

## 2 項テストにおける項目ラベル組み合わせの推定 (実践編)

### Identifying Hidden Category Labels in Dichotomous Tests: Practical Application

安間 一雄\*

Kazuo Amma

Email: ammakazuo@mac.com

前回の論文(安間, 2011)に引き続き, 構成要素の項目数と合計得点のみが得られている 2 項テストの項目を特定する方法論を検討する. 安間(2011)では小規模のサンプルデータを用い, 主成分分析による最小固有ベクトルから手掛かりを得る方法が有効であることを示した. しかしながら項目数が被験者数より大きい, いわゆる過飽和状態のデータに対しては解が得られないことが判明した. そこで今回はステップワイズ法による多重回帰分析を用いたところ, 若干の出力修正により適正な項目推定がなされた. 分析の対象にしたのは最近実施された TOEIC IP テストで, 受験者へのフィードバック事項である 'Abilities Measured' (AM) (技能別達成度診断) を構成する設問項目の特定を行った. これにより得られた項目毎の AM のラベルを基にクラスター分析を行ったところ, 約半数の設問項目が説明可能な範疇を成していることが判った.

Continued from Amma (2011) a rational solution was sought to detect the question items in the category for which only the percentage of the test takers' correct answers are known. The core element of the procedure developed by and described in Amma (2011) was to use a principal component analysis in which the non-null eigenvector scores of the smallest principal component indicate the items to be involved in the category in question. The same procedure was applied to the 'Abilities Measured' (AM) categories in a recent TOEIC IP test. This time, however, the results were not so easily obtained as in the trial version because of the insufficient subject size. All the AM members were obtained in the end with the help of the information of question types already known, but a general solution was beyond our reach. A new technique, then, was applied; a stepwise regression analysis successfully identified the items after a small adjustment. The items obtained were confirmed to be the correct members of the AM categories obtained manually. Additionally, the psychometric relevance of the Abilities Measured was evaluated by comparison with the result of a cluster analysis.

キーワード: 項目分析, TOEIC, Abilities Measured, 主成分分析, ステップワイズ重回帰分析

---

\*: 獨協大学国際教養学部

## 0. はじめに

本報告では、前回の論考<sup>1</sup>でサンプルデータを用いて有効性を提示した分析方法論を実際のテストデータ解析に応用した結果およびそれにより到達できなかった事項に対して新しい解析方法を用いて得た結果を述べ、併せて対象となったテストの範疇分けの妥当性を考察する。

本研究の目的は、一般的な表現では、変数の知られていない一連の試験において特定の範疇の項目合計得点を手掛かりにして当該範疇を構成する変数を特定することにある。具体的には、英語能力テストである TOEIC において設定されている内容別項目範疇である Abilities Measured (AM) に含まれる項目の特定を行った。前稿<sup>1</sup>で用いた主成分分析による項目の特定方法は被験者数が問題項目数より大きい場合にのみ解が得られる。逆の場合、即ち過飽和状態ではこの方法でなくステップワイズ法による重回帰分析を用いることで項目を特定できることが判明した。固有技術即ち問題項目の内容に関する情報を手掛かりとして項目数を減らした主成分分析により得られた項目とステップワイズ法により得られた項目が一致することが確認された。本稿ではこれら 2 つの分析方法を順を追

って説明する。

## 1. 分析対象と条件

TOEIC においては、表 1 に示すようにリスニング問題（第 1～100 項目）に 4 種類、リーディング問題（第 101～200 項目）に 5 種類の内容別項目範疇である Abilities Measured (AM) が設定されている。それぞれの範疇がどの項目により構成されているかは公表されていない。テスト受験者に返されるのはリスニングパート・リーディングパートそれぞれの換算得点と AM 毎の正答率（正答項目数を総項目数で割りパーセンテージ表示にして小数以下を四捨五入した値）である。

過去 5 年以内に実施されたある回の TOEIC IP テストについて別の研究企画のために募集した被験者 79 名の正誤解答データを入手した。また、被験者から提出されたスコアシート記載情報から 9 範疇の AM 正答率を得た。AM 項目総数の推定方法については前稿<sup>1</sup>ですでに説明している。このテストの項目総数は表 2 の通りである。これをもとに個人の範疇毎の正答数を推定した。

