

ベイジアンネットワークモデルを用いた 英語の比較変化の習得に関する検討

萩原 康 仁・山 森 光 陽

1. 問題

日本語を母語とする日本人が英語を学習するにあたっては、日本語にはみられない特有の文法体系を習得しなければならない。その中でも、名詞・代名詞・形容詞の性・数・格による変化および動詞の人称・数・時制・法・態による活用などの語形変化の習得は、多くの学習者にとって困難を感じせしめるものであると言えよう。特に、これらの習得は、日本の公教育における英語教育課程においては、中学校でなされるべきものであることが、現行の学習指導要領（文部省，1998）によって公示されている。

この学習指導要領による言語材料の取扱いのうち、中学校で指導すべき文法事項は、表1の通りである。現行の学習指導要領においては学年別の配当について定めはないが、検定済教科書7種では、おおよそ表1の通りに配当されている。この表を検討すると、文法事項という観点に限ってみれば、2年生が最も配当が多い。さらに、語形変化という観点から見ると、2年生においては、受動態と現在完了の学習に必要な、動詞の変化を覚えなければならないばかりか、比較変化も覚えることが必要となる。一方、1年生では3人称単数現在や過去形、現在分詞の変化が教育課程に含まれているが、3年生において語形変化は主たる対象とならない。そのため、2年生の英語学習においては、学習者の記憶の負担は大きいと言えよう。

そこで、これらの状況を鑑み、学習者に対する負担の大きい語形変化の学習について、指導に当たる者はより適切な指導方法を検討することが必要である。そのためには、単に定期試験などのテストによって求められた点数としての学習成果だけではなく、学習者の持つ知識構造がどのようになっているのか、また、指導によってそれがどのように変化するのを知りたいところである。そこで本研究では、中学校2年生に配当されている語形変化の単元である、比較における学習者の知識構造について取り扱った。

次に、知識構造を検討する一方法としての心理測定学的方法を概観する。わが国においても、項目応答理論（Item Response Theory, IRT）に基づいたコンピュータ適応型テスト（Computerized Adaptive Testing, CAT）が実用化されはじめた。このテストの特徴は、困難度などの各項目の母数を、IRTを用いて予め推定しておき、これを既知とした上で、受験者の前問までの解答パターンから推定された暫定的な能力水準に適した困難度を持つ項目を選択・呈示し、その解答結果から受験者の能力

母数を瞬時に更新するといった手続きを繰り返すことで、精度の高い能力母数の推定を効率的に行うことができることにある。CATでは、受験者によって異なった項目群が呈示されることになるが、推定された能力母数は、項目母数が同一尺度上で目盛りづけられているので、やはりその尺度上で解釈することができる。林（2001）は、自らが開発した英語能力を測定するCATを紹介し、従来の紙筆テストに比べて短時間で効率よく個人の能力を測定できたと報告している。IRTに基づいたCATで各受験者にフィードバックされる情報は通常、推定された能力母数を見やすいように整数の値で変換したスコアが主となる。このため、CATが受験者の配置や選抜に用いられる場合は効果的であるが、その受験者は一体何ができて何ができなかったのかという、質的な情報については乏しいため、診断的な評価をするのには向かないと考えられる。

