

マトリックスモデルの検証： 長期的な語彙知識変化の安定性に関する一考察

吉 井 誠

本研究の概要

長期的な語彙知識変化を予測する方法としてマトリックスモデルが提唱され、その妥当性について検証されてきた。また、予測に基づきシミュレーションを行い産出した仮想データより、語彙知識変化の詳細を探る試みも行われてきた。この論文では、もう一度原点に戻り、マトリックスモデルの前提である語彙知識変化率（推移確率）の安定性について検証する。実際の被験者22名から2ヶ月をかけてデータを収集し、5回に亘り行った各テスト間の推移確率を調べ安定性について検証した。本稿では、最初にマトリックスモデルの概要を紹介し、実験について説明する。研究目的、リサーチクエスションについて述べた後、研究デザインや手順を説明し実験の結果を報告する。各テスト間の推移確率は安定したものかどうかを検証した後、マトリックスモデルによる予測の妥当性についても検証する。考察を交え今後の研究課題に言及してまとめとする。

研究の背景

ここでは、マトリックスモデルについて簡潔に要点を紹介し、本研究を行うに至った経緯と理由について説明を行う。なお、モデルについてのより詳細な記述は吉井(2009a, 2009b)に掲載されているので、ここでは要点の簡略な説明にとどめる。

語彙知識をどう解釈するかは、それ自体が研究課題であるが、マトリックスモデルを採用する際には、語彙知識を不連続の状態から構成されている知識と定義する(Meara, 1997; Meara & Sanchez, 1993)。例えば、一番基本的な形として、単語を「知っている」(State 1)か「知らない」(State 0)の二つの状態からなるモデルが考えられる。以後、State 0はS0、State 1はS1の略式で記載する。単語の知識は、この二つの状態のいずれかに属し、途中の真

ん中の状態などを想定しない考え方である。そして、いつでも S_0 から S_1 へ、あるいは S_1 から S_0 へ推移が可能であるとする。このモデルにより予測を行う手順は以下の通りである。ある時点である分量、例えば 200 語の単語知識を測定したとする。学習者は、一つ一つの単語について S_0 か S_1 かの判断をしていくが、この場合自己評価方式がよく採用されている。この方式を用いることによって、短時間のうちに相当数(数百語)の単語の評価が可能である。そして、一定時間をおいて、再度同じテストを行なう。そして、最初のテストと 2 番目のテスト間で、 S_0 が S_0 に留まる確率、 S_0 が S_1 に推移する確率、 S_1 が S_1 に留まる確率、 S_1 が S_0 に推移する確率を計算する。そして、この変化の確率、すなわち推移確率(研究者によっては移行行列という呼び方をすることもあるが、本稿ではこの名称を使用)はそれ以後も不変であると仮定して、その確率に基づいて、3 番目、4 番目、そして将来的な n 番目のテストの結果を予測する。この考え方は、そもそもマルコフ連鎖理論(Bradley & Meek, 1986)と呼ばれ、数学、統計学、社会学など様々な分野で幅広く使われてきた。外国語教育の分野でも Meara(1989)、Meara & Rodriguez Sanchez(2001)がこの理論にいち早く着目し、語彙研究に応用した。

これまでの研究において、このモデルの信頼性について検証が行われてきたが(Fitzpatrick, Al-Qarni, & Meara, 2008; Horst & Meara, 1999; Meara & Sanchez, 2001)、その結果、かなり正確に予測できることが判明している。本研究者もこのモデルの検証をこれまでに行い、妥当なモデルであることを確認している。また、モデルに基づいてシミュレーションを行い産出した仮想データから語彙知識変化の詳細な動きの検証も行ってきた(吉井, 2009a, 2009b, 2010)。これらの研究の前提となっているのが推移確率の安定性である(Meara, 1989; Meara & Rodriguez Sanchez, 2001)。

