

クラスター分析による読書過程の研究

高山 草 二

Soji Takayama

A Study of Reading Processes by Cluster Analysis

読書過程の研究は、日常の読書、テキストからの学習といった実際的なテーマに深くかかわることからも深い関心を集めてきた。なかでも読書過程における個人差の研究は、テキストからの学習における差違というような実際的な問題意識と密接に関連している (Voss, 1984)。読書過程の個人差の研究は、単に個人差の研究に留まらず、個人差という現象を通して、背後に隠された読書過程自体を明らかにするという点においても重要である。読書過程は、パターン認識、意味記憶、推論、言語などの諸過程を含むきわめて複雑なものである。この複雑さゆえに、個人差の研究において、読書過程を反映する様々な変数の相関関係の分析が主要課題となる。

個人差を研究する場合、個人を読書過程にかかわる様々な側面から測定し、それらの側面と読書速度などのある一つの基準との相関関係を調べることが多い。基準としての読書速度の違いを、測定された側面のどれがうまく説明できるかを検討することによって、個人差を作り出している読書過程を明らかにしていこうとするのである。例えば、Baddelay, Logie & Nimmo-Smith (1985) は、読書テストの成績を基準とし、記憶スパン、語彙量、語彙判断、文字照合などの測定を行なっている。これらの測定結果と基準との相関パターンの分析から、語彙量、記憶スパン、語彙的接近の過程が読書テストの違いをよく説明できることを見出した。この研究では個人の特性を個別の課題を課して測定しているが、より自然な方法で測定するアプローチもある。

Graesser, Hoffman & Clark (1980) の研究では、被験者は一度に一文ずつ呈示される文を読み、理解できれば反応キーを押すという方法により、一文の理解に要する時間を測定した。この反応時間に文章の様々な性質がどのように影響するかを調べたのである。彼らは文章の

諸性質をマクロ構造とマイクロ構造とに分けて分析した。マクロ構造は文のレベルよりも高次の性質であり、文章のタイプ（物語文・説明文）、主題の熟知度、名詞（命題の変項）が文章内で初めて出現したかどうかの3変数であった。マイクロ構造は文のレベル以下の性質であり、各文に含まれている単語数、命題数、各文の統語的予測可能性の3変数が検討された。これら6変数を予測変量とし、一文の理解に要する時間を目的変量として重回帰分析を行ない、被験者別に各変数の偏回帰係数を求めた。そして、文章全体の理解に要する時間に関して速い被験者と遅い被験者とに2分して、偏回帰係数の比較を行なった。その結果、マクロ構造に関する変数には差が無く、マイクロ構造に関する変数において差がみられた。これから、読書過程における個人差はマイクロ構造の処理過程によって生ずると考えた。

彼らのアプローチにおいては、重回帰分析の結果得られた回帰係数が個人の特性を表わすことになる。Baddeley らが個別の課題を課して測定した特性を、間接的ではあるが回帰係数という形で求めるのである。このいくつかの回帰係数のうちどれが基準としての読書速度の違いに深く関与するかを調べることにより、個人差を明らかにする点で、Baddeley らと同一の論理といえよう。

さて、様々な課題の測定結果や回帰係数によって表わされる個人の諸特性をある基準を予測するために利用する方法を検討してきたが、これらの諸特性が個人の読書過程の特性を反映しているという点に注目するならば、例えば、回帰係数に基づいて個人を分類することも可能であろう。読書速度を予測できる特性を調べるという形ではなく、個人の特性のパターンの中に隠された読書過程とその個人差を探るのである。そのために、いくつかの特性に基づいたクラスター分析を行ない個人をいくつかの群に分類する。構成されたこれらの群がどのような

次元で異なっているかを検討することによって、個人差を作り出している読書過程を明らかにできるかもしれない。社会的判断理論の研究領域では判断の方略の違いを明らかにするためこのようなアプローチが応用されている (Adelman, Stewart & Hammond, 1975; Knight, 1984)。読書速度などへの寄与という形で読書過程の特性を表わす回帰係数を調べる方法が、外的基準がある場合の多変量解析に対応するのに対し、回帰係数そのものを分析の対象とするのは外的基準がない場合の多変量解析を行なうことともいえよう。

