日本語 Can - do - statements における DIF 項目の検出

野口裕之・熊谷龍一・脇田貴文・和田晃子
名古屋大学・新潟大学・京都大学・国際交流基金

Detection of DIF items in Japanese Can-do-statements

NOGUCHI Hiroyuki, KUMAGAI Ryuichi, WAKITA Takafumi, WADA Akiko
Nagoya University, Niigata University, Kyoto University, Japan Foundation

The purpose of this study was to conduct DIF analysis of Japanese Can-do-statements, focusing on the learner's mother tongue, either Chinese or Korean, and investigate (1) concrete verbal behavioral situations that cause DIF and (2) the extent to which differences in methods for detecting DIF affects the results.

Results indicated that (1) DIF was advantageous to a speaker whose mother tongue was Chinese, in the direction of "Can-do," for 5 question items on reading and writing skills. On the other hand, DIF was advantageous to the speaker whose mother tongue was Korean, in the direction of "Can-do," for 5 question items regarding speaking and listening skills. (2) Similar results, from the perspective detecting DIF, were obtained using SIBTEST, BILOG-MG DIF command, Likelihood ratio test, the Mantel-Haenszel statistical method, and Logistic regression method.

1. はじめに

Can-do-statements とは、学習者の言語能力を具体的に記述したものを云う（島田 他, 2003）。海外では、TOEIC(Test of English for International Communication)や ALTE(The Association of Language Testers in Europe: 欧州言語テスト協会)などで測定結果、すなわち得点に対する能力評価を具体的に記述するために、実際に活用されている。これらは、英語や欧州諸言語を対象としているが、外国語としての日本語に関しては、日本語能力試験の妥当性を検証するための外在基準の一つとして、「日本語 Can-do-statements」が開発されている。これは、日本語による言語行動を記述した項目に対して、回答者がどの程度できるかを自己評価するものであり、最終的に確定された版においては、

「読む」「書く」「聞く」「話す」の各技能 15 項目ずつ全部で 60 項目 (statements) から質問票が構成されている (三枝, 2004)。回答者は各項目に対する 7 段階評定尺度上のひとつのカテゴリを選択するように指示される。評定尺度の回答カテゴリは「1. 全然できない」から「7. 問題なくできる」までで、中央のカテゴリに「4. ある程度できる」という尺度表現が提示されている。調査対象者は日本への留学生に限定され、項目内容は日本にいる留学生の多くが現実遭遇する可能性のある場面が想定されている。この日本語 Can-do-statements に対して、野口・和田・熊谷・脇田 (2003) では、回答カテゴリを 2 値化した上で、60 項目の一次元性を確認し、2 パラメタ・ロジスティック・モデルを適用して IRT 尺度を構成している。

ところで、日本語学習者は異なるさまざまな背景を持っており、特に母語の違いが学習成果や習得速度に影響を与えると考えられる。そして、母語の違いの影響は日本語の学習に際して一様に現れるのではなく、特定の学習内容や言語使用場面に特異的に現れる可能性が高い。このため、日本語 Can-do-statements についても、回答者の母語の違いが特異的に影響する項目が存在する可能性がある。すなわち、特定の質問項目において、自己評価による日本語能力の水準が等しい回答者でも、ある特定の母語の回答者で「できる」と回答しやすいなど他の母語の回答者とは異なる回答傾向を示すことがある、というものである。このような項目を一般に DIF (Differential item functioning ; 特異項目機能) を示す項目と云う。このような項目は、日本語を学習する際に学習者の母語に応じたきめの細かい指導をするために、また、留学生等の日本での生活に適応するための指導に際して有用な情報となる。なお DIF は、i) ある特定の母語の回答者が一貫して「できる」と回答しやすいか、「できない」と回答しやすい場合の「均一 DIF (uniform DIF)」と、ii) 自己評価による日本語能力の一定水準までは、ある特定の母語の回答者が「できる」(または「できない」) と回答しやすいが、一定水準を超えると逆にその母語の回答者が「できない」(または「できる」) と回答しやすくなる場合の「不均一 DIF (non-uniform DIF)」とに分類される。