本研究ではマトリックスモデルの前提である推移確率に着目するが、推移確率が不変であるとは現実的には考えられない。しかし、各テスト間でかなり近似しており、安定しているために、かなり正確な推測が可能になると考えられる。モデルの検証を継続し、理論を発展させていくためには、前提となっている推移確率の安定性について日本人学習者を対象として実験を行い検証することが必要であると考えた。

本研究の目的

本研究では、日本人学習者を対象として推移確率の安定性と予測の妥当性について検証を行う。安定性を確認した後、マトリックスモデルによりどの程度正確に被験者のデータを予測できるのか調べていく。具体的には以下

の二つのリサーチクエスチョンを設定した。

1. 日本人学習者の各テスト間における推移確率は安定しているのか？
2. 日本人学習者を対象として、モデルによる予測は被験者の実際の語彙チェックテストの結果をどの程度正確に反映させているのか？

研究デザイン

被験者は日本人大学生 22 名であり、英文専攻の 1 年生で男性 1 名女性 21 名であった。運用基礎クラスを受講している二クラスを対象とした。実験は 40 名を越える被験者で始まったが、これから説明する手順に従い、すべての授業に出席し、すべての課題に取り組み、すべてのテストを受験した学習者に絞ったところ、最終的には 22 名となった。

実験の手順は以下の通りである。被験者は授業の一環として、1 ヶ月間に 200 語を集中的に学習した。目標単語は、GSL (General Service List) と呼ばれる語彙表 (West, 1953) から抽出した。これは英語の書き言葉に最も頻繁に出現する語彙のリストであり、かなり古いリストではあるがいまだに頻繁に使用されている。現在でも重要な語彙リストの 1 つであり、このリストの語彙によって既存のノンフィクションの書き言葉なら 75%、フィクションなら 90% をカバーしていると言われている (Nation & Hwang, 1995)。被験者には学期始めに、GSL リストから研究者によって抽出された 1,000 語の語彙表が渡された。これは主に GSL の 2000 語レベルと 1000 語レベルの一部から選ばれたもので、研究者のこれまでの経験より、この被験者群にとって未習のもの、あるいは既習でも知識がまだ定着していないと思われるものを抽出してリストにした。被験者に知っている単語にはチェックを入れるように指示し、その回答を集計した。その結果より、このグループの被験者が学習すべき単語、トップ 200 語のリストを作成し、これを今回の実験の目標単語とした。学習者には授業外の課題として、毎週 40 単語、5 週間かけて 200 語を覚えることが課せられた。学生には英語とその意味に相当する日本語が書かれてある Study Sheet が渡され、1 ヶ月の間で集中的に学習することが義務付けられた。その週に覚えて来た単語については、次の授業で多肢選択法の単語テストが課せられた。200 語の集中学習が終了後、約 2 ヶ月に亘って定期的に単語テストを行ない、学習した単語の知識の変化 (定着と忘却) について調査した。テストは、研究者が作成した Yes-No 形式のコンピュータプログラムで実施した。このプログラムでは、単語が 1 つずつスクリーンに登場し、知っている場合、Yes のボタンをクリックし、次の単語に進んだ。知らない場合、あるいは忘れてしまった場合に No のボタンを押すと、スク

リーンに単語の意味が日本語で提示された。これは、ただ単に単語の知識を測定するためのテストに終わらせず、忘却を防ぎ定着への効果を狙った教育的配慮から実施された処置であった。このプログラムで200語のテストを終了し、終了ボタンをクリックすると、YesがいくつでNoがいくつであるという今回のテストの結果が、前回のテストの結果と並んで表示され、学習者は前回からの語彙知識変化の概観が把握できるようになっていた。