本研究では、文章内の各文節を理解する時間と文節のもつ性質とに関して重回帰分析を行なう (高山, 1984)。その結果得られた回帰係数に基づいてクラスター分析を行ない、被験者を分類する。次に、正準相関分析を行ない、この分類の背後にある次元を明らかにする。この次元の分析から個人差を作っている読書過程について検討してみよう。この新しいアプローチによって、読書速度といった単一の基準の予測には必ずしも寄与しない特性が、個人差を作る要因として明らかになることも期待できる。

方 法⁽¹⁾

被験者 島根大学の学生30人。

装置 マイクロコンピュータ (富士通 FM-16 β) により刺激呈示とその制御を行なった。呈示した文字の字体は24×24ドットの精度であり、漢字 ROM (沖電気 MSM 38128) を用いた。CRT 上に呈示した各文字の大きさは縦 6.4mm×横 6.3mm、文字間隔は 0.5mm であり、被験者は 80cm の距離から見た。輝度は文字の部分が 130Nit、背景が 4.8Nit であった。マイクロスイッチを用いた反応キーから FM-16 β に入力することにより、反応時間を 0.54msec 単位で測定した。

刺激 菌類に関する説明文を科学事典 MEGA から抜粋して用いた。内容は菌類の種類、生態系における役割、繁殖力などについて解説したものであり、30文、333文節から成る。この文章に含まれる各文節について次の変数を測定した。(1)文章内における各文節の系列位置。(2)各文の中における各文節の系列位置。(3)文頭であるか否か。(4)文末であるか否か。(5)各文節の自立語の出現頻度。複合語の場合は出現頻度表 (国立国語研究所, 1962) の見出しにある語を単位とし、複数個の単位が文節に含まれているときは、そのうち最小の頻度をその文

表 1 偏回帰係数の平均と t 検定の結果

	平均	t	確率
文章内系列位置	-0.55	-3.32	P<.005
文内系列位置	-2.88	-2.12	P<.05
文末	-12.38	-0.53	ns
文頭	12.64	1.05	ns
文字数	52.16	7.37	P<.001
新出変項	64.26	5.14	P<.001
出現頻度	-9.68	-5.37	P<.001
切片	736.04	9.89	P<.001

節の出現頻度とした。この度数を自然対数変換した。頻度表にのっていない場合は度数を10として変換した。(6)新出変項。命題の変項となっている名詞が、初めて文章内で出て来たか否かを示す。(7)文節を構成する文字の個数。以上のうち、(3), (4), (6)は0と1のコーディングを行なった。実験用の文章の他に練習用に短い文章を用意した。

手続き 実験は個別に行なわれた。教示をした後、練習用文章を読んで手続きを理解させ、実験用の文章を始める。文章の始まる前に CRT の中央部に題目 (「菌類」) が呈示され、これを理解した被験者が反応キーを押すと、文章を構成する各文節が画面中央に呈示される。被験者はその文節を理解できたら反応キーを押す、それによって次の文節が呈示される。文節が呈示されてから反応キーを押すまでの時間を反応測定として記録した。反応キーを押してから次の文節が呈示されるまでの時間間隔は 1/60sec 以内であった。

被験者には文章の内容を理解するよう教示し、文章を読み終わった後、理解度についてテストすることをあらかじめ知らせておいた。実験用文章を読み終えた後、文章の内容について簡単なテストを行なった。このテストを終了した後、続けて短期記憶の課題を行なったが、本研究の目的とは関連しないものである。

結果と考察

理解度のテストは、多義的な質問が入っていたこと、理解しながら読むことを保証するためのものであったことから分析の対象にはしなかった。文節に対する反応時間は、100ms 以下と 5sec 以上のものを分析から除外した。その結果、全データの 0.69% が除外された。各文節に対する反応時間と文節の特性より 7 変数について、各被験者ごとに重回帰分析を行なった。表 1 は各変数と切

(1) 本実験は昭和60年度教育心理実験 No. 2 の一環として行なわれた。

表2 クラスタ別各変数に対する偏回帰係数の平均と標準偏差 (S D)

クラスター		文章内 系列位置	文内 系列位置	文 末	文 頭	文 字 数	新出変項	出現頻度	切 片
1	平均	-1.43	-2.04	-124.64	28.04	86.26	128.86	-16.99	1105.74
	S D	0.94	5.08	101.89	74.44	24.08	88.46	6.40	514.41
2	平均	-0.25	-16.01	190.63	-95.58	89.20	86.38	-16.53	541.24
	S D	0.76	8.42	166.26	18.55	20.09	18.82	12.18	199.68
3	平均	-0.23	-0.34	-7.61	29.85	28.77	30.63	-4.91	615.01
	S D	0.67	4.82	51.94	43.03	28.34	38.10	7.92	280.42

