

文字と単語の内部表現

牧岡 省吾

要旨

本研究では、学習による内部表現の獲得という観点から文字と単語の認知過程を包括的に理解することを目指し、先行研究のデータの再分析とシミュレーション、これまでに提案されたモデルに関する考察、そして心理実験を行った。

第1章では、文字とランダムドット・パターンの内部表現について検討した。まず、文字同士の混同確率やランダムドット・パターン同士の類似度を測定した先行研究のデータを、多次元尺度法(MDS)を用いて系統的に再分析した。MDSは、刺激同士の心理的類似度に関するデータから、多次元空間における刺激の布置を導出する。アルファベット文字間の混同確率を測定した Gilmore et al. (1979) 実験のデータを MDS によって分析した結果、幅や丸みなどの大局的な特徴にもとづく布置が得られた。また、ランダムドット・パターン間の類似度を測定した実験 (Blough, 1985) のデータから得られた布置も、大きさや規則性といった大局的な特徴に対応していた。これらの分析の結果から、文字やランダムドット・パターンのような2次元的な形は、脳内において大局的な特徴の組み合わせによって表現されていると考えられる。このような見方は、神経生理学的な研究から得られた証拠とも一致している。

次に、上記の心理実験で使用された文字やランダムドット・パターンをニューラルネットワークに学習させ、獲得された内部表現を分析した。シミュレーション1ではアルファベット文字の分類を学習させた。MDSによる分析の結果得られた布置は、心理実験のデータの分析から得られたものとは異なっていた。そこで、シミュレーション2では入力パターンの恒等変換をネットワークに学習させた。このようなネットワークが行う変換は主成分分析に近いことが知られている。アルファベット文字の恒等変換を学習したネットワークの内部表現は、心理実験のデータの分析によって導出された布置と類似した構造をもっていた。これは、ネットワークの獲得した内部表現が人間の脳内表現と類似していることを示唆している。ところが、ランダムドット・パターンを入力パターンとして用いた場合には、恒等変換を学習したネットワークの内部表現から得られた布置は、心理実験から得られたものと似ていなかった。この矛盾は入力表現を改良することで解消することができた。以上のようなシミュレーションの結果、入力パターンの恒等変換を学習する単純なニューラルネットワークによって、複数の心理実験の結果を説明可能であることが分かった。2次元的な形の脳内表現は、主成分分析による情報圧縮という枠組みで統一的に理解することができると考えられる。

第2章では、単語の認知過程に関する生理学・心理学的な研究について概説し、これまでに提案されたニューラルネットワーク・モデルについて考察を加えた。

単語認知過程と関係の深い脳損傷として、純粹失読がある。純粹失読の患者は、文章を読む能力が顕著に損なわれているが、その他の対象を認知する能力や、言語の理解に関する能力は比較的正常に保たれている。このような障害は、複数の対象を同時に認知するためのシステムが損傷を受けることによって生じるのではないかと考えられている (Farah, 1990)。

一方、認知心理学においては、様々な手法を用いて単語認知過程に関する実験が行われてきた。本論では、単語優位効果と文字の混合エラーという2つの現象に注目して議論を行った。単語や文字を瞬間的に呈示したとき、単語の中に含まれる文字は、単独で呈示された文字よりも認知が促進されることが知られている。これが単語優位効果である (Reicher, 1969)。この現象は、単語の表現と文字の表現との間に相互作用が生じていることを示唆している。一方、複数の単語を同時に瞬間呈示した場合には、文字が単語間で混合して見えることがある (Allport, 1977)。これは文字の混合エラーと呼ばれる。どのような条件において文字の混合エラーが生じやすくなるのかを探ることによって、単語の内部表現の性質を探ることができる。

以上のような生理学的・心理学的な知見にもとづいて、これまでに提案された単語認知過程に関する3つのニューラルネットワーク・モデルについて解説し、検討を加えた。McClelland & Rumelhart (1981) の Interactive Activation Model は、1個の単語の認知過程に関する定量的な説明を行うことに成功した。さらに、McClelland (1986) の Programmable Blackboard Model は、これを複数の単語に拡張した。しかし、神経生理学的な妥当性が低い、出力が不明確である、といった問題が生じることになった。これらの問題の原因は、入力表現において位置に関する不変性が実現されていないことにあると考えられる。Mozer (1991) の BLIRNET は、網膜像における位置のずれを徐々に許容していくような構造を取り入れることによって、入力表現に関する問題を解決した。さらに、BLIRNET に注意のメカニズム等を加えた MORSEL は、様々な心理実験の結果に対して明確な説明を与えることができた。現在のところ、BLIRNET は単語認知過程に関するもっとも優れたモデルであるといえるだろう。今後、単語認知過程に関するさらに一般的なモデルを構築するためには、学習の過程を取り入れることが必要である。人間は学習によって様々な言語で書かれた文章を読むことができるようになる。より一般的なモデルは、各種の文字体系に対して学習を行い、それぞれの言語を使用した心理実験の結果を説明することができなければならない。

第3章では、文字の位置に関する情報の表現に焦点を当てて、モデルの妥当性を検討するための実験を行った。Mozer (1991) のモデル BLIRNET においては、文字クラスターユニットと呼ばれる内部表現が用いられている。このユニットは隣

接する文字同士の位置関係にもとづいて文字の位置を表現する。実験1では、このような文字位置の表現が実際の単語認知過程で用いられているのかどうかを検討した。被験者の課題は、瞬間的に呈示されるプローブの中にターゲット単語があったかどうかを判断することであった。プローブは2つの単語から構成されていた。ターゲットを構成する2つの文字は、常にプローブの中に含まれていた（e.g. ターゲット：実用，プローブ：実証 応用）。これら2つの文字のプローブの中での位置関係を単語内・単語間で独立に操作し、フォールスアラーム率（実際にはターゲットがなかったのに「あった」と答える反応）に対する効果を検討した。その結果、2つの文字の単語内における位置関係がターゲットとプローブとで同じである場合の方が、異なっている場合より有意にフォールスアラーム率が高かった。これに対して、単語間での位置関係はフォールスアラーム率に影響を与えなかった。これらの結果は、文字の位置が単語内での相対的位置関係にもとづいて表現されていることを示しており、文字クラスターユニットのような内部表現の存在を支持している。