本研究においては、日本語 Can-do-statements に対して学習者の母語に着目した DIF 分析を実施して、

- i) 具体的にどのような言語行動場面で DIF が生ずるか、
- ii) DIF 検出の方法の違いがどの程度結果に影響するか、

を明らかにすることを目的とするが、特に留学生の日本語学習者に多く見られる「中国語母語話者」と「韓国語母語話者」の 2 つの下位集団を取り上げて分析する。

2. DIF 検出の方法

DIF 検出の方法には大きく分けて、i) IRT モデルを適用して DIF 検出のための統計量を計算する「パラメトリック法」と、ii) 特に測定モデルを仮定することなく、正答数 (正反応数) と項目に対する応答とを用いて DIF 検出のための統計量を計算する「ノン・パラメトリック法」の 2 種類がある。

パラメトリック法には、例えば次のような方法がある。

i) BILOG-MG の DIF コマンドによる方法

BILOG-MG (Zimowski, Muraki, Mislevy, & Bock, 1996) に含まれる DIF コマンドを利用して DIF の検出を行う。すなわち、下位集団 1 と下位集団 2 について、両集団における識別力パラメタは同一となるように制約を置いて項目パラメタ値を推定した後に、両集団の困難度パラメタについて等化 (equating) を行い、等化調整を行った後の両集団の困難度パラメタ推定値の差をもって、DIF の程度を示す。

ii) Thissen 他の尤度比検定法

DIF の存在を検討したい項目について、「下位集団間で項目パラメタ値は等しい」というモデルと、「下位集団間で項目パラメタ値は異なる」というモデルとを比較する検定統計量を構成し、実際に得られたデータから「下位集団間で項目パラメタ値は等しい」という検定仮説が棄却された時に当該項目で DIF が生じていると判断する (Thissen, Steinberg, & Wainer, 1993)。

また、ノンパラメトリック法には次のような方法がある。

iii) Mantel-Haenszel 法

受験者集団を能力水準で K 段階に分けて、各段階毎に下位集団 × 回答 (はい - いいえ) の 2×2 の分割表を作成する。これら K 個の分割表が全体として同じ傾向を示すか否かを M-H 統計量を用いて検定する。検定結果が有意な場合に当該項目に DIF が生じていると判断する。

iv) SIBTEST (Simultaneous Item Bias Test)

テストに含まれる項目の中で DIF がないと確信される項目群から個人の得点を計算して、受験者集団を K 段階に分割する。段階毎に DIF 検討項目の通過率の差を求め、段階を通して β 推定量と呼ばれる統計量を計算して、DIF の判断をする (Shealy & Stout, 1993)。

v) Logistic 回帰による方法

Logistic 回帰分析を利用して DIF の検出を行う方法で、基準変数に正答確率を、予測変数に i) 得点, ii) 下位集団, iii) 得点と下位集団との交互作用, を投入して分析する (適用例としては、例えば、前田・野口 (2006) 参照)。

これら 5 つの DIF 検出法のうちで、i) iii) iv) の方法を適用した場合には均一 DIF が、ii) v) の方法を適用した場合には均一 DIF および不均一 DIF が検出される。

本研究では、日本語 Can-do-statements60 項目に対して以上の 5 つの方法を用いて分析

を進める。

3. 方 法

3.1 データの概要

本研究に用いる日本語 Can-do-statements データは、国内の日本語学校（8 校）、および大学（3 校）に所属する外国人学生 794 名を対象とし、2002 年 12 月に実施されたものである。これらの学生は、大学進学を目指す外国人学生および、大学で学ぶ留学生で日本語能力試験を受験している。そして、項目分析や信頼性に関する分析などは、既に三枝（2004）などで公表されている。なお、本研究では、回答者の中から、中国語母語話者集団（269 名）と韓国語母語話者集団（383 名）に対して分析を行う。

3.2 分析の手順

DIF 検出の方法の多くは、DIF が生じていないことが既に確認されている項目群を基準として、DIF の可能性について検討したい項目を 1 項目ずつ分析する、という手順をとる。前者は matching item（以下、基準項目群と云う）、後者は suspect item（以下、DIF 検討項目と云う）と呼ばれるが、DIF 分析を開始する時点に於いては多くの場合、何れの項目に DIF が生じているのかは未知である。そのため、基準項目群を確定することから分析を始めなければならない。そこで本研究では、以下のような手続きで分析を進めた。