[表 1 Abilities Measured (AM) の定義。括弧内は略号]

- L1 短い会話・アナウンス・ナレーションなどの中で明確に述べられている情報をもとに要点・目的・基本的な文脈を推測できる。(L-Sskim)
- L2 長めの会話・アナウンス・ナレーションなどの中で明確に述べられている情報をもとに要点・目的・基本的な文脈を推測できる。(L-Lskim)
- L3 短い会話・アナウンス・ナレーションなどにおいて詳細が理解できる。(L-Sscan)
- L4 長めの会話・アナウンス・ナレーションなどにおいて詳細が理解できる。(L-Lscan)
- R1 文書中の情報をもとに推測できる。(R-Inf)
- R2 文書中の具体的な情報を見つけて理解できる。(R-Scan)
- R3 ひとつの文書の中でまたは複数の文書間でちりばめられた情報を関連付けることができる。(R-Integ)
- R4 語彙が理解できる。(R-Voc)
- R5 文法が理解できる。(R-Gram)

[表 2 Abilities Measured (AM) 毎の総項目数]

L1	L2	L3	L4	R1	R2	R3	R4	R5
L-Sskim	L-Lskim	L-Sscan	L-Lscan	R-Inf	R-Scan	R-Integ	R-Voc	R-Gram
24	23	16	37	16	16	23	29	25

## 2. 主成分分析による項目の特定

この方法においては、範疇正答数と該当項目（リスニング／リーディングの大区分）を対

象として項目間の共分散行列に基づいた主成分分析を行う。最小主成分の固有ベクトル値が 0 でない項目がこの範疇に属する項目である。リ

スニング／リーディングのパートはそれぞれ 100 項目あり、被験者は 79 名しかいないため、このままでは過飽和状態となり意図した解は求まらない。分析対象とする項目には無関係の（範疇に帰属しない）項目が含まれているのでそれを極力対象から外して項目数を被験者数より少なくすれば主成分分析を行うことができる。TOEIC に関する資料<sup>2</sup>からテスト項目のおよその内容が知られている。例えば Part 1（項目番号#1-10）は短文を聞きそれに合う写真を選ぶ問

題、Part 2（#11-40）は対話が適正な談話となるよう応答の発言を選ぶ問題で、いずれも音声テキストを聞いて概要を把握することが求められている。すなわち L1(L-Sskim)もしくは L2(L-Lskim)が関わっている。Part 3（#41-70）は長めの会話、Part 4（#71-100）は長いナレーションもしくは説明文なので L3(L-Sscan)もしくは L2(L-Lscan) が関わっていることがわかっている。表 3 に AM 毎の推定される項目番号を示す。

[表 3 Abilities Measured (AM) 毎の推定項目番号]

L1	L2	L3	L4	R1	R2	R3	R4	R5
L-Sskim	L-Lskim	L-Sscan	L-Lscan	R-Inf	R-Scan	R-Integ	R-Voc	R-Gram
#1-40	#41-100	#1-40	#41-100	#153-200	#153-200	#143-200	#101-200	#101-153

この知識に基づき、リスニング問題については項目#1-40 に限定して L1(L-Sskim)と L3(L-Sscan)の帰属項目を、項目#41-100 に限定して L2(L-Lskim)と L4(L-Lscan)の帰属項目をそれぞれ特定した。リーディング問題については同様の方法で R1(R-Inf), R2(R-Scan), R3(R-Integ) は得られたが、R4(R-Voc), R5(R-Gram) は対象項目を限定できなかつたため帰属項目を特定できなかった。表 4 に項目#1-40 を用いて L1(L-Sskim)の帰属項目を推定した際の主成分分析の因子ベクトル値（最初の 19 項目分）を示す（最大主成分 **Prin1** と最小主成分 **Prin40** のみ表示、この間の主成分値は省略）。分析には統計ソフト *JMP* (version 10)<sup>3</sup>を使用した。

L014	0.00506	0
L015	0.04952	0.20412
L016	0.01795	0
L017	0.04594	0.20412
L018	0.01637	0
L019	0.03272	0.20412
L020	0.01527	0