ところで、日本語の文章は、英語などとは異なり、空白によって単語を区切らないで表記される。実験2では、このような表記の特性が単語認知過程に対してどのような影響を与えているのかを検討した。実験2で使用したプローブは、プローブを構成する単語に隣接する空白を、ランダムにひらがなで置き換えたものであった（e.g. あ実証き応用せ）。実験1と同様な手続きで実験を行った結果、実験1と同じく、単語内における位置関係の効果が有意にみられた。これは、漢字とひらがなの区別が文字同士の相対的位置関係に関する情報を抽出する以前に行われていることを示している。日本語の認知過程においては、漢字とひらがなの境界が、文を単語に分節化するための手がかりとして用いられているのかもしれない。一方、実験1ではみられなかった単語間における位置関係の効果も有意にみられた。これは、複数の単語から構成される刺激全体の大局的特徴を処理する過程が存在することを示唆している。

第4章では、主成分分析による情報圧縮という枠組みが、単語の内部表現の生成過程に対しても適用できるのかどうかについて検討した。さらに、ニューラルネットワークを用いた単語の内部表現に関する研究が、言語という記号系の理解のためにどのような意味を持っているのかについての考察を行った。

Internal representation of letters and words

Abstract

In chapter 1, we investigated the internal representation of 2-dimensional patterns. Psychological space of shapes has been studied by many experiments. However, how shapes are represented in the brain has not been a major issue in psychological literature. We considered the characteristics of internal representation and how it was formed, and tried to explain the results of experiments in a unified manner. First we reexamined the data of similarity of alphabetic characters and random dot patterns. Multivariate analysis suggested that those patterns were represented by the combination of global features. Secondly we trained 3-layer neural networks to perform categorization or identity transformation of the same sets of patterns as used in psychological experiments, and analyzed activation pattern of the hidden units. When the network learned categorization of the patterns, its internal representation was not similar to the representation suggested by psychological experiments. But a network which learned identity transformation of the patterns could acquire such an internal representation. The transformation performed by this kind of network is similar to principal component analysis in that it projects the input image onto a lower dimensional space. From these results we propose that 2-dimensional shapes are represented in human brain by a process like principal component analysis. This idea is compatible with the findings in neurophysiological studies about higher visual areas.

In chapter 2, first we reviewed the studies on visual agnosia (impairment in the higher visual processes), and considered how the impairment is related with the word recognition process. Secondly, we reviewed psychological studies on word recognition. Especially, we attended to word superiority effect and letter migration errors. Thirdly, we examined the neural network models of word recognition. The interactive activation model (McClelland & Rumelhart, 1981) successfully explained the recognition process of one word. McClelland (1986) improved the model in order to explain the recognition process of multiple words. The model was called programmable blackboard model, which could explain how letter migration

errors occur. However, the model was too complex and lost compatibility with neurophysiological evidence. The problem of this model lies in its input representation. Mozer (1991) developed the model called BLIRNET, which used more reasonable input representation. BLIRNET had the simple architecture which is compatible with neurophysiological findings, and could explain the results of many psychological experiments, including letter migration errors. In order to develop a more general model on word recognition, we have to realize the process of learning. Such a model should be able to learn multiple languages.

In chapter 3, two experiments investigated the representation of letter position in visual word recognition process. In experiment 1, subjects were asked to detect a target word in a briefly-presented probe. Probes consisted of two *kanji* words. The letters which formed targets (critical letters) were always contained in probes. (e.g. target:実用 probe:実証 応用) High false-alarm rate was observed when critical letters occupied the same within-word relative position (left or right within the word) in the probe words as in the target word. The effect of within-word relative position was explained by BLIRNET, which used letter cluster units as internal representation of words. In experiment 2, spaces adjacent to probe words were replaced by randomly chosen *hiragana* letters (e.g. あ実証き応用せ), because spaces are not used to separate words in regular Japanese sentences. As in experiment 1, the effect of within-word relative position was observed. This result suggest that segmentation between *kanji* and *hiragana* was performed before the information about within-word relative position was extracted. On the other hand, the effect of between-word relative position (left or right between the word) was also observed. This result suggests the existence of the process which represented the shape of whole stimuli.

In chapter 4, we considered whether the internal representation of letters and words is generated by the process like principal component analysis. Furthermore, we considered the significance of our study in terms of evolution.

目次

序.....	1
第1章 文字の内部表現.....	3
第1節 神経生理学的な研究.....	3
1. 高次視覚野における形の表現.....	3
2. 高次視覚野における顔の表現.....	4
3. 心理実験との関連.....	5
第2節 文字の類似度に関する心理実験.....	8
1. 26個のアルファベット文字の類似度.....	8
2. 6個のアルファベット文字の類似度.....	11
3. ハトを被験体として用いた実験.....	12
4. ランダムドット・パターンの類似度.....	13
5. 心理実験のまとめ.....	14
第3節 シミュレーション.....	15
1. シミュレーション1：分類学習.....	16
1.1. 26個のアルファベット文字に関するシミュレーション.....	19
2. シミュレーション2：恒等変換.....	23
2.1 26個のアルファベット文字に関するシミュレーション.....	26
2.2 6個のアルファベット文字に関するシミュレーション.....	28
2.3 ランダムドット・パターンに関するシミュレーション.....	29
3. シミュレーション3：複雑型細胞を用いた入力表現.....	31
3.1 ランダムドット・パターンに関するシミュレーション.....	34
4. シミュレーションのまとめ.....	35
第4節 考察.....	36
1. 心理実験のデータに関する再考.....	36
2. 主成分分析との関係.....	37
3. 学習規則について.....	38
4. 神経生理学的な研究との関係.....	39
第2章 単語認知過程のモデル化.....	42
第1節 文字認知から単語認知へ.....	42
1. 単語認知と文字認知との関係.....	42
2. 失読と単語認知.....	44
3. 注目すべき心理現象.....	49
3.1 単語優位効果.....	49
3.2 文字の混合エラー.....	50
第2節 単語認知過程のモデルの発展.....	55
1. Interactive Activation Model (IA)	55
2. Programmable Blackboard Model (PABLO)	60
3. BLIRNETとMORSEL	65
4. まとめ.....	73