- 1) SIBTEST を適用して DIF 分析を進める。具体的には、
 - i) 日本語 Can-do-statements の第 1 項目を DIF 検討項目、残りの項目すべてを基準項目として β 推定量を算出する。続いて第 2 項目を DIF 検討項目として β 推定量を算出する。このようにして、日本語 Can-do-statements に含まれる項目を順次単独に DIF 検討項目として β 推定量を算出する手続きを 60 項目分繰り返す。
 - ii) β 推定量に基づいて、DIF が生じていると考えられる項目を分析から除外する。残った項目について i) を再度実施し、 β 推定量を算出する。DIF が生じていると判断する基準は、Bolt & Stout (1996) の結果を参照して、 β 推定量の絶対値が 0.1 を超えたものを DIF 項目とする。
この手続きを DIF 項目が検出されなくなるまで繰り返す。
 - iii) 最終的に DIF が検出されなくなった項目群を基準項目群として確定する。この基準項目群を用いて、ii) で分析から除外された項目について順次 β 推定量を算出し、DIF を検討する。

以上の手続きは、 β 推定量の算出を複数回繰り返す必要があり極めて煩雑になるが、DIFPACK (Stout, 1999) を用いることにより、容易に計算を行うことが可能である。
- 2) また、SIBTEST とは異なり、基準項目群を予め確定しておく必要のない BILOG-MG の DIF コマンドによる検出法を適用して分析する。
- 3) SIBTEST で確定された基準項目を用いて、SIBTEST で DIF が生じていると判断された項目について、

Thissen 他の尤度比検定法

Mantel-Haenszel 法

Logistic 回帰による方法

を適用して分析する。

- 4) 以上の5つの方法で DIF の検出される項目の異同を確認することにより、日本語 Can-do-statements における DIF 項目を最終的に確定する。

4. 結 果

4.1 SIBTEST による分析結果

日本語 Can-do-statements の 60 項目に対して SIBTEST を実施した結果、項目 4, 7, 16, 19, 22, 25, 29, 39, 46, 51 を除く 50 項目が基準項目とされた。

さらに、これらの基準項目を用いて、いま除外した各項目について再度 β 推定量、および、標準誤差を計算した結果は表 1 に示すとおりである。すべての項目で β 推定量の絶対値が 0.1 を超えており、これら 10 項目で DIF が検出された。

項目 4, 7, 16, 19, 25 で β 推定量の値が正であり、中国語母語話者が「できる」と回答しやすく、項目 22, 29, 39, 46, 51 で β 推定量の値が負であり、韓国語母語話者が「できる」と回答しやすいことが示された。

表 1 SIBTEST による分析結果

DIF 検出項目	β 推定量	標準誤差
4	0.210	0.036
7	0.131	0.033
16	0.121	0.034
19	0.267	0.032
22	-0.118	0.037
25	0.157	0.036
29	-0.113	0.032
39	-0.120	0.034
46	-0.127	0.033
51	-0.134	0.034

4.2 BILOG-MG の DIF コマンドによる分析結果

BILOG-MG の DIF コマンドにより全 60 項目に対して DIF の検出を行った結果得られた、等化調整後の困難度パラメタ推定値の差（中国語母語話者－韓国語母語話者）は表 2 に示すとおりである。

表 2 BILOG-MG の DIF コマンドによる困難度パラメタ推定値の差
(中国語母語話者集団－韓国語母語話者集団)

項目 01	0.150	項目 21	0.024	項目 41	0.149
項目 02	-0.050	項目 22	0.260	項目 42	-0.029
項目 03	-0.145	項目 23	0.040	項目 43	-0.066
項目 04	-0.591	項目 24	0.018	項目 44	-0.020
項目 05	0.244	項目 25	-0.447	項目 45	0.037
項目 06	-0.049	項目 26	-0.109	項目 46	0.348
項目 07	-0.295	項目 27	-0.084	項目 47	0.009
項目 08	0.019	項目 28	0.264	項目 48	0.124
項目 09	-0.046	項目 29	0.362	項目 49	0.012
項目 10	-0.241	項目 30	0.038	項目 50	0.106
項目 11	-0.070	項目 31	0.081	項目 51	0.297
項目 12	-0.125	項目 32	0.143	項目 52	-0.021
項目 13	0.185	項目 33	0.028	項目 53	-0.011
項目 14	-0.008	項目 34	0.154	項目 54	-0.015
項目 15	-0.112	項目 35	0.012	項目 55	0.141
項目 16	-0.298	項目 36	-0.002	項目 56	0.146
項目 17	-0.240	項目 37	0.044	項目 57	-0.139
項目 18	-0.159	項目 38	0.190	項目 58	-0.014
項目 19	-1.094	項目 39	0.345	項目 59	0.017
項目 20	-0.023	項目 40	0.344	項目 60	0.171