片に対する偏回帰係数の平均と t 検定の結果である。文末・文頭に関する変数をのぞき、全て有意となった。

求めた偏回帰係数を変数ごとに標準化してから、クラスター分析にかけた。用いたクラスター分析は非階層的方法に属する⁽²⁾ (SAS USER'S Guide, 1982)。2, 3, 4の3種類のクラスター数について分析した結果、クラスター数2の場合は5, 25人, 3の場合は4, 8, 18人, 4の場合は2, 4, 9, 15人サイズに分けられた。クラスター数4の場合に小さなサイズのクラスターが生じたこと、クラスターから変数を予測する R^2 をプールした値がクラスター数を3から4に増してもあまり変わらないこと(各々, 439, 481)から、クラスター数3の場合について更に分析を行なった。各クラスター別の各変数の平均を表2に示した。

次に3つのクラスターの各々の性質を明らかにするために、3群をダミー変数としてコーディングした場合の正準相関分析を行なった。これは記述的目的のために用いているので、検定についてはふれない。その結果、負荷行列は表3のようになった。2つの正準変量を座標として正準スコアをクラスター別にプロットしたのが図1である。この図から明らかなように、第3クラスター(18人)と第1, 第2クラスター(合計12人)は、第1正準変数によって区別される。これに対し、第2正準変数では、第1クラスターが最も高いスコアを持ち、そのすぐ下に第1クラスターと重なる形で第3クラスターが位置し、この2つのクラスターから少し離れて第2クラスターが位置している。つまり、第2正準変数は3個のクラスターを区別しうることになる。

2つの正準変数の意味を考えてみよう。第1正準変数に対する各変数の負荷を表3で調べると、大きいものは文字数、新出変項、出現頻度の3変数にみられる。文字数と出現頻度は共に語認知過程を規定する変数である。新出変項は、文章内の諸命題のある変項を記憶システム

表3 負荷行列

変 数	正 準 変 量	
	1	2
文章内系列位置	0.267	-0.245
文内系列位置	0.276	0.429
文 末	0.049	-0.593
文 頭	0.194	0.398
文 字 数	-0.537	-0.034
新 出 変 項	-0.378	0.115
出 現 頻 度	0.356	0.002
切 片	-0.217	0.256

内にノードとして形成する必要があるかどうかにかかわるものであり、その変項が既出のものでなければ、ノード作成にある程度の時間を要することが考えられる。すなわち、新出変項は文章内の諸命題の統合過程に関与する変数として想定されている。しかし、こういった役割の他に、新出の変項は既出のものと同語認知過程においても違っているであろう。既出の語は繰返し表示されることで、新出のものより語認知が促進されていると考えられる。語認知における反復効果(Mitchell, 1982)の存在はこれを支持するものであろう。したがって、文字数、出現頻度と共に、新出変項も語認知過程に関与するもので、第1正準変数は、語認知過程のレベルの次元を表わすものと解釈できよう。この次元によって被験者は大きく2分されており、読書過程の個人差の要因として語認知レベルがかかわっているといえよう。ここには、文字数に関与すると思われる視覚的弁別過程や、出現頻度が規定するとされる語彙的接近過程が含まれると考えられる。負荷の正負を考慮すれば、第3クラスターは文字数、新出変項の係数が小さく、第1・第2クラスターはその逆である。第3クラスターの方が、この次元に関して効率的なグループを表わしている。上記の関係は、各クラスター別の回帰係数の平均(表2)を見ても確認