第3章 単語認知過程における文字位置情報の表現.....	75
第1節 文字位置情報と文字の混合エラー.....	75
1. 文字位置情報の役割.....	75
2. 文字の混合現象に対する文字位置の効果.....	75
第2節 実験.....	77
1. 実験1：単語内文字位置と単語間文字位置の効果.....	77
2. 実験2：漢字とひらがなの分節化.....	86
第3節 考察.....	93
1. モデルに与える示唆.....	93
2. より一般的なモデルに向けて.....	94
第4章 総合的な考察.....	97
1. 単語認知過程と主成分分析.....	97
2. より広い意味での情報圧縮.....	99
3. 進化という観点から.....	101
第5章 総括.....	104
注.....	107
参考文献.....	110
補遺.....	115
研究業績.....	118
謝辞.....	119

序

文字や単語の視覚的な認知過程は、認知心理学が成立して以来、その主要なトピックの一つであり続けてきた⁽¹⁾。様々な刺激や課題を用いて実験が繰り返され、数多くのモデルが提案されている。

文字の使用が、わずか数千年前という、人類の歴史全体から見れば比較的最近に始まったことであることを考えると、このような状況はいささか奇妙なものに感じられる。(いわゆる)先進国に住む現代人にとっては、文字や単語を読むことはごく当たり前の行為であり、日常生活の中で大きな比重を占めている。しかし、動植物の種類を見分けること、人の顔を覚えること、道具を作ること、道順を記憶すること、そして会話することなどに比べれば、読むことは人間の認知的能力の中でも比較的特殊な技能の部類に入るのではないだろうか。それにも関わらず、文字や単語の認知過程に関する研究がさかんに行われている背景には、他の認知的能力を対象とする場合と比べて統制された実験を行うことが容易であるという理由があるだろう。

しかし、理由はそれだけではない。文字や単語の認知過程を研究することには、知覚と認知のつながりを探る、という重要な意味がある。知覚的な過程によって再構成された外界の像の中から、われわれは様々な対象を認知する。文字という後天的に学習され、限定された集合を構成する対象を認知するとき、われわれの脳内にはどのような表現が形成されるのだろうか。そして、文字の組み合わせによって定義される単語に関してはどのような表現が用いられているのだろうか。さらに、文字や単語の表現と、意味や文法などのよ

り高次の表現との間には、どのような関係があるのだろうか。

文字や単語の認知過程を、そのためだけに特殊化したシステムによって説明することは、それらが学習によって獲得されるものであることを考えると不自然に思われる。他の対象の認知に用いられているものと共通のシステムが、文字や単語を学習することによってそれらの効率的な認知を行うことができるようになっていくと考えた方が説得力がある。そこで本論では、ニューラルネットワークの学習による内部表現の獲得という観点から、文字と単語の認知過程を包括的に理解することを目指す。第1章では、文字やランダムドット・パターンの類似度に関する先行研究のデータを系統的に再分析し、ニューラルネットワークによるシミュレーションの結果と比較することによって、これらのパターンの内部表現がどのような性質をもつのかを検討する。第2章では、単語の認知過程に関する生理学・心理学的な研究について概説し、これまでに提案されたニューラルネットワーク・モデルについて考察を加える。第3章では、文字の位置に関する情報の表現に焦点を当てて、モデルの妥当性を検討するための実験を行う。第4章では、総合的な考察と今後の展望に関する議論を行う。

第1章 文字の内部表現

第1節 神経生理学的な研究

1. 高次視覚野における形の表現

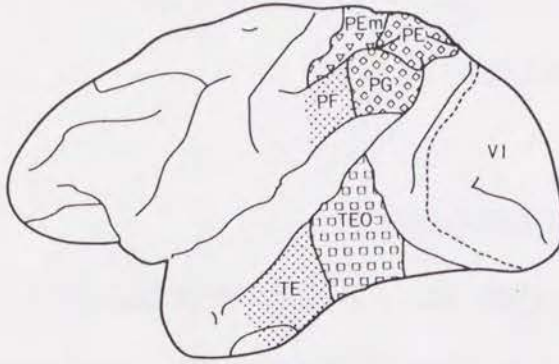


図 1.1 サルの脳における下部側頭葉の位置. TE, TEO と記された部分が下部側頭葉である. VI と記された部分は一次視覚野と呼ばれる.
(図は乾(1993)より抜粋.)

近年の神経生理学における知見は、脳内において、対象^②の形がある程度複雑な特徴の組み合わせによって表現されていることを示唆している。Fujita, Tanaka & Cheng (1992) は、サルの下部側頭葉（図 1.1）における細胞の反応を微小電極を用いて測定した。下部側頭葉は脳内において形を分析する経路の最終段階に位置している。彼らは、まず動物のぬいぐるみなどの複雑な形状をもつ対象に対する細胞の反応を測定し、刺激の形を徐々に単純化していくことによって、その細胞が反応するために必要な最小限の特徴を特定した。その結果、下部側頭葉の細胞は、上下逆向きの T 字型や突起のついた円盤などの、比較的複雑な形に対して選択的に反応することが分かった。これらの細胞は、いわば中程度に複雑な特徴を表現していると

考えられる。さらに、彼らはこれらの細胞が受容野の類似度に応じてコラム状に並んでいることを発見した。すなわち、それぞれのコラムの中には、互いに類似した形の刺激に対して反応する細胞が集まっていた。このような結果から、対象の形は単一の細胞の活動ではなく、複数の細胞の活動パターンによって表現されているのではないかと彼らは主張している。

2. 高次視覚野における顔の表現

Young & Yamane (1992) は、サルに顔の弁別課題を行わせ、その間の下部側頭葉前部 (AIT) と側頭葉多感覚野上部 (STP) における細胞の活動を記録した。彼らはこれらの細胞の活動パターンを多次元尺度法 (MDS) を用いて分析した。MDS は高次元空間における刺激間の距離関係を、元の構造をできるだけ保存しながら低次元空間に再現する。AIT の場合、それぞれの刺激に対して 41 個の細胞の活動パターンが記録された。このデータをそのままの形で把握することは困難であるが、MDS を用いて低次元空間における刺激の布置を導出することによって、細胞の活動パターンが全体として何を表現しているのかを理解することが可能になる。各領野ごとに分析を行った結果、STP の細胞は顔の親密度 (familiarity) を表現していることが分かった。これに対して、AIT の細胞の活動パターンは、顔の物理的な特性を表現していた。MDS によって導出された 2 つの次元は、それぞれ髪の量と顔の細長さに対応していた。Fujita, et al. (1992) の結果と同様、顔の形は比較的複雑な特徴の組み合わせによって表現されていると考えられる⁽⁹⁾。本論では、このような特徴のことを、対象の全体的な特徴を表現しているという意味で大局的な特徴 (global

feature) と呼ぶことにする。

3. 心理実験との関連

一方、心理学においては、形の心理的な類似度を測定することによって形の心理空間の性質を探ろうとする試みが数多く行われている。たとえば、Gilmore, Hersh, Caramazza, & Griffin (1979) は文字の同定課題を用いてアルファベット文字間の混同確率を測定した。その結果得られた混同行列を MDS を用いて分析することによって、彼らはアルファベット文字の心理空間がどのような構造をもっているのかを検討した。彼らの実験から得られた混同行列の一部を表 1.1 に示す。

