0%であり、やはり、これも語圏間に差はなかったと言える。下から移動と上から移動の差を見ると、15%程度が上から移動していた。

3. 2. Sレベルの重回帰分析

ここまでで観察されたのは、どの語圏の受験者も文法テストは同一のレベルの得点を得る一方で、非漢字圏の学習者は漢字圏や準漢字圏の受験者よりも、得点が低く、会話能力ではどの語圏の受験者も等しく、OPTの得点の方が高く判定されていることが観察された。それでは、Sレベルのレベル変動をOPTの得点が予測するかどうかを、重回帰分析の手法で検討していく。語圏別に、Sレベルのレベル変動（順序変数）を従属変数とし、その変動を、文法・聴解・読解の各テストの得点（間隔尺度）を独立変数とした、5%有意の強制投入法による重回帰分析によって検討した。表6は非漢字圏の受験者のSレベルをOPTの得点で予測した結果である^{注6)}。

まず、回帰式の妥当性を確かめるために多重共線性（multicollinearity）^{注7)}の疑いがないかどうかを診断する。VIF（Variable Inflating Factor：分散拡大要因）と許容度を見る。VIFは5以下が望ましく、10よりも大きい変数であれば、他の

独立変数と強い相関を有している可能性が高く、多重共線性が原因である可能性が高いと診断される。表6では、VIFは1.759～2.377であるため、全く問題ない。許容度は、最高値が.999であり、それに近ければ許容度が高いことになるが、ここでは最も高い数値でも.568であるので、許容度も十分な数値であると言え、多重共線性の疑いはなかった。それでは、重回帰分析の結果の数値を見ていく。決定係数は $R^2=.666$ （調整済み $R^2=.662$ ）^{注8)}で、十分な値だと思われる。この数値は「Sレベルの変動の要素をこの回帰式によって66.6%説明できる」ということを示している。独立変数は、文法・聴解・読解のどのテストの得点も0.01%水準で有意であり、Sレベルの変動には、文法・聴解・読解の得点のどれかが関わっていることを示している。

同様の手法で、準漢字圏のSレベルも重回帰分析によって検討していく。結果を表7に展開した。表7では、VIFは2.202～2.601であり、許容度は最も高いものでも.454であり、VIFも許容度も十分な数値であると言え、多重共線性の疑いはなかった。重回帰分析の結果、決定係数は $R^2=.624$ （調整済み $R^2=.609$ ）で、十分な値であった。独立変数は、文法・聴解は、5%水準で有意であり、読解は0.01%水準で有意であり、Sレベルの変動には、文法・聴解・読解の得点の

表6 非漢字圏受験者のSレベルを予測する重回帰分析

$R^2=.666$, 調整済み $R^2=.662$		共線性の統計量				
$n=244$	<i>M</i>	<i>SD</i>	β	<i>t</i> 値	許容度	VIF
予測変数（従属変数）						
Sレベル						
説明変数（独立変数）						
文法	42.58	12.76	.235	***	.568	1.759
聴解	44.49	21.00	.333	***	.421	2.377
読解	37.27	17.75	.358	***	.434	2.307

注) *M* = 平均, *SD* = 標準偏差, $p < .05$ *, $p < .01$ **, $p < .001$ ***, *ns* = not significant.

表7 準漢字圏受験者のSレベルを予測する重回帰分析

		$R^2=.624$, 調整済み $R^2=.609$				共線性の統計量	
		M	SD	β	t 値	許容度	VIF
予測変数 (従属変数)							
Sレベル							
説明変数 (独立変数)							
	文法	45.56	13.90	.212	*	.430	2.324
	聴解	65.74	27.38	.243	*	.454	2.202
	読解	58.74	20.20	.422	***	.384	2.601

注) M = 平均, SD = 標準偏差, $p<.05$ *, $p<.01$ **, $p<.001$ ***, ns =not significant.

表8 漢字圏受験者のSレベルを予測する重回帰分析

		$R^2=.538$, 調整済み $R^2=.528$				共線性の統計量	
		M	SD	β	t 値	許容度	VIF
予測変数 (従属変数)							
Sレベル							
説明変数 (独立変数)							
	文法	45.62	11.95	.321	***	.500	1.999
	聴解	64.29	22.52	.400	***	.423	2.362
	読解	63.98	18.26	.103	ns	.346	2.890

注) M = 平均, SD = 標準偏差, $p<.05$ *, $p<.01$ **, $p<.001$ ***, ns =not significant.