表 2 で強調されている 10 項目は SIBTEST において DIF 項目とされたものである。BILOG-MG の DIF コマンドにおいて示される DIF の程度は困難度パラメタ推定値の差で示されるが、DIF 項目と判断する基準値を設定することは難しい。しかしながら、SIBTEST の結果と比較すると、SIBTEST で DIF 項目とされたものは、相対的にその他の項目よりも困難度パラメタ推定値の差が大きくなっていることが確認できる（最大値-1.094，最小値 0.260）。項目 40 は差が 0.344 と大きい。これは SIBTEST においても比較的大きな β 推定量(-0.092)を示していた。これら 10 項目の中で、項目 4, 7, 16, 19, 25 で困難度パラメタ推定値の差が負の値を示し、中国語母語話者集団の方が韓国語母語話者集団よりも困難度は低く、中国語母語話者の方が「できる」と回答しやすく、項目 22, 29, 39, 46, 51 で困難度パラメタ推定値の差が正の値を示し、韓国語母語話者集団の方が中国語母語話者集団よりも困難度は低く、韓国語母語話者の方が「できる」と回答しやすいことが示された。

4.3 Thissen 他の尤度比検定による分析結果

この方法では、DIF 検討項目ごとに基準項目群と合わせたデータ行列を作成して、当該項目に関して下位集団別にパラメタの推定を行わなければならない、基準項目群が確定されていない状況においては、基準項目群を確定するための分析から始める必要がある。このため計算の繰り返しがきわめて煩雑になる。そこで、本研究では、SIBTEST の結果得られた基準項目群に対して、DIF 検討項目とされた 10 項目（項目 4, 7, 16, 19, 22, 25, 29, 39, 46, 51）に対して順次分析を行った。

表 3 は「下位集団間で項目パラメタ値は異なる」というモデル D における DIF 検討項目の項目パラメタ推定値と $-2 \times$ 対数尤度 の値、および、「下位集団間で項目パラメタ値は等しい」というモデル E における $-2 \times$ 対数尤度 の値である。

表 3 Thissen 他のデザインを用いた時の項目パラメタ推定値および対数尤度

		識別力	困難度	モデル E $-2 \times$ 対数尤度	モデル D $-2 \times$ 対数尤度	差
項目 4	韓国	1.176	-0.131	28195.29	28005.25	190.0449
	中国	1.030	-0.802			
項目 7	韓国	1.409	0.065	28146.72	27977.07	169.6516
	中国	1.420	-0.271			
項目 16	韓国	1.132	-0.578	28090.93	27923.56	167.3783
	中国	1.494	-0.824			
項目 19	韓国	0.874	-0.463	28156.72	27929.45	227.2691
	中国	1.270	-1.332			
項目 22	韓国	1.245	-0.493	28151.15	27990.01	161.1394
	中国	1.314	-0.243			
項目 25	韓国	1.165	0.396	28203.09	28028.26	174.8286
	中国	0.963	-0.078			
項目 29	韓国	1.552	0.425	28004.42	27829.94	174.4744
	中国	1.631	0.742			
項目 39	韓国	1.847	-0.801	28021.40	27846.90	174.5005
	中国	1.418	-0.525			
項目 46	韓国	1.755	-0.722	28030.87	27853.76	177.1101
	中国	1.757	-0.382			
項目 51	韓国	1.595	-0.370	28084.42	27918.33	166.0848
	中国	1.764	-0.081			