表 1.1 Gilmore et al. (1979) の実験で得られた混同行列の一部。

		反応				...
		A	B	C	D	
刺 激	A	.278	.018	.007	.008	...
	B	.022	.155	.022	.078	...
	C	.005	.007	.545	.017	...
	D	.011	.039	.045	.379	...

.	
.	

この行列は、特定の文字が呈示されたときに、被験者がそれぞれの文字を報告した比率を表す。この行列を行方向（横方向）にたどっていくと、特定の刺激文字に対して被験者が各アルファベット文字を答えた割合がわかる。たとえば、"B" という刺激が呈示されたときに被験者が "D" と答えた比率は.078 である。この行列を刺激間の類似度と対応づけるためには、対称な形に変換する必要がある。彼

らは、Luce の選択モデル (Luce, 1963) を用いて、この行列を対称行列と各文字に対するバイアスを表すベクトル⁽⁴⁾に分解した。このようにして得られた対称行列は、文字間の心理的な類似度を反映しているともみなすことができる。

このような行列を MDS (多次元尺度法) を用いて分析することによって、文字間の類似度がもつ構造を把握することが容易になる。

MDS は、多数の項目間の類似度 (距離) に関するデータから、多次元空間における項目の布置を導出する手法である。このとき、もとのデータにおける類似度と、導出された空間における距離は、できるだけ一致している必要がある。すなわち、類似度の高い項目同士は導出された布置においても互いに近くに位置し、類似度の低い項目同士は互いに遠くに位置しなければならない。両者の不一致の度合いはストレスと呼ばれる値によって表される (ストレスの値が小さいほど適合度が高い)。MDS は、ストレスを最急降下法を用いて最小化することによって、もとのデータとの適合度ができるだけ高い解を算出する。一般に、最終的な解のストレスの大きさは、導出する次元の数が多くなるほど低下する傾向がある。一方、結果の解釈を容易にするためには、次元の数は少ない方が望ましい。

Gilmore et al. (1979) は、MDS による分析にもとづいて、アルファベット文字の心理空間について検討した。しかし、彼らの分析は主として記述的なものであり、脳内表現と関連付けようとする試みは行われていない。これまでの心理学的研究全般においても、3次元の形状の脳内表現に関する研究 (e.g. Edelman, 1995) は数多くあるものの、2次元的な形の脳内表現に関してはあまり検討されてこなかった。2次元パターンに関する従来の研究においては、形は主として

それを構成する特徴という観点から分析され、その特徴は多くの場合実験で使用された刺激に固有のものであった。複数の心理実験の結果を統一的に説明することができるような、より一般的な枠組みが必要とされている。文字パターンが脳内においてどのように表現され、さらにその表現がどのように生成されるのかを検討することによって、そのような枠組みを構築することが可能なのではないだろうか。

この章では、最初に、アルファベット文字やランダムドット・パターンの類似度を測定した心理実験の結果を MDS を用いて分析する。前述のように、神経生理学における研究は対象の形が複数の細胞の活動パターンによって表現されていることを示唆している。これは形が脳内における多次元空間の中の点として表現されていることを意味する。この空間における距離は心理学的な類似度に反映されると考えられる。したがって、心理実験の結果を MDS を用いて分析することによって得られる空間の構造は、脳内表現の構造を反映していると仮定することができる。

次に、ニューラルネットワーク⁶⁾を用いてシミュレーションを行う。人工的に構築したニューラルネットワークに心理実験で使用されたものと同じ刺激を学習させ、獲得された内部表現を分析する。ネットワークの内部空間における文字間の類似度を MDS を用いて分析することによって、ネットワークの内部空間の構造を記述し、さらに心理実験の結果から推定される脳内表現との比較を行うことができる。このシミュレーションの目的は、心理実験の結果そのものを再現することではなく、そこから推定される内部表現を再現することにある。

第2節 文字の類似度に関する心理実験

1. 26個のアルファベット文字の類似度

Gilmore et al. (1979) は 26 個のアルファベット文字間の混同確率を測定した。実験で使用されたフォントを図 1.2 に示す。被験者の行った課題は、瞬間呈示される文字を同定して口頭で報告することであった。刺激の呈示時間は平均正答率が 50% に近づくように被験者ごとに調整された。実験の終了後、被験者がある文字を別の文字と間違えた確率から、文字間の混同行列が作成された。