実験の大まかな計画は、図1のとおりである。被験者は、200語の集中学習を終了1週間後にTest 1を受け、その1週間後にTest 2を受けた。Test 2を受けた直後に冬休みに入ったため、Test 3が実施されたのは休み明けの最初の授業であった。それ以後はTest 4、Test 5とそれぞれ1週間の時間を置いてテストが行なわれた。その結果、Test 2とTest 3の間のみが、他のテスト間の期間と比べ長くなっている。被験者が受講していた二つのクラスのスケジュールの違いから、Test 2とTest 3の間は1つのクラスは4週間、もう1つのクラスは5週間となった。推移確率の安定性を調べるために、一人ひとりの被験者の各テスト間の推移確率を算出し、これらの確率の相関関係を分析した。マトリックスモデルの予測に関しては、Test 1とTest 2の間の推移確率からTest 5時点での語彙テストの結果を推測した。そして、Test 5の学習者の語彙テストの結果（実測値）とTest 5時点でのマトリックスモデルによる予測（予測値）との比較も、両データの相関関係を用いて分析した。

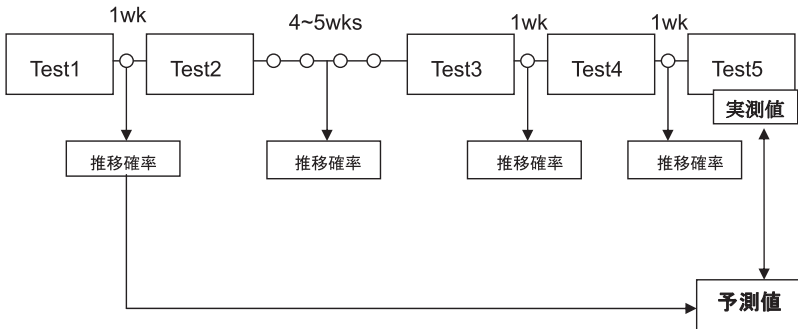


図1 本実験の計画手順

結果

最初に各テスト間の推移確率の安定性についての検証結果を報告し、次に、実測値と予測値の関係を調べた結果について報告する。推移確率の安定性について調べるために、一人一人の被験者の各テスト間の推移確率を算出した。表1では、例として3人分の推移確率が記載されている。表の左端に推移パターンが記載されているが、ここでは、Yes-No形式の語彙テストにおいて、4つの推移パターンが存在することを示している。すなわち、最初のテストでS0という状態がS0に留まるパターン、またS0がS1状態へと変化した場合、逆にS1であったものが、S0に逆戻りする場合、そして最後に、S1が次のテストでもS1の状態である場合の4つのパターンである。そして、これをTest 1とTest 2の間(T1-T2)、Test 2とTest 3の間(T2-T3)、Test 3とTest 4の間(T3-T4)、そしてTest 4とTest 5の間(T4-T5)とそれぞれのテスト間で算出した。図1でも説明したように、T1-T2、T3-T4、T4-T5はそれぞれ1週間の時間を置いて実施し、T2-T3のみ4週間から5週間と長い間隔を置いた。しかし、参考までにすべてのテスト間における推移確率を分析することにした。

表1 3人の被験者の各テスト間の推移確率

推移パターン	T1-T2	T2-T3	T3-T4	T4-T5
s0s0	.73	.83	.81	.69
s0s1	.27	.17	.19	.31
s1s0	.14	.13	.10	.06
s1s1	.86	.87	.90	.94
s0s0	.62	.76	.37	.41
s0s1	.38	.24	.63	.59
s1s0	.03	.12	.04	.11
s1s1	.97	.88	.96	.89
s0s0	.56	.57	.59	.58
s0s1	.44	.43	.41	.42
s1s0	.10	.19	.15	.12
s1s1	.90	.81	.85	.88

表1のような形で、被験者22名全員の各テスト間の推移確率を算出し、それぞれの推移パターンが安定したものであるかどうかを、各テスト間の相関関係を用いて分析した結果を表2に示す。

表2 各テスト間の推移確率の相関関係

	T1-T2 (1wk)	T2-T3 (4~5wks)	T3-T4 (1wk)	T4-T5 (1wk)
T1-T2(1wk)	1.00			
T2-3(4~5wks)	.84	1.00		
T3-T4(1wk)	.93	.88	1.00	
T4-T5(1wk)	.86	.80	.94	1.00