おおむね、識別力パラメタは下位集団間でほぼ等しいが、困難度パラメタには違いが見

られる。ただし項目 16, 19 および項目 39 は識別力パラメタにも違いが見られ、これらの 3 項目では不均一 DIF が生じている可能性がある。

尤度比検定については、今回の分析では 2 パラメタ・ロジスティック・モデルを採用しているため、検定仮説のもとで対数尤度を 2 倍した結果の差が自由度 2 の χ^2 分布に漸近的に従うことを利用する。その結果、10 項目すべてが 1%水準で有意であった。

困難度を見ると、項目 4, 7, 16, 19, 25 で中国語母語話者集団の方が韓国語母語話者集団よりも困難度は低く、中国語母語話者の方が「できる」と回答しやすく、項目 22, 29, 39, 46, 51 で韓国語母語話者集団の方が中国語母語話者集団よりも困難度は低く、韓国語母語話者の方が「できる」と回答しやすいことが示された。

4.4 Mantel-Haenszel 法による結果

この方法においても、SIBTEST の結果得られた基準項目群に対して、DIF 検討項目とされた 10 項目（項目 4, 7, 16, 19, 22, 25, 29, 39, 46, 51）に対して順次分析を行った。

まず、基準項目群で「できる」と回答した数に基づいて各回答者集団を層に分けて、DIF 検討項目毎に層間に共通なオッズ比の推定値 α_{MH} を計算した上で、さらに対数変換した値 β_{MH} を求めた。 β_{MH} が負の場合中国語母語話者集団が「できる」と回答しやすい傾向、正であれば韓国語母語話者集団が「できる」と回答しやすい傾向になる。さらに、 β_{MH} の標準誤差を計算して、 β_{MH} のこれに対する比を計算した。これらの値は表 4 に示した通りである。

表 4 Mantel-Haenszel 法による結果

	項目番号	$\beta_{MH} = -2.35 \times \ln(\alpha_{MH})$	β_{MH} の標準誤差	比	>2.576
1	4	-1.171	0.208	-5.630	*
2	7	-0.778	0.207	-3.758	*
3	16	-0.785	0.226	-3.473	*
4	19	-1.875	0.257	-7.296	*
5	22	0.583	0.205	2.844	*
6	25	-0.921	0.201	-4.582	*
7	29	0.794	0.246	3.228	*
8	39	0.823	0.234	3.517	*
9	46	0.928	0.232	4.000	*
10	51	0.816	0.223	3.659	*

当該項目で DIF が生じているか否かの判定に関しては、前田・野口(2006)と同様の判定基準を用いた。すなわち、 β_{MH} のその標準誤差に対する比が 2.576 を超える場合に、当該

項目に DIF が生じているという判断をした。これは 1%水準で両側検定を行っていることに相当する。

その結果、DIF 検討項目 10 項目すべてで有意な結果が得られ、DIF 項目と判断された。項目 4, 7, 16, 19, 25 で β_{MH} の値が負で中国語母語話者が「できる」と回答しやすく、項目 22, 29, 39, 46, 51 で β_{MH} の値が正で韓国語母語話者が「できる」と回答しやすいことが示された。

4.5 Logistic 回帰による結果

この方法においても、SIBTEST の結果得られた基準項目群に対して、DIF 検討項目とされた 10 項目（項目 4, 7, 16, 19, 22, 25, 29, 39, 46, 51）に対して順次分析を行った。

ロジスティック回帰の予測変数として、i) 基準項目群における「できる」と回答した数 X のみを入れた場合（モデル 1）、ii) モデル 1 に加えて母語別（中国語か韓国語か）の情報 Q を入れた場合（モデル 2）、iii) モデル 2 に加えて両者の交互作用項 $X \cdot Q$ を入れた場合（モデル 3）、それぞれの $-2 \times \text{対数尤度}$ の値は表 5 に示す通りである。

表 5 Logistic 回帰による結果

	項目番号	モデル1	モデル2	モデル3	モデル1－ モデル3	モデル1－ モデル2	母語の 係数
1	4	667.191	631.631	631.360	35.831	35.560	>0
2	7	618.411	603.625	603.584	14.827	14.786	>0
3	16	572.500	561.858	560.776	11.724	10.642	>0
4	19	642.922	576.207	573.160	69.762	66.715	>0
5	22	626.854	619.569	619.522	7.332	7.285	—
6	25	687.927	667.561	666.182	21.745	20.366	>0
7	29	508.861	497.689	496.047	12.814	11.172	<0
8	39	516.460	503.339	498.938	17.522	13.121	<0
9	46	518.299	499.415	499.387	18.912	18.884	<0
10	51	566.017	551.157	551.015	15.002	14.860	<0