L1(L-Sskim)の場合、得られた帰属項目（**Prin40** の列で値が 0 でなくかつ合計得点である範疇項目 L-Sskim でない項目）の数は 23 であるが、全員が正答したため除外した項目 #001 (=L001)を加えると表 2 に示す総項目数 24 と一致する。

さらに表 5 に項目#101-200 を用いて R4 (R-Voc) の帰属項目を推定した際の主成分分析の因子ベクトル値（最大主成分 **Prin1** と最小主成分 **Prin101** のみ表示、この間の主成分値は省略、最初の 20 項目分）を、表 6 に項目 #101-200 を用いて R5 (R-Gram) の帰属項目を推定した際の主成分分析の因子ベクトル値（最大主成分 **Prin1** と最小主成分 **Prin101** のみ表示、この間の主成分値は省略、最初の 20 項目分）をそれぞれ示す。これらは被験者数が変数数（設問項目数）より少ないいわゆる過飽和状態のデータであるため、因子ベクトル値が 0 に収束する項目はない。

[表 4 L1 (L-Sskim) の因子ベクトル値]

Variable	Prin1	Prin40
L-Sskim	0.96513	-0.20412
L002	0.01014	0.20412
L003	0.01383	0.20412
L004	0.04419	0.20412
L005	0.0106	0.20412
L006	0.02833	0.20412
L007	0.00257	0.20412
L008	0.0158	0.20412
L009	0.03641	0.20412
L010	0.00401	0
L011	0.0307	0
L012	0.02134	0.20412
L013	0.0268	0

[表 5 R4 (R-Voc) の因子ベクトル値]

Variable	Prin1	Prin101
R-Voc	0.94967	-0.04809
R101	0.031	0.03642
R102	0.02673	0.044
R103	0.02716	0.03535
R104	0.03765	0.08305
R105	0.02962	0.02863
R106	0.01671	-0.11154
R107	0.00907	-0.16575
R108	0.0233	0.138
R109	0.03361	0.01321
R110	0.00755	0.01594
R111	0.03171	0.06843
R112	0.04235	0.19445
R113	0.04133	-0.08749
R114	0.0378	0.07294
R115	0.04855	0.15057
R116	0.04184	0.14826
R117	0.00398	0.11686
R118	0.02828	-0.07945
R119	0.02707	0.10796
R120	0.04122	-0.03453

[表 6 R5 (R-Gram) の因子ベクトル値]

Variable	Prin1	Prin101
R-Gram	0.94757	-0.05943
R101	0.04436	-0.00819
R102	0.02204	0.01513
R103	0.04473	0.00465
R104	0.03808	0.04474
R105	0.0399	0.12524
R106	0.01998	-0.10008
R107	0.00766	-0.13692
R108	0.0342	0.11871
R109	0.04723	0.02342
R110	0.01777	0.05738
R111	0.02262	0.07503
R112	0.05021	0.15468
R113	0.03826	-0.02625
R114	0.02271	-0.04004
R115	0.05062	0.25647
R116	0.02984	0.09279
R117	0.00914	0.02942
R118	0.03262	-0.06452
R119	0.03118	0.14157
R120	0.05017	0.01999

ここで範疇項目と各設問項目の相関係数に注目してみる。当該範疇に帰属する項目の合計得点が範疇項目の値になるので両者の間にはある程度の相関関係があることが期待される。相関係数が低い項目を除外して主成分分析の対象項目数を被験者数より少なくすることができれ

ば、主成分分析を適用することができるはずである。この可能性を探るためすでに結果がわかっている範疇 L1(L-Sskim) に対するリスニング問題全項目 (#1-100) の相関係数を観察した (表 7, 値の低い順にソートしてある。また, 上位 20 項目と下位 20 項目を表示)。表中第 1 列 (Category) に既知の範疇帰属項目 (=1) を示す。この範疇の場合, 相関係数値が下位から 6 個目の項目 (L007) が帰属項目であることからわかるように, 相関係数値が低い項目を安全に除外することはできない。この段階での結論として, 主成分分析の適用には限界があり, 被験者数を変数数より少ない場合には解が求まらないことが判明した。

[表 7 L1 (L-Sskim) の範疇項目と設問項目の相関係数]