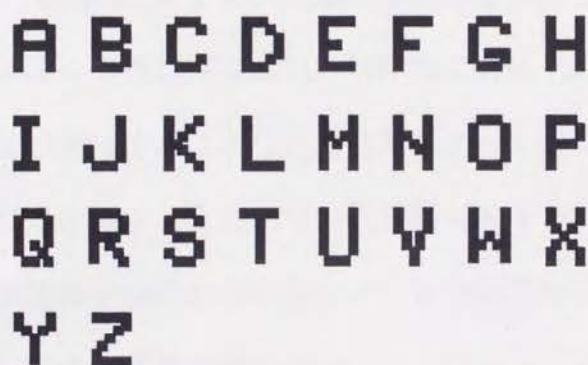


図 1.2 Gilmore et al. (1979) の実験で使用されたフォント。

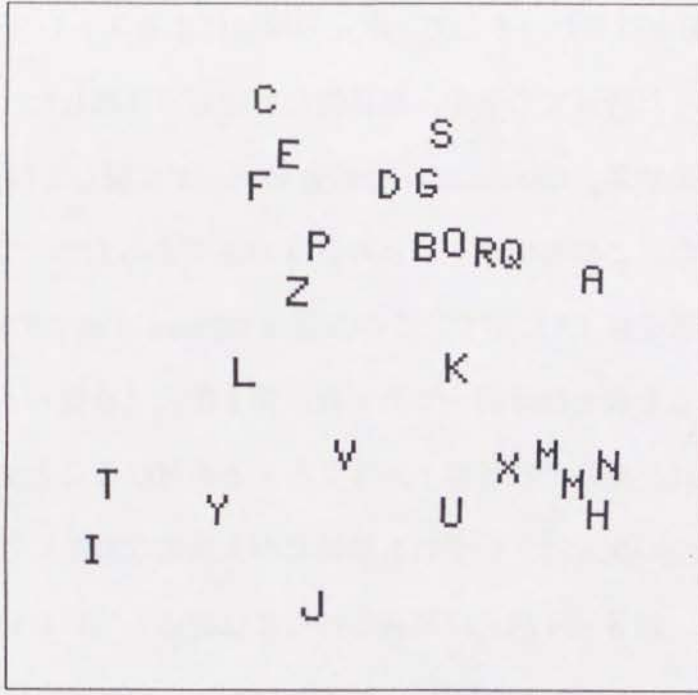
前述のように、Gilmore et al. (1979) 自身も MDS による分析を行っている。しかし、他の分析との一貫性を保つために、本論では統計パッケージ SAS の MDS プロシジャ (SAS Institute Inc., 1992) を用いて彼らのデータを再分析することにした。分析には順序尺度モデルを使用した。このモデルは、もとのデータにおける各刺激の類似度の順序をできるだけ保存しながら、低次元空間における刺激の順序を導出する。順序尺度モデルにおいては、もとの空間における順序と、分析によって導き出された布置における順序の一致する度合いにもとづいてストレスの値が算出される (Kruskal の公式 1)。ここ

では、まず1次元解から開始して、次元を1個ずつ増やしながら繰り返しMDSを実行した。そして、次元の増加によるストレスの低下がほぼ止まった時点での解を、最終的な解として採用した。このような手続きの結果、Gilmore et al. (1979) のデータに関しては5次元解を採用した。この解のストレスの値は.008であった⁶⁾。

得られた布置を図1.3に示す。この布置はGilmore et al. (1979) 自身の分析によるものとほぼ同一であった。第1次元（横軸）と第2次元（縦軸）は、それぞれ「幅」、「丸み」と名付けることができる。これら2つの次元は、いずれも単純だが大局的な特徴に対応している。一方、第3・4次元は単純な特徴には対応していない。Gilmore et al. (1979) は、この第3・4次元について「知覚的な構造は特定できるが、それに対応する物理的な次元を名付けることができないという状況に直面している (pp.430)」と述べている。

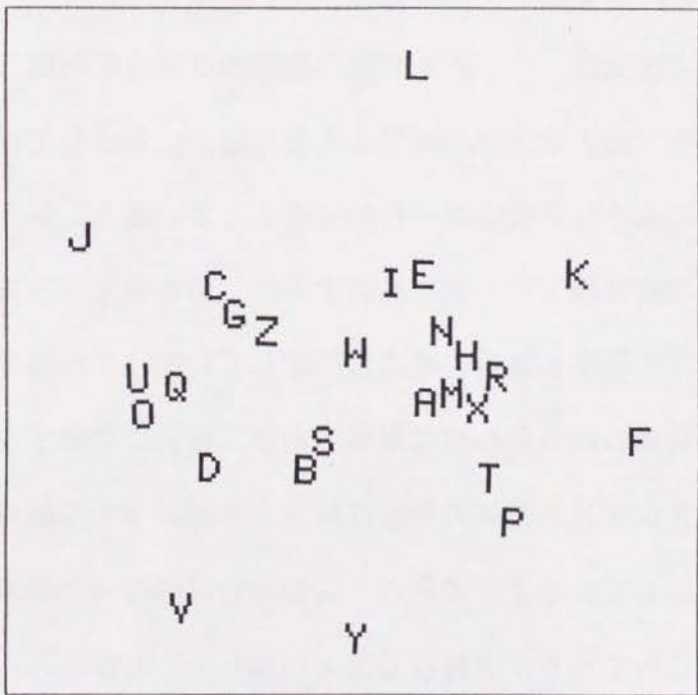
これら4つの次元は、いずれも刺激の物理的な特性に対応している。第1次元の「幅」は文字の両側の垂直な線分に対応している。ある文字がそのような特徴の組み合わせをもっていると、第1次元の値が大きくなる（図上で右側に移動する）。第2次元の「丸み」は文字の四隅に位置する線分の端点に対応している。これら2つの次元の場合は特徴の組み合わせは単純であり、容易に命名することができる。第3次元は、左側の垂直な線分と上部の水平な線分の組み合わせに対応しているようである。この場合は異なった種類の特徴が組み合わさっており、名付けることが難しい。しかし、この次元も1つの単純な特徴に還元できないことは確かである。第4次元に関しても同じことがいえる。4つの次元は、いずれも文字の大局的な特徴に対応しているという点で共通している。

第2次元



第1次元

第4次元



第3次元

図 1.3 Gilmore et al. (1979) のデータから得られたアルファベット文字の布置.
 (上) 第1・第2次元. (下) 第3・第4次元.

2. 6個のアルファベット文字の類似度



図 1.4 Rumelhart (1971) のデータから得られたアルファベット文字の布置. 実験で使用されたフォント・パターンを下に示す.