各項目について DIF の判断をするのに、i) まず、モデル 1－モデル 3 の値が自由度 2 の χ^2 分布することを利用して検定する、ii) 次に i) で有意な場合に モデル 1－モデル 2 の値を計算する。この値が自由度 1 の χ^2 分布に従うことを利用して検定する、という手順を踏む。i) で有意でなければ当該項目に DIF はないと判断でき、i) で有意、かつ、ii) で有意な場合には母語別情報 Q のみが関与する DIF（均一 DIF）、ii) で有意でない場合には、交互作用 $X \cdot Q$ も関与する DIF（不均一 DIF）が生じていると判断される。

自由度 2 の χ^2 分布の 1 パーセント点の値は 9.210 であるが、項目 5 を除いて 9 項目が

有意であった。また、自由度 1 の χ^2 分布の 1 パーセント点の値は 6.635 であるが、これら 9 項目のすべてで有意であった。従って、DIF 検討項目の中で項目 22 では DIF は検出されず、残りの 9 項目についてはすべて均一 DIF が検出された。

本研究では母語別情報の変数 Q に対する係数の値が、正ならば中国語母語話者が「できる」と回答しやすいことを、負ならば韓国語母語話者が「できる」と回答しやすいことを表しているが、項目 4, 7, 16, 19, 25 が前者に、項目 29, 39, 46, 51 が後者に該当する。

5. 考 察

5.1 DIF 検出方法間の一致

SIBTEST, BILOG-MG DIF コマンド, Thissen 他の尤度比検定, M-H 統計量, Logistic 回帰の 5 つの方法で DIF 分析を実施した結果を表 6 にまとめて示した。その結果, SIBTEST による分析の結果から DIF 検討項目とされた 10 項目 (項目 4, 7, 16, 19, 22, 25, 29, 39, 46, 51) に関しては、Logistic 回帰の項目 22 を除いて全ての DIF 検出法で一致して DIF が生じている項目と判断できることがわかる。SIBTEST と BILOG-MG DIF コマンドの場合には DIF が生じているとする基準が必ずしも十分な根拠を持って確定できていないという点でありまいさも残してはいるが、これらの場合も他の項目に比べて、DIF 項目と判断するための数値が大きいことは確かである。

表 6 DIF 検討項目に対する各 DIF 検出法の結果

項目 番号	項目内容	DIF 検出法				
		SIBTEST	BILOG- MG	尤度比 検定	M-H 統 計量	Logistic 回帰
4	図書館の本棚にある本の背表紙を見て、必要な本をさがすことができますか。	中国	中国	中国	中国	中国
7	勉強に必要な本や論文を読んでわかりますか。	中国	中国	中国	中国	中国
16	日本語で履歴書がかけますか。	中国	中国	中国	中国	中国
19	封筒やはがきの住所が正しい書き方でかけますか。	中国	中国	中国	中国	中国
22	日本語で日記が書けますか。	韓国	韓国	韓国	韓国	——
25	ワープロ・コンピュータを利用して日本語で文を書くことができますか。	中国	中国	中国	中国	中国
29	サークルやイベントのちらしやパンフレットを作ることが出来ますか	韓国	韓国	韓国	韓国	韓国
39	相手のいいたいことがわからないとき、聞き返すことが出来ますか。	韓国	韓国	韓国	韓国	韓国
46	テレビのドラマがわかりますか。	韓国	韓国	韓国	韓国	韓国
51	授業・講演などを聞いて、全体の流れがわかりますか。	韓国	韓国	韓国	韓国	韓国