Category	Item	L-Sskim
	L062	-0.0829
	L047	-0.0042
	L043	0.0339
	L076	0.0366
	L010	0.0512
1	L007	0.0574
	L014	0.0812
	L075	0.092
	L042	0.1057
	L054	0.1257
	L099	0.1358
	L020	0.1373
	L025	0.1493
	L084	0.155
1	L002	0.1556
	L073	0.1597
	L068	0.1632
	L063	0.1668
	L016	0.1718
	L069	0.1742
	.....	
1	L015	0.4133
	L081	0.4151
	L097	0.4176
	L055	0.4183
1	L040	0.4226
	L050	0.4267
1	L023	0.4275
	L100	0.4347
	L045	0.4385
	L080	0.4403
	L077	0.4477
1	L009	0.4512
	L059	0.4528
	L093	0.4557

	L066	0.4671
1	L004	0.4794
	L034	0.4812
1	L029	0.5092
1	L024	0.5673
1	L038	0.5831

目を目的変数，設問項目を説明変数とし，編入・排出基準をそれぞれ  $p < 0.2$  と  $p > 0.2$  に設定して双方向（編入・排出）で変数の特定を行った．範疇 R4(R-Voc) に対する分析結果を表 8（最初の 20 項目分）に示す．Entered 欄に X が記された項目（Intercept を除く）が帰属推定項目である．

### 3. 重回帰分析（ステップワイズ法）による項目の特定

過飽和状態における解を得るため，ステップワイズ法による重回帰分析を行った．範疇項

[表 8 R4 (R-Voc) に対するステップワイズ法による範疇帰属項目推定]

Lock	Entered	Parameter	Estimate	nDF	SS	F Ratio	Prob>F
X	X	Intercept	-1.65E-15	1	0	0	1
		R101	0	1	8.68E-30	0	1
	X	R102	1	1	2.734714	1.83E+14	0.00E+00
		R103	0	1	4.00E-28	0	1
	X	R104	1	1	12.19668	8.18E+14	0
		R105	0	1	1.96E-27	0	1
		R106	0	1	1.10E-27	0	1
	X	R107	1	1	4.385759	2.94E+14	1e-314
		R108	0	1	4.35E-28	0	1
		R109	0	1	6.07E-29	0	1
		R110	0	1	2.28E-29	0	1
	X	R111	1	1	10.83946	7.27E+14	0
		R112	0	1	4.36E-30	0	1
		R113	0	1	7.36E-28	0	1
	X	R114	1	1	10.11513	6.79E+14	1e-323
		R115	0	1	2.01E-27	0	1
	X	R116	1	1	11.61044	7.79E+14	0
	X	R117	1	1	6.930193	4.65E+14	1e-319
		R118	0	1	6.41E-28	0	1
		R119	0	1	2.84E-28	0	1
		R120	0	1	3.03E-28	0	1

この方法により推定された非帰属項目 #152-180 を除外して，主成分分析を行った結果（表 9，最初の 20 項目分），範疇帰属項目はステップワイズによる推定項目と全て一致した．

R109	0.03395	0
R110	0.00771	0
R111	0.03238	0.18257
R112	0.04284	0
R113	0.04209	0
R114	0.03856	0.18257
R115	0.04897	0
R116	0.04208	0.18257
R117	0.00399	0.18257
R118	0.02845	0
R119	0.0273	0
R120	0.04155	0

[表 9 R4 (R-Voc) の因子ベクトル値]

Variable	Prin1	Prin101
R-Voc	0.95976	-0.18257
R101	0.0312	0
R102	0.0271	0.18257
R103	0.02732	0
R104	0.03775	0.18257
R105	0.03021	0
R106	0.01676	0
R107	0.00922	0.18257
R108	0.02339	0

範疇 R5 (R-Gram) に対しても同じ方法を用いて帰属項目を特定することができた（表 10，最初の 20 項目分），Entered 欄に X が記された項目（Intercept を除く）が初期帰属推定項

目である。但しこの範疇の場合は推定値が極端に小さい項目（R102 など）を除外することで帰属項目が得られた。

[表 10 R5 (R-Gram) に対するステップワイズ法による範疇帰属項目推定]