Rumelhart (1971) は Gilmore et al. (1979) と同様な手続きで実験を行い、A から F までの 6 個のアルファベット文字間の混同確率を測定した. 使用されたフォントを図 1.4 (下) に示す. Gilmore et al. (1979) のデータの場合と同一の手続きで混同行列を対称行列に変換し、MDS による分析を行った. 採用した解は 2 次元解で、ストレスの値は .01 であった. 得られた布置を図 1.4 (上) に示す. 第 1・第 2 次元とも、単一の線分の有無のような単純で局所的な特徴には対応していないことが分かる. 各次元に対応する特徴は Gilmore et al. (1979) の場合と異なっているが、それぞれの次元が文字の大局的な

特徴に対応しているという点では一致している。

3. ハトを被験体として用いた実験

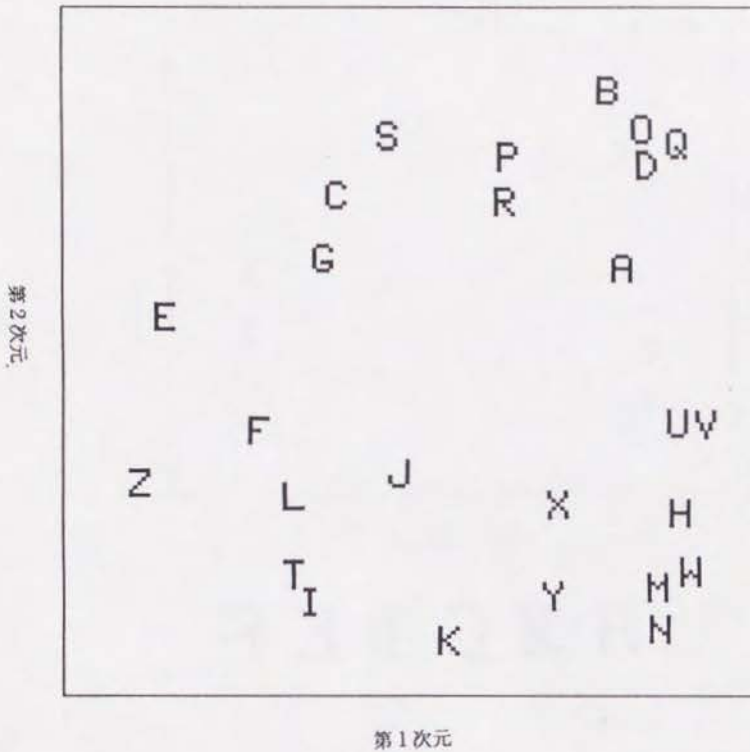


図 1.5 Blough (1985) のデータから得られたアルファベット文字の布置. この実験では被験体としてハトが用いられた.

Blough (1985) はハトがアルファベット文字を弁別する際の混同確率を測定した. 被験体のハトは, 個々のアルファベット文字を他のすべての文字と弁別できるように条件付けされた. テスト試行における弁別時のエラーにもとづいて, 文字間の混同行列が作成された. 使用されたフォントは Gilmore et al. (1979) のものとほぼ同じであった. 前回と同様な分析によって得られた布置を図 1.5 に示す. 採用した解は 5 次元解で, ストレスの値は .06 であった. この布置は Gilmore et al. (1979) のデータから得られたものとよく似ている. 両者の布置が類似していることは, アルファベット文字の心理空間が

主として刺激の物理的属性によって決まっていることを意味している。発音や意味などの人間に固有の経験は、文字の内部表現に対してあまり影響を与えていないようである。

4. ランダムドット・パターンの類似度

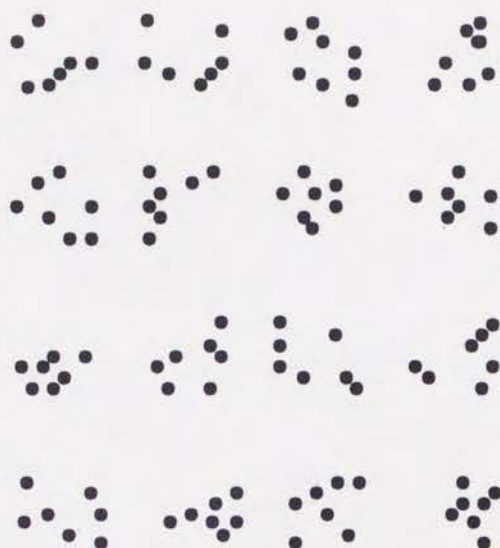


図 1.6 Blough (1985) の実験で使用されたランダムドット・パターン。

Blough (1985) は、人間の被験者を対象としてランダムドット・パターン間の心理的な類似度を測定した。使用された刺激を図 1.6 に示す。被験者の課題は、同時に呈示された2つのランダムドット・パターン間の類似度を評定することであった。実験の終了後、すべてのパターン同士の組み合わせに関する評定値にもとづいて、パターン間の類似度行列が作成された。類似度は混同確率と同じく刺激間の距離を反映しているので、これまでと同様な手続きで MDS による分析を行うことができる。得られた布置を図 1.7 に示す。採用した解は4次元解で、ストレスの値は .06 であった。第1次元（横軸）

は刺激の「大きさ」に、第2次元（縦軸）は刺激の「規則性」に対応している。ここでも、得られた次元は刺激の大局的な特徴に対応している。

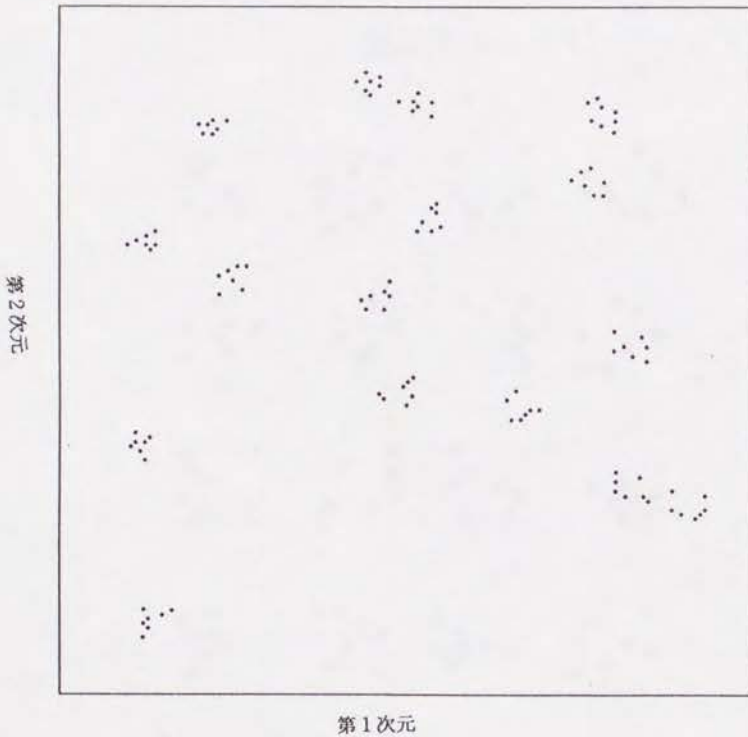


図 1.7 Blough (1985) のデータから得られたランダムドット・パターンの布置.