一方で、近年のテスト理論の分野では、受験者の質的な評価をするための研究もなされている。例えば、Doignon & Falmagne（1999）は、テスト項目間の関連を、knowledge structureという部分的な包含関係で表し、それを確率的にモデル化する方法を提唱した。また、Doignon & Falmagne（1999）は、このモデルを用いて、各受験者の当該分野におけるknowledge stateを明らかにするための効率的な手法を紹介した。これは、IRT以外の理論を基にした適応型テストの雛型であると考えられる。なぜなら、IRTを基にした適応型テストでは、次問として、受験者のそれまでの解答パターンから推定された暫定的な能力母数に近接した困難度が付与された項目を呈示することで、推定精度を確保したまま受験者の能力母数を効率的に推定することが目的となるのと同様に、Doignon & Falmagne（1999）の手法を基にした場合は、受験者のそれまでの解答パターンから、各knowledge stateに属する確率を暫定的に推定し、次問として、いずれかのknowledge stateにその受験者が属する確率が最も高くなるように更新されると期待される項目を呈示することによって、受験者の属するknowledge stateを効率的に推定することが目的となるからである。また、Tatsuoka（1995）は、各テスト項目を解くのに必要とされる要件をattributeと呼び、当該テスト分野における諸々のattribute×テスト項目という（要素が0-1である）Q-Matrixを形成し、これを用いて、「このattributeを有していないときはこの解答パターンが得られる」といった理想的な解答パターンを形成する方法を唱えた。またTatsuoka（1995）は、IRTで推定された各受験者の能力母数と、その受験者が示した解答パターンとIRTモデルとの間の乖離度を利用して、実際の各受験者の解答パターンが、Q-Matrixから得られる理想的な解答パターンの内、どれに最も近いかを分類するための方法を提唱した。これをRule Spaceの手法という。Tatsuoka（1995）は、この手法を用いることで、各受験者について、分類された理想的な解答パターンから引き出される各attributeの有無とIRTにおける能力母数が得られることによって、診断的な評価と総括的な評価をすることができるとしている。さらに椎名（1997）は、潜在クラス分析を応用した手法を用いて、小学校4、5年生を対象とした4桁割る2桁の6問からなる割り算のテストの解答状況から、割り算を解くのに必要な下位スキルの内、各児童がどの下位スキルを習得している確率が小さい、すなわち苦手になっているのかを推定する方法を述べた。また椎名（1997）は、「掛け算はできるのに繰り下がりのある引き算は苦手にしたままである」とい

表 1. 中学校各学年における文法事項の配当の現状

文法事項 (中学校学習指導要領(平成10年3月)第2章9節「外国語」による)		配当学年		
		1	2	3
(ア) 文				
a	単文, 重文及び複文	○	○ 重文・ 複文	○ 重文・ 複文
b	肯定及び否定の平叙文	○		
c	肯定及び否定の命令文	○		
d	疑問文のうち, 動詞で始まるもの, can, do, may などの助動詞で始まるもの, or を含むもの及び how, what, when, where, which, who, whose, why の疑問詞で始まるもの	○		
(イ) 文型				
a	[主語+動詞] の文型	○		
b	[主語+動詞+補語] の文型のうち,	名詞 (a) 主語+ be 動詞+代名詞 形容詞	○	
		(b) 主語+ be 動詞以外の動詞+名詞 形容詞	○	
c	[主語+動詞+目的語] の文型のうち,	名詞 代名詞 (a) 主語+動詞+動名詞 how (など) to 不定詞 that で始まる節	○	○
		(b) 主語+動詞+ what などで始まる節		○
d	[主語+動詞+間接目的語+直接目的語] の文型のうち,	(a) 主語+動詞+間接目的語+名詞 代名詞		○
		(b) 主語+動詞+間接目的語+ how (など) to 不定詞		○
e	[主語+動詞+目的語+補語] の文型のうち,	(a) 主語+動詞+目的語+名詞 形容詞		○
f	その他の文型のうち,	(a) there + be 動詞+～	○	
		(b) It + be 動詞+～ (+ for ～) + to 不定詞		○ ○
		(c) 主語+ tell, want など+目的語+ to 不定詞		○
(ウ) 代名詞				
a	人称, 指示, 疑問, 数量を表すもの	○		
b	関係代名詞のうち, 主格の that, which, who 及び目的格の that, which の制限的用法の基本的なもの			○
(エ) 動詞の時制など				
	現在形, 過去形, 現在進行形,	○		
	過去進行形,		○	
	現在進行形		○	○
	及び助動詞などを用いた未来表現		○	
(オ)	形容詞及び副詞の比較変化		○	
(カ)	to 不定詞のうち基本的なもの		○	
(キ)	動名詞のうち基本的なもの		○	
(ク)	現在分詞及び過去分詞の形容詞としての用法			○
(ケ)	受け身のうち現在形及び過去形		○	○

(注) 多くの教科書では, 2年生に現在完了形, 3年生に受け身が配当されているが, その逆に配当されているものもある。

う潜在クラスを仮定したモデルの方がデータとの適合が良かったという分析結果から、小学校4、5年の児童は、小学校の低学年における「引き算を習ってから掛け算を習う」というカリキュラムとは異なったスキルの横断的発達の様相を呈している可能性を示唆した。