しかしながら、この結果から方法間での DIF 項目の検出に関する判断が必ず一致するとは言いきれない。これら 10 項目以外の項目では判断が分かれる可能性が残されている。本研究では基準項目群を確定する作業がきわめて煩雑であるため、5 つの方法すべてで基準項目群を確定する作業をするのではなく、SIBTEST の結果を用いて基準項目群および DIF 検討項目を決定した結果を他の方法にも適用したが、この点について厳密な比較をするためには、各方法について基準項目群を確定する段階から計算過程を繰り返す必要がある。基準項目群が異なった場合には、本研究で DIF 検討項目とした以外に検討する必要がある項目が出てくることもあり得るし、また、DIF 検討項目に対する DIF の判断も異なる可能性もある。

現実には、これらの 10 項目に関しては 5 つの DIF 検出法の間で判断がほとんど一致したことから、少なくとも 2 つ以上の DIF 検出法を適用して得られた結果をもとに内容的な考察に入れば大きな問題が生ずるとは思われない。逆に言うと、単一の方法のみで DIF 分析を実施するのではなく、常に複数の方法を適用して DIF 分析を実施する必要があると考えられる。

5.2 DIF 項目の内容分析

DIF の方向についても、5 つの DIF 検出法の間で一致が見られたが、中国語母語話者にとって、「できる」と回答しやすい項目は 5 項目すべてが、「読む」と「書く」技能の項目である。具体的には、「4 図書館の本棚にある本の背表紙を見て、必要な本をさがすことができますか」、「7 勉強に必要な本や論文を読んでわかりますか」、「16 日本語で履歴書が書けますか」、「19 封筒やはがきの住所が正しい書き方で書けますか」、「25 ワープロ・コンピュータを利用して日本語で文を書くことができますか」である。項目 4 は、本の「背表紙」の題目のうちで「漢字」の部分で探すことができることが、項目 7 についても本や論文の文章中で「漢字」の部分を手がかりに意味を理解することが可能なことが多い、という「漢字」という共通文字が影響していると考えられる。また、項目 16 と項目 19 とは、日本と中国とで、郵便物の住所の書式や履歴書の書式が近いことが影響していると考えられる。項目 25 はワープロでは漢字変換に際して正しい漢字を選ぶ必要があるが、この点では明らかに中国語母語話者にとって有利である。

中国語母語話者にとって有利に働く DIF は、漢字と文化的な共通性の両方が影響しているが、いずれの項目の場合も「漢字」に負うところが大きいと考えられる。

これに対して韓国語母語話者にとって「できる」と回答しやすい項目については、「書く」が 2 項目、「話す」が 1 項目、「聞く」が 2 項目である。逆に見ると、「話す」と「聞く」技能の項目で DIF が検出された項目は、すべて韓国語母語話者にとって「できる」と回答しやすい項目である。具体的には、「22 日本語で日記が書けますか」、「29 サークルやイベントのちらしやパンフレットを作ることが出来ますか」、「39 相手の言いたいことがわからないとき、聞き返すことが出来ますか」、「46 テレビのドラマがわかりますか」、「51 授業・講演などを聞いて、全体の流れがわかりますか」である。項目 46 は、韓国

のテレビ・ドラマが日本のテレビ・ドラマと似た構成をとっており、類推が利きやすいということが、項目 51 は、言語構造が似ているため、授業や講演の内容を逐次的に認知処理することができることが影響していると考えられる。項目 29 は、韓国と日本との学校生活の類似性が高いこと、項目 39 は、韓国語話者の相手との距離のとり方が日本人と似ていて、効果的に聞き返せること、などが理由として考えられるが、これらについてはさらに情報が必要である。項目 22 に関しては、現段階ではその理由は明らかではない。

DIF が生じた原因については、さらに各母語話者の国・地域の文化、歴史的経緯、生活環境など、そして、その国・地域で行われている日本語教育の特徴などの幅広い視点から検討する必要がある。

6. 今後の課題

本研究では、Can-do-statements に含まれる 60 項目の中で、「読む」技能と「書く」技能に関する項目で中国語母語話者が、「話す」技能と「聞く」技能に関する項目で、韓国語母語話者が「できる」と回答しやすい項目が DIF 項目として検出されることが明らかになった。「読む」技能が一般に中国語母語話者にとって「漢字」の影響により、韓国語母語話者よりも一般に「できる」ことが多いことは確かであるが、自己評価による日本語能力水準が同一の回答者間において、中国語母語話者の方が韓国語母語話者よりも「できる」と回答することが多いことを DIF は示しており、なぜ特定の項目で DIF が生じているのかについての原因は、さらに詳細に検討を続ける必要がある。「書く」「話す」「聞く」の各技能についても同様である。