さらに、Blough (1985) はこれと同じ刺激を用いてハトを対象とした実験も行っている。アルファベット文字の場合と同様、ハトはランダムドット・パターンの弁別を行うように条件づけられ、テスト試行における混同確率が測定された。MDS による分析の結果得られた布置は、人間の場合とほぼ同様であった。

5. 心理実験のまとめ

文字やランダムドット・パターン間の混同確率や類似度を測定し

た先行研究の結果を再分析したところ、すべての場合において、MDSによって得られた布置は刺激の大局的な特徴に対応していた。この結果は神経生理学における知見と一致している。文字やランダムドット・パターンなどの2次元的な形は、脳内において大局的特徴の組み合わせによって表現されているのではないかと考えられる。

第3節 シミュレーション

文字間の混同確率を説明するためのモデルは、これまでもいくつか提案されている。たとえば、Rumelhart (1971) は一群の特徴検出器の活動によって混同確率を予測するモデルを提案した。このモデルは、先に分析した彼自身の実験の結果をうまく説明することができた。しかし、仮定された特徴検出器は実験で使用されたフォントに合わせて作られたものであり、他の実験の結果にそのまま適用することはできない。

本論では、前節で再分析を行った一連の実験の結果を統一的に説明できるような枠組みを構築することを目指す。神経生理学的な研究は、高次視覚野において対象の形が複数の細胞の活動パターンによって表現されていることを示している。個々の対象は、それらの細胞によって構成される多次元空間内の点として表現されていると考えられる。一方、心理実験によって得られた混同行列や類似度行列は、刺激パターン間の心理的な距離関係を表している。このような心理的な距離関係は、脳内の多次元空間の構造を反映していると仮定できる。この仮定にもとづくと、混同行列や類似度行列をMDSを用いて分析することによって得られた布置を、脳内表現の構造を示す指標とみなすことができる。そのような脳内表現の生成過程に

ついて検討するために、ニューラルネットワークに心理実験で使用されたものと同一の刺激パターンを学習させ、獲得された内部表現を分析した。

1. シミュレーション1：分類学習

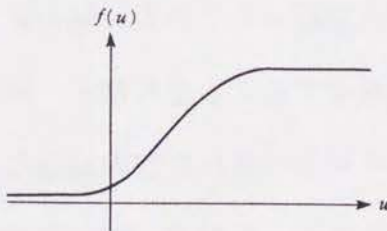
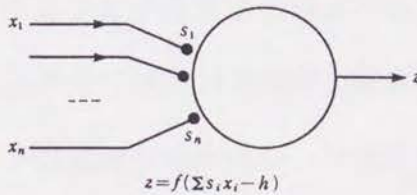
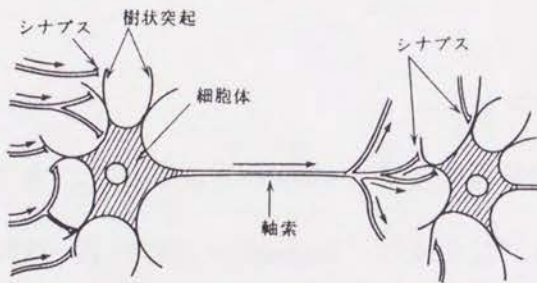


図1.8 (上) 神経細胞の模式図。神経細胞は軸索を通して他の神経細胞に信号を送る。送られた信号は、樹上突起上のシナプスと呼ばれる部分で受け取られる。

(中) 神経細胞の数理モデル。各ユニットへの入力信号は、信号の強度と結合の強度の積和によって決まる。すなわち、信号の強度を x_1, x_2, \dots, x_n 、結合の強度を s_1, s_2, \dots, s_n 、ユニットの閾値を h とすると、ユニットへの入力 u は、

$u = \sum_{i=1}^n s_i x_i - h$ となる。(下) ユニットの出力関数。ユニットの出力を z とすると、 $z = f(u)$ となる。図のような出力関数はシグモイド型と呼ばれる。

(図は甘利(1989)より抜粋。)

ニューラルネットワークは生体の神経系をモデル化したものであり、神経細胞と同様なふるまいをする多数のユニットから構成される。個々のユニットは他のいくつかのユニットからの信号を受け取り、それを総合して他のユニットへ信号を送る（図 1.8）。ユニット間の結合は、神経細胞同士が信号の受け渡しを行う部分であるシナプスに相当する。ユニットは並列に働き、ユニット同士の相互作用がネットワーク全体のふるまいを決める。個々のユニットは非常に単純なものであるが、多数のユニットから構成されるネットワークは複雑な情報処理を行うことができることが知られている（e.g. 甘利, 1989）。ニューラルネットワークは、ユニット間の結合の強度を変化させることによって学習を行う。ネットワークの獲得した知識は、多数のユニット間の結合パターンとして表現される。

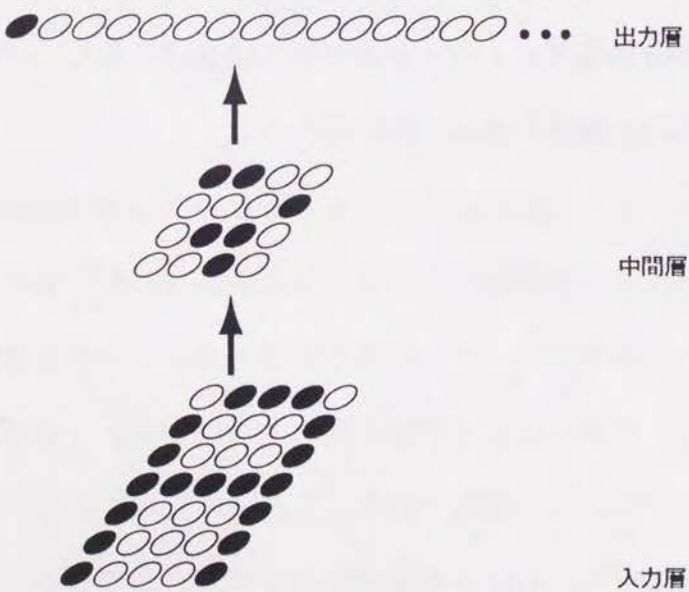


図 1.9 刺激パターンの分類を行うニューラルネットワーク。実際には、入力層のユニットは中間層のすべてのユニットと結合しており、中間層のユニットは出力層のすべてのユニットと結合している。