本研究では、比較変化における学習者の知識構造を検討するために、ベイジアンネットワークモデルを用いた。ベイジアンネットワークモデルとは、現象の因果関係を、確率変数をノードとする有向非循環グラフとして表現し、ノード間の関係を条件つき確率として示すモデルである。このモデルの特徴は条件つき独立にある。すなわち、親ノードの状態を所与とするとき、当該ノードの反応は他のノードの状態とは独立となる（繁樹, 1999）。このモデルを適用した例として、Mislevy (1995) がある。Mislevy (1995) は、分数の引き算を解くのに必要となる下位スキルを特徴づける潜在クラス変数と、各潜在クラス変数の影響を受ける顕在変数との間の関係を有向非循環グラフで表現した。本研究では、有向非循環グラフを用いて、授業内で行われた比較の語形変化クイズ結果の一部から、その習得の様相を適切に表現することを試みた。

2. 方法

東京都内の公立中学校在籍の2年生95名を対象に、授業時間内の10分間で行われた。クイズの内容は、One World 2（教育出版）の巻末に掲載されている比較変化の表から20単語を取り出し、それぞれ原形を示し、比較級・最上級を書かせるものであった。

3. 分析と結果

各単語とも、比較級と最上級の両方が書けることで正答とみなされた。クイズの各項目内容と、項目正答率、および双列相関係数を表2に示す。次に、これらから単に語尾に-erや-estをつける語形変化以外の5種類の変化について、本クイズ内で初出の単語を取り出し、2つのベイジアンネットワークモデルを構成した。5項目のみをモデル構成に使用した理由は、標本数の少なさにある。すなわち、項目数を増やすと、何らかの制約を入れない限り推定すべき母数の数も増えるが、小標本でこれを推定することは困難だということである。構成したモデルのひとつは、語形変化の種類にかかわらず、比較変化の習得がなされているか否かを表現したもの（図1）であり、もうひとつは、語尾屈折の習得とそうでない語形変化の習得を分け、前者が後者に先行することを表現したモデル（図2）であった。両モデルとも、1つの潜在変数に設けたクラス数は2であった。すなわち、各潜在クラス変数内では、習得状況の有無が表された。ここで、モデルの数理的表現と母数となる（条件つき）確率の推定に用いたアルゴリズムについて述べる。表記は、Vermunt & Magidson (2003) に倣った。

モデルの数理的表現について述べる。図1のモデルで表される各解答パターンの同時確率は、

$$f(y_1 y_2 y_3 y_4 y_5) = \sum_{x_1} \pi(x_1) \prod_{k=1}^5 f(y_k | x_1) \quad (1)$$

表2. 比較級/最上級クイズの基本統計量
(比較級と最上級共に正答で正解とした。太斜体字の項目をモデル構築に使用した。)

ITEM	WORD	解答者数	正答人数	正答率	双列相関
Q1(Y4)	<i>important</i>	95	60	0.632	0.941
Q2(Y1)	<i>happy</i>	95	57	0.600	0.918
Q3	fast	95	81	0.853	1.220
Q4(Y5)	<i>good</i>	95	50	0.526	0.804
Q5	tall	95	80	0.842	1.244
Q6(Y2)	<i>red</i>	95	51	0.537	0.781
Q7(Y3)	<i>nice</i>	95	72	0.758	0.923
Q8	hard	95	79	0.832	1.221
Q9	long	95	75	0.789	1.046
Q10	popular	95	62	0.653	0.955
Q11	little	95	33	0.347	0.677
Q12	small	95	76	0.800	1.077
Q13	big	95	72	0.758	1.110
Q14	many	95	55	0.579	0.944
Q15	high	95	80	0.842	1.244
Q16	fine	95	75	0.789	1.159
Q17	difficult	95	65	0.684	1.028
Q18	easy	95	65	0.684	1.057
Q19	beautiful	95	62	0.653	0.941
Q20	hot	95	66	0.695	0.996