どれも関わっていることを示していた。

漢字圏のSレベルの重回帰分析の結果は表8の通りである。表8では、VIFは1.999～2.890であり、許容度は最も高いものでも.500であり、VIFも許容度も十分な数値であると言え、多重共線性の疑いはなかった。重回帰分析の結果、決定係数は $R^2=.538$ (調整済み $R^2=.528$)であった。独立変数は、文法・聴解は、0.01%水準で有意であり、Sレベルの変動を予測していたが、読解の得点は有意な因果関係が見られなかった。

4. 考察

OPTの得点は、非漢字圏は、文法＝聴解＝読解、準漢字圏と漢字圏は、文法<聴解＝読解で

あった。推測されるのは、準漢字圏・漢字圏の受験者は「漢字の識字力」で有利にテキストを理解してしまうために、基本的な文法力よりも高く得点できるのだと思われる。識字力のみではなく、漢字語彙の意味や用法などについても共通項が多いため、そのぶん有利だったのだと思われる。

そして、OPTとインタビューテストの比較では、3割の被験者が、基礎的な日本語能力(OPT得点)よりも会話力が高く判定され、インタビューの結果、下のレベルへ移動させられる結果になっていた。これは、表現力や音感などで上手く聞こえてしまうが、実際に正確に使用できているかどうかは別の問題であり、受動的な筆記テストと会話能力には差が生じる。そして大抵は、同等か低く判定される傾向が窺え

た。しかし、語圏別には、その差はなかった。

次にSレベルを予測する重回帰分析では、非漢字圏と準漢字圏の全ての項目が会話能力を予測する変数となっていた一方、漢字圏の受験者のSレベルの変動に読解の得点は関連性が見られなかった。これは恐らく、漢字圏の受験者は読解テストを漢字による識字力で有利に進めており、それにより、得点が不自然に高く出ているため、Sレベルとの連動が出なかったのではないかと考えられる。即ち、識字力と会話能力は直結せず、「漢字」というのは、言語能力の中でも、特殊な能力であると考えられる。全体として、四肢選択式のテストで測定される能力よりも会話能力は低く、やはり文法能力や語彙能力を測定するテストは言語知識を測定することはできるものの、会話能力のような言語運用力を正確に測定することは困難であり、15%程度の誤差が生じることがわかった。

注

- 1) 「語圏」という用語は一般的に定着している用語ではないが、母語で漢字を常用しているか否かで分けられた区別で、一部の論文では用いられている用語である。漢字圏は、中国・台湾・香港・日本のように、日常的に漢字表記が用いられる国を漢字圏とし、準漢字圏は韓国のように部分的に漢字表記が用いられ、国民もある一定の漢字の知識を持っている国を指す。非漢字圏はヨーロッパやアフリカなどの漢字とは全く縁のない国々を指している。
- 2) あくまで識字力の観点から分類を行なっている。そのため、ベトナムやシンガポールやマレーシアは非漢字圏の中にも含めた。ベトナムは本来であれば母語に漢字語を大量に有しているのであるが、漢字を表記として用いていない。また、シンガポールやマレーシアの中華系の民族も漢字知識については個々人の環境差によるところが大きいので、非漢字圏に分類した。
- 3) 読解テストや聴解テストはJLPTを参考に上級レベルの分類のために作成されている。文法テストも

出題形式はJLPTを参考にしているが、九州大学の日本語コースで出現する文型や語彙のレベル基準を配慮して作成されている。

- 4) 聴解と読解の得点は、J-5以下ではプレースメントに直接的な関係はないものの、参考値として考慮し、その得点によってレベル分けが調整されることもある。
- 5) インタビューテストの時間は1名あたり10分間で、会話クラスを担当する教員が判定を行なっている。
- 6) 文法の得点は60点満点、読解は38点満点で聴解は18点満点であるが、読解と聴解は、100%に尺度を直してある。
- 7) 重回帰分析の場合、1つの従属変数を複数の独立変数で予測するため、独立変数の数が多くなると、それらが相互に干渉し合っ、誤った結果を導く現象があり、これを多重共線性と呼ぶ。通常、2つか3つ程度の独立変数であれば、多重共線性の疑いは生じないが、その診断のためにVIFや許容度の数値から判断し、多重共線性の問題を回避する。
- 8) 決定係数は、従属変数の数値の変動を独立変数で予測する割合を示したものである。となれば、独立変数の数が多ければ多いほど、決定係数が高くなってしまおうという問題が生じる。これは多重共線性と同じタイプの問題である。そのため、独立変数の数を考慮して調整されたのが調整済み決定係数である。通常、決定係数と調整済み決定係数の両方を報告するが、どちらか片方のみを報告する際には調整済み決定係数の方を報告する。

参考文献

- 尾崎茂 (2008) 言語テスト学入門—テスト作成の基本理念と研究法—。大学教育出版。
- 小森和子 (2011) プレースメントテストのオンライン化の試みと問題項目の分析評価, 九州大学留学生センター紀要19, 89-106.
- 斉藤信浩 (2012) オンラインプレースメントテスト問題項目の分析評価, 九州大学留学生センター紀要20, 101-114.
- 国際交流基金・日本国際教育協会 (2007) 『日本語能力試験出題基準【改訂版】(第4版)』凡人社。
- 日本テスト学会編 (2010) 『見直そうテストを支える基本の技術と教育』金子書房。
- Arbuckle, J. L. (2009). *AMOS 18.0 User's Guide*. Chicago, IL: SPSS Inc.