表2が示しているように、各テスト間の推移確率は強い相関関係を示しており、推移確率が安定していたことを示している。テスト間の間隔が他のテスト間と比べかなり長かったT2-T3においても、他のテスト間の推移確率と高い相関関係を維持しており、テスト間の間隔の長さに関わらず安定していたことを裏付ける形となった。これにより、マトリックスモデルの前提である推移確率は不変ではないにしても、かなり安定したものであることが分かった。

次に、実測値と予測値の関係についての結果を図2に示す。これは、Test 5における22名の学習者の語彙テストの結果（ここではYesと認識された単語の数）が線で示され、マトリックスモデルによるそれぞれの予測値が棒グラフで表されている。図2では、結果が分かりやすいように、被験者の語彙テストの得点が高いものから低いものへと順番に置き換えている。図2が示すように、実測値よりも高い予測値（棒が線を越えている場合）もあれば、逆に実測値よりも低い予測値（棒が線よりも下に位置するもの）も存在するが、全体的に見ると二つの値は近似しており、かなり正確に予測ができていることを表している。予測値と実測値の相関関係を調べたところ、ピアソンの係数で $r = .85$ と強い相関関係を表しており、統計的にも予測がかなり正確であったことが確認できた。

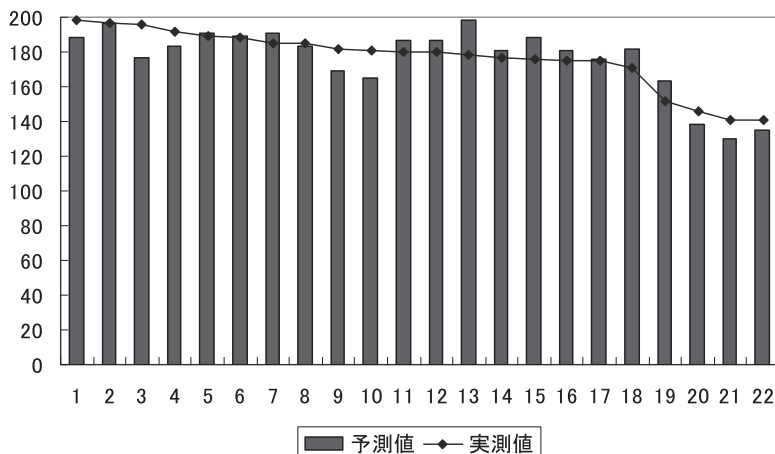


図2 Test 5における実測値と予測値の比較

結論とまとめ

本実験の結果より、マトリクスモデルの前提となっている推移確率はかなり安定したものであることが分かった。また、モデルによって実測値をかなり正確に予測することが可能であることが示された。これにより、マトリクスモデルによる長期の語彙知識変化の予測が可能であることを再確認することが出来た。

しかし、本研究はまだ始まったばかりであり、課題が残されている。今回は Yes か No の二つの知識の状態を想定した一番基本的な 2-State モデルを採用したが、これは、研究を始めるにあたりできるだけシンプルなものから始め、徐々に複雑なものへと発展させることが大切だと考えたからである。今後は、「どちらともいえない」「はっきりしない」という Yes と No の中間に位置する状態を含んだ 3-State モデルや、さらに、「知らない」「知っているがはっきりしない」「知っていると思う」「はっきりと知っている」の 4 つの状態を想定した 4-State モデルなど、より複雑なモデルの検証も望まれる。その際には、細分化された一つ一つの語彙知識の状態を適切に定義することが重要である。また、3-State、4-State などのモデルにおいても、推移確率は安定しているのかどうかを検討していく必要がある。モデルに基づく予測を、実際の被験者から収集したデータと照らし合わせ、モデルの信頼性、妥当性について引き続き検証していかなければいけない。