と表される。ここで、 $\pi(x_1)$ は、潜在クラスの確率、 $f(y_k | x_1)$ は、潜在クラスが所与のときの条件つき確率である。同様に、図2のモデルで表される各解答パターンの同時確率は、

$$f(y_1 y_2 y_3 y_4 y_5) = \sum_{x_1} \sum_{x_2} \pi(x_1) \pi(x_2 | x_1) \prod_{k=1}^3 f(y_k | x_1) \prod_{k=4}^5 f(y_k | x_2) \tag{2}$$

と表される。

次に、母数の推定方法について述べる。解答パターン*i*が固定されたときの尤度は、ベクトル表記を用いて、

$$L = \prod_i f(y_i | \theta)^{n_i} \tag{3}$$

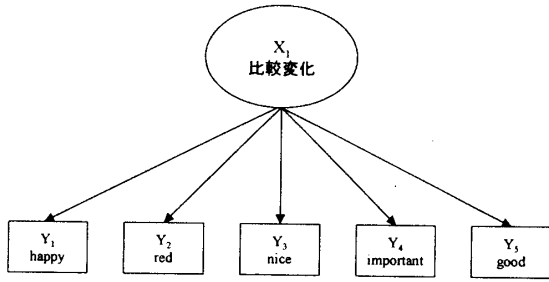


図1. 潜在変数が1つの場合の有向非循環グラフによる表現

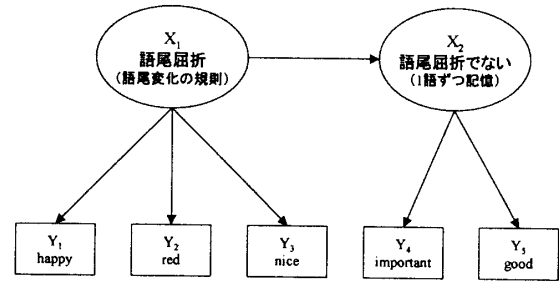


図2. 潜在変数が2つの場合の有向非循環グラフによる表現

と表される。ここで、 $f(y_i | \theta)$ は母数ベクトルが所与のときの、解答パターン i の同時確率

$$f(y_i | \theta) = \sum_{\mathbf{x}} \pi(\mathbf{x} | \theta) f(y_i | \mathbf{x}, \theta) \quad (4)$$

である。(3)式の対数は、

$$\log L = \sum_i n_i \log f(y_i | \theta) \quad (5)$$

と表されるが、観測されない潜在変数ベクトル \mathbf{x} があるため不完全データの対数尤度である。

(5)式は、解答パターン i を示した n_i 人が各潜在クラスのパターンに属する期待人数

$$n_{ix} = n_i \pi(\mathbf{x} | y_i, \theta) = n_i \frac{\pi(\mathbf{x} | \theta) f(y_i | \mathbf{x}, \theta)}{f(y_i | \theta)} \quad (6)$$

を求めることで擬似的に完全データの対数尤度として扱うことができ、

$$\log L^C = \sum_i \sum_{\mathbf{x}} n_{ix} \log f(y_i | \theta) = \sum_i \sum_{\mathbf{x}} n_{ix} \log \pi(\mathbf{x} | \theta) f(y_i | \mathbf{x}, \theta) \quad (7)$$

と表される。完全データの対数尤度を、

$$\frac{\partial \log L^C}{\partial \theta} = \sum_i \sum_{\mathbf{x}} n_{ix} \frac{\partial \log \pi(\mathbf{x} | \theta) f(y_i | \mathbf{x}, \theta)}{\partial \theta} = 0 \quad (8)$$

と最大化することで、母数が推定される。具体的な方法としては、EMアルゴリズムがある。 ν 回目のEMアルゴリズムのEステップでは、

$$\hat{n}_{ix}^{\nu} = n_i \pi(\mathbf{x} | y_i, \hat{\theta}^{\nu-1}) = n_i \frac{\pi(\mathbf{x} | \hat{\theta}^{\nu-1}) f(y_i | \mathbf{x}, \hat{\theta}^{\nu-1})}{f(y_i | \hat{\theta}^{\nu-1})} \quad (9)$$