Lock	Entered	Parameter	Estimate	nDF	SS	F Ratio	Prob>F
X	X	Intercept	-3.99E-13	1	0	.	.
	X	R101	1	1	1.31E-01	.	.
	X	R102	1.57E-13	1	8.74E-28	.	.
	X	R103	1	1	3.29E-01	.	.
		R104	0	0	0.00E+00	.	.
	X	R105	1	1	3.80E-01	.	.
	X	R106	1	1	1.82E-01	.	.
	X	R107	-1.43E-13	1	4.04E-27	.	.
	X	R108	1	1	2.28E-01	.	.
	X	R109	1	1	4.95E-01	.	.
	X	R110	1	1	3.43E-01	.	.
	X	R111	1.84E-13	1	9.28E-27	.	.
	X	R112	1	1	0.608286	.	.
	X	R113	1	1	1.09E+00	.	.
	X	R114	1.33E-13	1	1.02E-26	.	.
	X	R115	1	1	2.08E-01	.	.
	X	R116	8.63E-14	1	1.23E-27	.	.
		R117	0	0	0.00E+00	.	.
	X	R118	1	1	1.58E-01	.	.
	X	R119	1	1	4.42E-01	.	.
	X	R120	1	1	2.11E-01	.	.

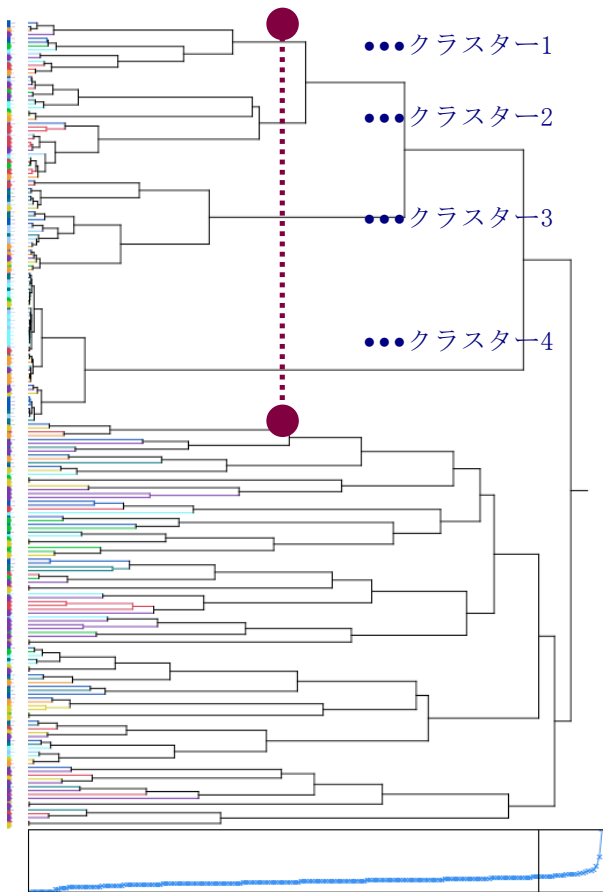
#### 4. 範疇帰属項目特定方法のまとめ

過飽和状態のデータにも適用できる一般的な解法はステップワイズ法による重回帰分析であることが確認された。しかしながらデータによっては推定値の誤差を容認する場合があるので初期推定項目を調整する必要がある。被験者数を多くして過飽和状態を解消し主成分分析を用いることができればそれに越したことはない。

的事実とは離れた恣意的な範疇である可能性がある。そこで、今回得られた設問項目毎の AM のラベルを用いて設問 200 項目に対するクラスター分析を試みた。分析には同じく統計ソフト JMP (version 10) を使用した。結果を図 1 に示す。