(2) 分析は全て京都大学大型計算機センターのSASを利用。

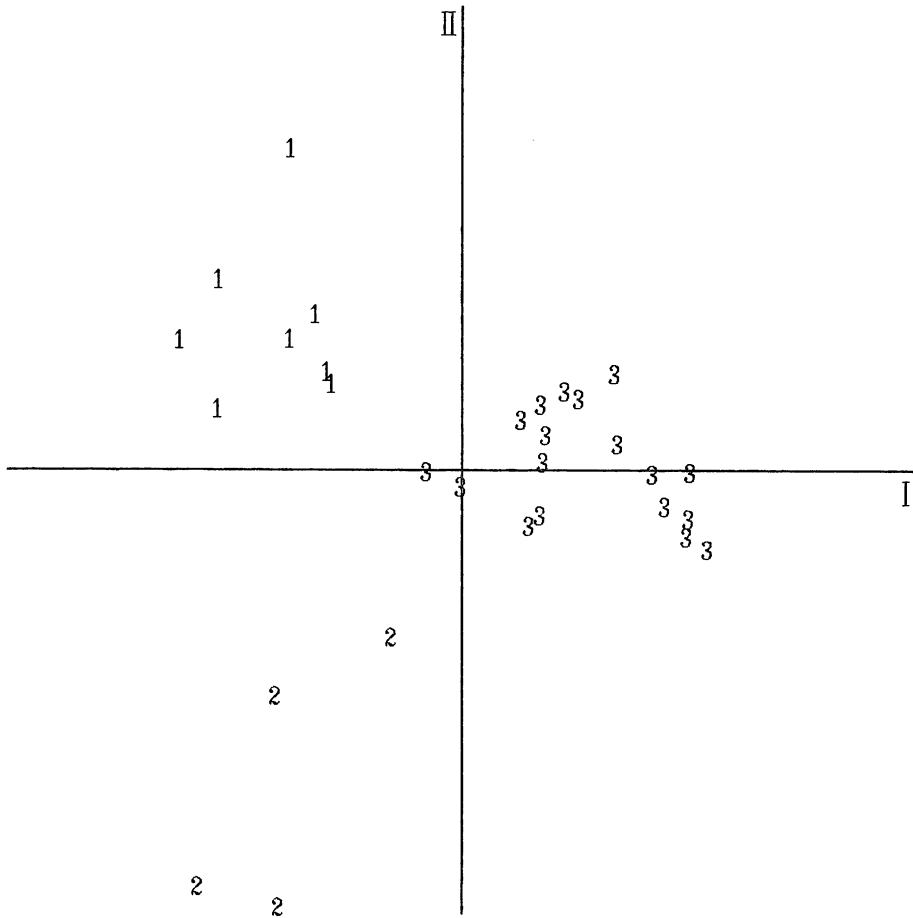


図1 3クラスターの正準スコアによる表示。図中の数字は各クラスターを示し、I、IIの軸はそれぞれ第1、第2正準変量を表わす。

できる。

第2正準変量の負荷を調べてみると、文内の系列位置、文末、文頭の各変数で大きい。これら3変数は全て文内における各文節の位置を表わしている。その位置が文頭から文末に移動するにつれて、文法的、意味的制約が増し、文脈効果が生ずる。文末では、一つの文の意味的統合を形成するため、遅れが生ずる場合がある(Just & Carpenter, 1980; Muncer & Bever, 1984)。また、文頭では、テーマの切り換え、新たな意味領域の設定などの余分な過程が含まれることもある。このように、これら系列位置の問題は複雑であるが、いずれにしても、単語レベルを越えた、文処理のレベルを基礎としている。それゆえ、この第2正準変量は文の処理過程が関与する次元として解釈できよう。そして、この次元は

第1の語認知のレベルとは独立した形で、被験者を区別することができる。

3変数を個別に調べると、反応の文末での遅れが大きく、文頭では逆に早く、系列位置が進むに従って早くなる場合が、第2正準変量の小さい方に対応する。これらの結果のパターンは複雑であり、文処理過程と系列位置との関係が単純ではないことを示している。この点は、今後の研究が必要であろう。

回帰係数に基づいてクラスター分析を行ない、3群を作った。これらを分類する次元として、語認知過程を表わす次元と、より高次の文処理過程にかかわる次元が認められた。読書過程にはこれら2つの次元が独立に関与しており、またこの2つの次元が個人差をも作り出している。Graesser et al (1980) は、読書の速さに関する

個人差は文章のマクロ構造ではなくて、ミクロ構造によって作られるとしている。本研究の結果得られたクラスターを弁別する2つの次元はいずれも文レベル以下と考えられるので、Graesserらの結果との矛盾はない。さらに本結果でマクロのレベルにかかわる文章内の系列位置が2つの次元に大きくは関与していないことは、マクロ構造が個人差には無関係であることを支持するものである。