のように、 $\nu-1$ 回目のMステップで得られた暫定的な母数ベクトルを利用して、期待人数を更新する。 ν 回目のEMアルゴリズムのMステップでは、 ν 回目のEステップで得られた期待人数を用いて、完全データの対数尤度を最大化する。ただし、本研究では、母数の推定にLEMWIN (Vermunt,

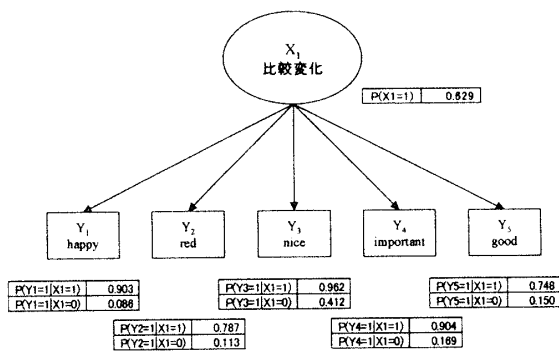


図3. 図1のモデルの推定された(条件つき)確率
 $(\chi^2_{(20)}=24.908, p=0.205, AIC=-15.092)$

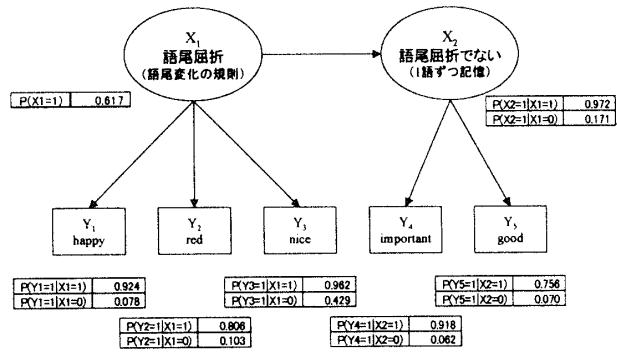


図4. 図2のモデルの推定された(条件つき)確率
 $(\chi^2_{(20)}=17.890, p=0.463, AIC=-18.110)$

1997) を用いたが、このプログラムが採用しているアルゴリズムでは、現状での期待対数尤度よりも ν 回目の M ステップで得られる期待対数尤度の方が大きくなっていけば、 $\nu+1$ 回目の E ステップへ移行する。すなわち各回の M ステップにおける反復計算の回数は 1 である。これを特に、GEM アルゴリズム (Little & Rubin, 1987) という。

本研究の対象となった学習者は、語尾屈折による語形変化をする単語とそうでないものとに二分して覚えることを指導されていたため、図2の方がデータとの適合が良いという仮説を立てていた。図1, 図2で示されたモデルの母数の推定結果をそれぞれ図3, 図4に示す。標本数は少ないものの、両モデルの AIC を比較した結果、図2のモデルが採択された。ここで、図3と図4にある推定された母数について比較すると、nice 以外の単語はいずれも、図3より図4の方が、各潜在クラス変数が示す種類の語形変化を習得していることを所与としたときの項目正答確率が大きくなっていて、それを習得していないことを所与としたときの項目正答確率が小さくなっていることが分かる。このことは、潜在クラス変数の個々の顕在変数に与える影響力は、図4のモデルの方が大きいことを表している。一方で nice においては、図4の語尾屈折を習得していなくとも正答する確率が、図3の語形変化を習得していなくとも正答する確率よりも大きくなっているため、特異な項目とみなせる。この結果が生じた理由は、図4における語尾屈折が測定する nice 以外の 2 単語の方が、nice よりも低い正答率だったことにあると考えられる。すなわち、図4における語尾屈折の困難度が、図3における語尾変化のそれより高いため、語尾屈折を習得していなくとも、正答率の高い nice は正答できるということである。

次に、採択されたモデルの妥当性を検証した。まず、解答パターン i を示した学習者の、図2におけるそれぞれの潜在クラスのパターンに属する確率を、ベイズの定理

$$\pi(x | y_i, \hat{\theta}) = \frac{\pi(x | \hat{\theta}) f(y_i | x, \hat{\theta})}{f(y_i | \hat{\theta})} \tag{10}$$