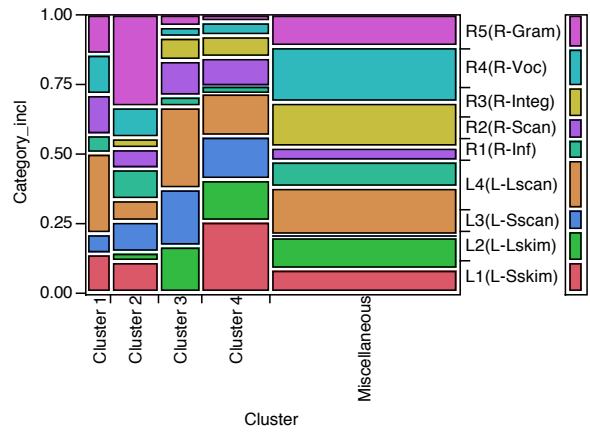
#### 5. 考察

ここで範疇指定の妥当性を考える。表 7 で見たように、範疇項目と設問項目の相関係数は当該範疇への帰属を推定する根拠としては非常に弱い。相関係数の低かった L007 や L002（それぞれ 0.057 と 0.156）は範疇の合計得点に対する寄与が低かったと言える。実際、これらの項目の正答率 FV はそれぞれ 0.95 と 0.97 と高かったが、それより高い L005 (FV=0.99) や L003 (FV=0.97) は相関係数は低くない（それぞれ 0.310 と 0.288）。言い換えれば、これらの項目は、指定された範疇が想定する構成概念とは異なる言語事項をテストしていると想定される。もとより AM はテスト作成者が指定した所与のものであるので、心理測定



【図1 TOEIC IP テスト設問項目に対するクラスター分析】

全体として、クラスターが明確に組織される項目（図の上半分）とランダムな項目（下半分）に大別される。上半分については大まかに4個のクラスターに分かれると読み取れる（縦破線位置）。次に、これら4個のクラスターを本図上方より「クラスター1 (Cluster 1)」、 「クラスター2 (Cluster 2)」、 「クラスター3 (Cluster 3)」、 「クラスター4 (Cluster 4)」と名付け、これらのクラスターにおけるAMの構成配分をモザイク図に表示した（図2）。下半分のクラスターは「その他 (Miscellaneous)」とした。縦軸が各クラスター内のAMの配分比を表す。



【図2 TOEIC IP テストのクラスターにおけるAMの構成配分】

各クラスターにつきおよそ次の特徴が観察される。

- (1) クラスター1：L-Lscan の配分比が大きいため、リスニングの長いテキストの処理に特有の能力が関わると考えられる。ワーキングメモリと関連している可能性がある<sup>4</sup>。
- (2) クラスター2：リーディング能力、特に文法と語彙に関する能力が関わる。知識とその運用が関連していると考えられる。
- (3) クラスター3：L-Sscan, L-Lscan, R-Scanの配分比が大きいため、リスニング・リーディング共にスキミングすなわちテキスト内から特定の情報を検索・照合する能力が関わると考えられる。
- (4) クラスター4：リスニングの配分比が大きいためリスニング全般の能力が関わると考えられる。その中でも特にL-Sskimの配分比が大きいが、これにはPart 1の写真判別問題 (#1-10) が多く含まれることから同種の問題項目でクラスターを成していると言える。

クラスター間で帰属項目数が大きく隔たらないことはバランスの取れた問題作成が行われているものとして評価できる。しかしながら項目の半数はクラスターによる説明が困難なもので、上記観察は部分的な項目の傾向を説明するに過ぎない。今後はより客観的な心理測定的事実に基づいたAMの設定が求められる<sup>5</sup>。

## 注

- \* 本研究は 2011-2012 年度獨協大学特別研究助成（共同研究者：安間一雄・中谷安男・日野克美）により支援を受けて実施された。

## 参考文献

- 1) 安間一雄： 2 項テストにおける項目ラベル組み合わせの推定. 獨協大学情報科学研究, 28, 65-72 (2011).
- 2) 財団法人国際ビジネスコミュニケーション協会 TOEIC 運営委員会： TOEIC®テスト新公式問題集 Vol.4. 財団法人国際ビジネスコミュニケーション協会 TOEIC 運営委員会, 9 (2009).
- 3) *JMP* (version 10.0). Cary, NC: SAS Institute, Inc. (2012).
- 4) Miyake, A. and Friedman, N. P.: Individual differences in second language proficiency: working memory as “language aptitude”. In A. F. Healy & L. E. Bourned (eds.), *Foreign Language Learning: Psycholinguistic Studies on Training and Retention*. Lawrence Erlbaum Associates, New Jersey, 339-364 (1999).
- 5) 安間一雄： 標準テストを授業学習に連繫させる：TOEIC スコアの有効活用. 外国語教育研究, 29, 17-35 (2011).

(2012 年 9 月 21 日受付)

(2012 年 12 月 19 日採録)