本結果から得られた2つの次元は、相互に独立に個人差を説明するという点が重要であろう。ミクロ構造が個人差を作る場合にも、その個人差は、語認知レベルに基づくこともあるが、文処理のレベルに基づくこともあり、または両方がかかわることもある。高山(1984)において、反応時間で分類した場合、本研究での第1正準変量に関係する、出現頻度と新出変項に有意な個人差はみられなかった。これは、反応時間による分類と、本研究で得られた2次元とが対応していなかったことによるとも考えられる。反応の遅速といった一つの分類基準で分割した場合、読書過程とその個人差を面的にしか検討できないこともありうるであろう。Kintsch & Mross (1985)は、語認知を意味の賦活と文脈に適合した意味の選択という2つの過程に分け、前者には文脈などは影響しないと考えている。Forster (1981)も言語システムの語彙、統語処理、解釈処理のレベルは、相互作用をもたない自律的なものとしている。本研究で得られた語認知・文処理レベルの独立的な2つの次元は、言語システムの自律性、モジュール性にかかわるものかもしれない。

本研究では30人の被験者についてクラスター分析を試みたが、十分なサンプル数とはいえないかもしれない。クラスター分析の結果、そのサイズは18人、8人、4人となっており、小さなサイズのクラスターに一般性からの偏りの可能性が考えられる。十分なサンプルによる今後の検討を要する。検討した諸変数は文章の諸特性を十分代表するものではない。その意味で、本結果は取り上げた変数において成立するものに過ぎないかもしれない。検討した変数には、文章レベルの特性を表わすものがほとんど含まれていなかった。唯一検討した文章内の系列位置は常に有意な効果を持つが、もっと文章の意味内容にかかわる変数の検討が必要である。さらに、文のレベルについても、系列位置的なものばかりでなく、文法的、意味的特性を表わす変数の工夫が望まれる。

また、本研究ではGraesserらの方法に従って、理解しながら読むという教示が守られたと仮定して反応時間のみの分析をした。しかし、今後、理解ということの意

味も含めて、この仮定の十分な検討が必要であろう。Graesserら(1984)は、詳細な理解度テストの結果と一文の読書時間との間に負の相関関係を得た。つまり、読書の速い方が理解度が高いということになり、理解を無視して速度を上げるということはみられていない。今後は、むしろ時間的側面と理解の側面とを1変量として個別に研究するのではなく、多変量として同時に検討できるアプローチも展開すべきであろう。

本研究は、個人の特性に基づく分類から、その背後に隠された個人差の次元を探るものであった。いくつかの問題点があり、予備的分析の域を出ないともいえるが、このようなアプローチも読書過程とその個人差の研究にとって、有望なものと思われる。

参 考 文 献

- Adelman, L., Stewart, T.R. and Hammond, K. R. 1975 A case history of the application of social judgment theory to policy formation. *Policy Science*, 6, 137-159.
- 渥美和彦他(編)1985 MEGA. 講談社.
- Baddeley, A., Logie, R., and Nimmo-Smith, I. 1985 Components of fluent reading. *Journal of Memory and Language*, 24, 119-131.
- Forster, K.I. 1981 Priming and the effects of Sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. *Quarterly Journal of Experimental Psychology*, 33A, 465-495.
- Graesser, A. C., Hoffman, N. L., and Clark, L. F. 1980 Structural components of reading time. *Journal of Verbal Learning and Verbal Behavior*, 19, 131-151.
- Graesser, A. C. and Riha, J.R. 1984 An application of multiple regression techniques to sentence reading times. In D.E. Kieras and M. A. Just (Eds.) *New methods in reading comprehension research*. Hillsdale, N. J.: Lawrence Erlbaum Associates.
- Just, M. A. and Carpenter, P.A. 1980 A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87, 329-354.
- Kintsch, W and Mross, E.F. 1985 Context effects in word identification. *Journal of Memory and Language*, 24, 336-349.
- Knight, G.P. 1984 A survey of some important techniques and issues in multiple regression. In D. E. Kieras and M. A. Just (Eds.) *New methods in reading comprehension research*. Hillsdale, N. J.: Lawrence Erlbaum Associates.
- Mitchell, D. C. 1982 The process of reading: A cognitive analysis of fluent reading and learning to

- read. Chichester: John Wiley & Sons.
- Muncer, S. J. and Bever, T. G. 1984 Sensitivity to propositional units in good reading. *Journal of Psycholinguistic Research*. 13, 275-279.
- 高山草二 1984 読書時間の諸成分と文脈効果—個人差を通しての予備的分析. 島根大学教育学部紀要(人文・社会科学編), 18, 1-8.
- Voss, J. F. 1984 On learning and learning from text. In H. Mandl, N. L. Stein, and T. Trabasso (Eds.) *Learning and comprehension of text*. Hillsdale, N. J.: Lawrence Erlbaum Associates.