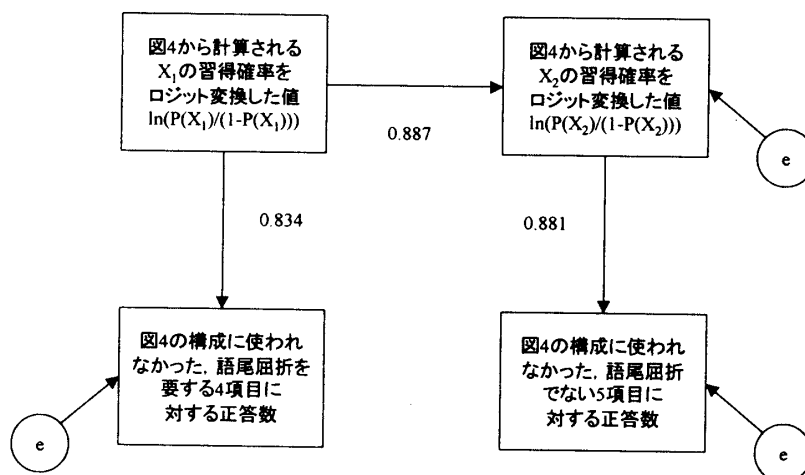


図5. 採択されたモデル(図2)の検証用のモデルのパス図による表現と標準化解
 $(\chi^2_{(3)}=4.140, p=0.246, CFI=0.999, RMSEA=0.033)$

を用いて計算した。(10)式では、図4で示した推定された母数を用いていることに留意されたい。それぞれの潜在クラスのパターンに属する確率から、各学習者について、「少なくとも語尾屈折は習得している確率 $P(X_1)$ 」と「少なくとも語尾屈折でない語形変化は習得している確率 $P(X_2)$ 」を求め、それぞれをロジット変換した。次に、図2の構成に使われなかった語尾屈折を要する4項目に対する正答数と、図2の構成に使われなかった語尾屈折でない5項目に対する正答数を求め、これら4変数をすべて連続変数として扱い、図2と同じ構造を持ったパス図を描き、パス解析を行った。妥当性検証用のモデルと標準化解、ならびに適合度指標を図5に示す。図5のモデルの、データとの適合は良く、図2のモデルの妥当性が部分的には示された。

4. 考察

まず、本研究において検討した、比較変化の習得に対するベイジアンネットワークモデルそのものについて考察を行い、次に、本研究の結果によって得られる教育的含意について論じる。

第1に、本研究の結果、実際に行われた教授処理に沿ったモデルが、データとの適合が良いということが示された。このことは、他の教授処理がなされた場合は、それに合致したモデルが良い適合を示す可能性を示唆している。本研究は標本数の少なさや、1校の生徒のみを対象をしているという問題点があるが、今後は教授法ごとにまとめられた比較的大規模な集団に対して同様の調査を行うことにより、教授法と学習者の知識形成との関係の安定した表現が可能となるであろう。そのためには、教授法に関する情報の共有化が求められる。

第2に、本研究で用いられたモデル構築とその検討の方法は、指導と評価の一体化という観点から見ても有益であると言える。児童生徒の学習と教育課程の実施状況の評価の在り方についての、教育課程審議会答申(教育課程審議会, 2000)は、これからの評価においては、観点別学習状況の評価を基本とした現行の評価方法を発展させ、目標に準拠した評価(いわゆる絶対評価)を一層重視すると

ともに、児童生徒一人一人のよい点や可能性、進歩の状況などを評価するため、個人内評価を工夫することが重要であると述べている。

目標準拠評価を行うに当たっては、具体的な評価規準と基準を設定する必要がある。現在、この規準と基準の設定は、授業担当者の裁量によってなされているが、橋本（1976）は、これが容易ならざる仕事であり、ややとすると主観に陥り、信頼がおけなくなると指摘している。さらに、複雑高度の目標・内容に関して何段階かのスケール化した評価のスタンダードを設定する必要がある場合においてこれをどんな手続きで決めるか、その方法に関してまだ一般に承認されるような理論が見あたらないとも述べている。本研究によって検討された方法は、授業担当者の裁量のみによらず、学習者の知識構造の形成過程をも加味しながら、評価規準と基準の設定が可能となることを示唆している。すなわち、定期テストの項目を、本研究で用いた方法によって検討をした上で配列し、1年間のテストを作成し、実施することによって、より教育的な目標準拠評価を行うことを可能とすると考えられる。