また、自己評価方式の質問紙調査に対する回答結果が、回答者の日本語能力に関してどの程度正確に反映しているのかに関して、本研究とは別の流れの中で検討する必要がある。

さらに、今回は中国語母語話者と韓国語母語話者との違いを DIF 分析で検討したが、非漢字圏、特に欧米系の言語の母語話者や、文法構造が日本語と似ていると言われるトルコ語の母語話者など、より多くの母語話者について研究対象を拡げる必要がある。そうすることにより、DIF 分析の成果を日本語教育の内容・方法・教材・カリキュラムなどに反映させることが可能になる。

以上のことなどが、今後の検討課題として残されている。

文 献

- Bolt, D. & Stout, W. 1996 Differential item functioning: its multidimensional model and resulting SIBTEST detection procedure, *Behaviormetrika*, 23, 67-95.
- 熊谷龍一・脇田貴文 2003 特異項目機能検出方法の比較 -BILOG-MG と SIBTEST を用いた検討-. 名古屋大学大学院教育発達科学研究科紀要, 50, 83-90.
- Lord, F. M. 1980 *Applications of item response theory to practical testing problems*. Hillsdale, NJ:Lawrence Erlbaum Associates.
- 前田忠彦・野口裕之 2006 「法科大学院統一適性試験」の DIF 分析 (適性試験委員会

- 編 法科大学院統一適性試験 テクニカル・レポート 2005 商事法務) 64-86.
- 日本語教育学会試験分析委員会 1999 『Can-do-statements 調査報告』. 国際交流基金
- 日本語能力試験実施委員会・日本語能力試験企画小委員会 2003 平成 13 年度日本語能力試験 分析評価に関する報告書. 国際交流基金・日本国際教育協会
- 野口裕之 2002 IRT モデルによる DIF の検出. 日本心理学会第 66 回大会ワークショップ配布資料
- Roussos, L. A., & Stout, W. F. 1996 Simulation studies of the effects of small sample size and studied item parameters on SIBTEST and Mantel-Haenszel type I error performance. *Journal of Educational Measurement*, 33, 215-230.
- 三枝令子 編著. 2004. 日本語 Can-do-statements 尺度の開発. 科学研究費補助金研究成果報告書.
- Saegusa, R. et al., 1999 Verifying the Validity of the Japanese Language Proficiency Test Using Can-do Statements. The 21st Language Testing Research Colloquium, Tsukuba International Congress Center.
- Shealy, R., & Stout, W. 1993 A model-based standardization approach that separates true bias/DIF from group ability differences and detects test bias/DTF as well as item bias/DIF. *Psychometrika*, 58, 159-194.
- 島田めぐみ、三枝令子、野口裕之 2006 日本語 Can-do-statements を利用した言語行動記述の試みー日本語能力試験受験者を対象としてー 『世界の日本語教育』 79-92, 国際交流基金.
- Stout, W. 1999 *DIFPACK*: Dimensionality-based DIF/DBF package: SIBTEST, Poly-SIBTEST, Crossing SIBTEST, DIFSIM, DIFCOMP [Computer program]. IL: Assessment System Corporation.
- Thissen, D., Steinberg, L., & Wainer, H. 1993 "Detection of differential item functioning using the parameters of item response models." in P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 67-113). Hillsdale, NJ: Lawrence Erlbaum.
- 渡辺直登・野口裕之 1999 組織心理測定論ー項目反応理論のフロンティアー, 白桃書房
- Zimowski, M. F., Muraki, E., Mislevy, R. J., & Bock, R. D. 1996 *BILOG-MG* [Computer program]. IL: Scientific Software International.

附 記

本研究は、日本言語テスト学会第 8 回大会 (2004 年) で口頭発表した研究をもとに、追加分析を実施し、考察を加えたものである。

また、本研究で用いた日本語 Can-do-statements の回答データは、科学研究費補助金基盤研究 (B) (1) 「日本語 Can-do-statements 尺度の開発」 (平成 13~15 年度 代表者: 三枝令子, 課題番号 13480068) の成果の一部である。