今後は、これらの研究結果をどのように応用していけばいいのか検討することも課題である。応用としては、予測に基づく早期診断とそれに沿った適切な指導が可能性として挙げられる。これまでの語彙テストでは、テスト受験時点での語彙力の評価が基本であったが、このモデルを使用することによって、未来へ向けての評価、予測が可能となり、テスト時点で学習方法を修正する必要がある場合に早期に指導を行うことが可能となる。今回採用した2-Stateモデルには、S0がS0に留まる（学習がなされていない）、S0がS1に推移する（学習が行われている）、S1がS0に戻る（忘却がおきている）、S1がS1に留まる（定着している）の4つのパターンが存在した。各学習者の推移確率の特徴を分析し、それに基づいた指導助言が可能となる。たとえば、S0からS0の確率が高い学習者の場合、学習を怠っている、もしくは学習はしても新出語がなかなか記憶に定着しない、という二つの可能性が考えられる。指導の際は、どちらが原因かを学習者との面談の中でつきとめ、前者の場合、未知語を積極的に学ぶことをすすめたり、後者の場合、学習困難な単語について単語個別の学習方法を提示したりすることができる。また、S1からS0への確率が高い学習者の場合、学習したものを忘れやすい傾向が強く、最初に覚える段階の問題なのか、それとも学習後の復習の問題なのか、原因を究明する必要がある。いずれにしても、推移確率による学習状況の説明やモデルによる予測値の紹介など、具体的な数値を提示することで、より具体的で、より現実味のある助言が学習者に提供できる。

語彙知識の変化については、被験者からのデータ分析のみに留まらず、シミュレーションを用いた検証（吉井、2009a、2009b、2010）を行うことも必要である。実際の学習者から収集できるデータの量にはおのずと制限が生じる。数百語の単語を学習させ、その知識を長期間に亘って定期的に何回もテストを実施することは容易ではない。シミュレーションを行うことによって、被験者を用いた実験を想定した仮想データを産出し、それを用いて理論の検証や、予備実験を行うことができる。シミュレーションの役割が今後益々重要となってくることが予想される。

謝辞 本研究は、科学研究費（課題番号 20520510）の助成を受けたものである。

参考文献

- Bradley, I., & Meek, R. L. (1986). *Matrices and society*. Harmondsworth: Penguin.
- Fitzpatrick, T., Al-Qarin, I., & Meara, P. (2008). Intensive vocabulary learning: a case study. *Language Learning Journal*, 36(2), 239-248.
- Horst, M., & Meara, P. (1999). Test of a model for predicting second language lexical growth through reading. *Canadian Modern Language Journal*, 56(2), 308-328.
- Meara, P. (1989). Matrix models of vocabulary acquisition. *AILA Review* 6, 66-74.
- Meara, P. (1997). Towards a new approach to modeling vocabulary acquisition. Models of vocabulary acquisition. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 109-121). Cambridge: Cambridge University Press.
- Meara, P., & Sanchez, I. R. (1993). Matrix models of vocabulary acquisition: An empirical assessment. *CREAL Occasional Paper No. 1*. Ottawa: University of Ottawa.
- Meara, P., & I. Rodriguez Sanchez (2001). A methodology for evaluating the effectiveness of vocabulary treatments. In M. Bax & J-W Zmert (Eds.), *Reflection on language and language learning*. (pp.267-278). Amsterdam: John Benjamins Publishing Co.
- Nation, I.S.P., & Hwang, K. (1995). Where would general service vocabulary stop and special purposes vocabulary begin? *System*, 23(1), 35-41.
- West, M. (1953). *A general service list of English words*. London: Longman.
- 吉井 誠 (2009a). 「語彙知識変化の検証～シミュレーションと実際のデータとの比較」『学術情報センター紀要』第13, 14, & 15巻第1号, pp. 1-22.
- 吉井 誠 (2009b). 「マトリックスモデルを用いたシミュレーションスタディー：長期的な語彙知識変化の検証」『熊本県立大学文学部紀要』第15巻第68号、pp. 65-79.
- 吉井 誠 (2010). 「長期的語彙知識変化パターンの検証：被験者データとシミュレーションデータの比較」『熊本県立大学文学部紀要』第16巻第69号、pp. 33-46.