また、個人内評価を行うに当たっても、本研究の方法による習得モデルの構築は有効であると考えられる。つまり、本研究で用いたモデル構築の手法によって、習得順序を具体的に記述することを可能にすると考えられる。これが実現すれば、個々の学習者が、現在どの段階まで習得しているのか、また次に学習すべき内容は何かを理解するための一助となるであろう。そうすることにより、理想的な個人内評価を可能とし、ひいては指導と評価の一体化が図られると考えられる。

さらに、本研究では教授処理後における学習者の知識の様相を表現するにとどまったが、事前と事後とにおけるモデルを検討することにより、具体的に学習者の知識構造がどのように変容したかについての検討をも可能にすると考えられる。また、学習内容を首尾よく習得した群とそうでない群の知識構造を検討し、学習内容の習得という観点から、より適切な教授法を検討することも可能になるであろう。

引用文献

- Doignon, J.-P., & Falmagne, J.-C. (1999) *Knowledge Spaces*. Springer-Verlag.
- 橋本重治 (1976) 新・教育評価法総説 (上巻). 金子書房.
- 林規生 (2001) 英語能力測定における CAT の適応例と効果測定. 計測と制御, 40, 8, 572-575.
- 教育課程審議会 (2000). 児童生徒の学習と教育課程の実施状況の評価の在り方について (答申).
- Little, R. J., & Rubin, D. B. (1987) *Statistical Analysis with Missing Data*. John Wiley & Sons.
- Mislevy, R. J. (1995) Probability-based inference in cognitive diagnosis. In Nichols, P. D., Chipman, S. F., & Brennan, R. L. (Eds.), *Cognitively Diagnostic Assessment*. Lawrence Erlbaum Associates.
- 文部省 (1998) 中学校学習指導要領. 大蔵省印刷局.
- 繁榎算男 (1999) ペイジアン・ネットワーク・モデル. 中島義明・安藤清志・子安増生・坂野雄二・繁榎算男・立花政夫・箱田裕司 (編) 心理学辞典. 有斐閣.
- 椎名乾平 (1997) 児童の問題解決とルール. 長縄久生・椎名乾平・川崎恵里子 (編) 認知心理学の視点—理論と測定法—. ナカニシヤ出版.
- Tatsuoka, K. K. (1995) Architecture of knowledge structures and cognitive diagnosis: A statistical pattern recognition

- and classification approach. In Nichols, P. D., Chipman, S. F., & Brennan, R. L. (Eds.), *Cognitively Diagnostic Assessment*. Lawrence Erlbaum Associates.
- Vermunt, J. K. (1997) *LEM: A General Program for the Analysis of Categorical Data* [Computer software]. Retrieved from <http://www.uvt.nl/faculteiten/fsw/organisatie/departementen/mto/software2.html>
- Vermunt, J. K., & Magidson, J. (2003) *Technical Appendix for Latent Gold 3.0* [Manual]. Retrieved from http://www.statisticalinnovations.com/products/lg_app3.pdf

附記

この論文の一部は、第45回日本教育心理学会総会で発表された。

A study to examine the states of the acquisition of comparison with using Bayesian network model

ABSTRACT

The purpose of this study is to examine the states of the acquisition of comparison with using Bayesian network (BN) model. The model represents causal relationships among variables as a directed acyclic graph. We applied the model to the data on the quiz of comparison conducted for 8th grade Japanese students in the English class.

The two BN models were examined; one is hypothesized that the students acquires both regular and irregular comparison simultaneously, the other is that the acquisition of regular comparison is preceded irregular one. The result indicated the latter model that represented the teacher's treatment fit the data better.

The applicability of the BN model to the data on periodic examinations conducted to evaluate achievement is discussed. Furthermore, we emphasize the result implies that the data of the examinations can be utilized to assess learners' knowledge states diagnostically.