シミュレーションで使用したネットワークは、入力層・中間層・出力層の3つの層から構成されていた(図 1.9)。学習はバックプロパゲーション法(誤差逆伝搬法)を用いて行った(Rumelhart, Hinton, & Williams, 1986)。バックプロパゲーション法は教師あり学習規則の1種であり、ネットワークの出力パターンと教師パターンとの間の誤差(出力パターンと教師パターンの差の2乗和)が減少するようにユニット間の結合強度を変化させる(図 1.10)。このようなアルゴリズムは最急降下法の一種とみなすことができる。

ネットワークに入力パターンが呈示されると、ユニットの活性化度(activation)は中間層を経て出力層へ伝達される。出力層においてネットワークの生成した出力パターンと正しい出力パターン(教師パターン)が比較され、両者間の誤差が計算される。次に、この誤差の大きさに従って、出力層と中間層間の結合の強度が修正される。さらに、誤差信号はネットワークの中を逆向きに伝搬し、中間層と入力層間の結合強度の修正に用いられる。

シミュレーション1においては、ネットワークは26個の出力ユニットのうちただ1つを活性化させるように訓練された。学習の終了後、獲得された内部表現、すなわち中間層のユニットの活動パターンを分析した。学習された文字は中間層のユニットから構成される多次元空間内の点として表現される。これらの点同士の距離は、ネットワークの内部空間における類似度とみなすことができる。したがって、心理実験の場合と同様にMDSを用いて分析することによって、ネットワークの内部空間の構造を記述し、さらに心理実験の結果から推定される脳内表現との比較を行うことができる。

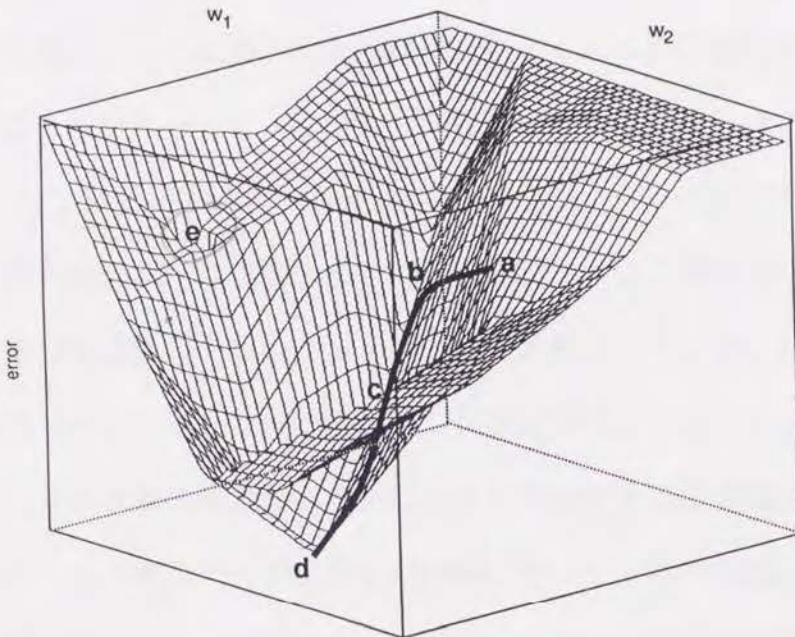
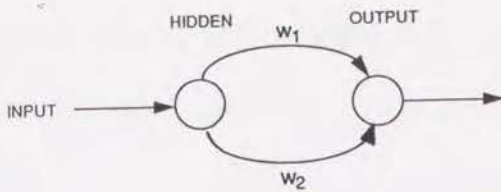


図 1.10 結合を 2 つしか持たない単純なネットワークを想定する。2 つの結合の強度によって構成される平面に対して、誤差の大きさを高さとしてプロットすると、上のような曲面を描くことができる。ネットワークの結合強度の初期値が点 a のような値をとっているとす。バックプロパゲーション法は、その地点から見て誤差が最も減少する方向へ結合強度を変化させる。その結果、矢印のような軌跡をたどって結合強度が変化していき、最終的に誤差の大域的な最小値に到達する。結合パターンが局所的な最小値（図中の点 e ）に落ち込んでしまう可能性も存在するが、初期値を変えてシミュレーションを繰り返すことによって、局所的な最小値を最終的な解として採用してしまふことを避けることができる（図は Elman (1992) より抜粋）。

1.1. 26 個のアルファベット文字に関するシミュレーション

ネットワーク Gilmore et al. (1979) の実験で使用されたフォントにもとづいて入力パターンを作成した。このフォントは、縦 $7 \times$ 横 5 個の、計 35 個の点から構成されていた。そこで、各文字を縦 $7 \times$

横 5 個の升目で区切り，その中の点が存在すれば 1 を，存在しなければ 0 を当てはめることによって 35 次元の 2 値ベクトルを作成し，入力パターンとして使用した。

出力層はアルファベットの各文字に対応する 26 個のユニットで構成されていた。入力層の各ユニットは中間層のすべてのユニットと結合し，中間層の各ユニットは出力層のすべてのユニットと結合していた。これらの結合はすべて前向きであり，フィードバック型の結合は存在しなかった。結合強度の初期値は，-3 から .3 までの一様乱数の中から選択した。中間層のユニットの個数は，結合強度の初期値に関わらず安定した結果が得られるような最小の個数を探索的に求めることによって決定した（これ以降のシミュレーションにおいても，同様な手続きで中間層のユニットの個数を決定した）。中間層と出力層のユニットの出力関数はシグモイド型であった。その後の分析を容易にするために，各ユニットの閾値（バイアス）は 0 に固定した。入力パターンはランダムな順序でネットワークに呈示され，結合強度の更新はその都度行われた。学習は出力層における誤差が十分に小さくなるまで続けられ，総計 20 万回の学習が行われた。シミュレーションの実行にはワークステーション (SUN: Sparc 330) を用いた（補遺 1 参照）。

結果と考察 学習の終了後，それぞれの入力パターンを呈示した際
の中間層のユニットの活性度パターンを記録した。中間層は 16 個の
ユニットから構成されるので，ある文字を入力したときのユニット
の活性度パターンは 16 次元空間における点とみなすことができる。
そこで，文字間の距離をこの空間内でのユークリッド距離として定
義し，距離行列を作成した。この行列は，心理実験から得られた混

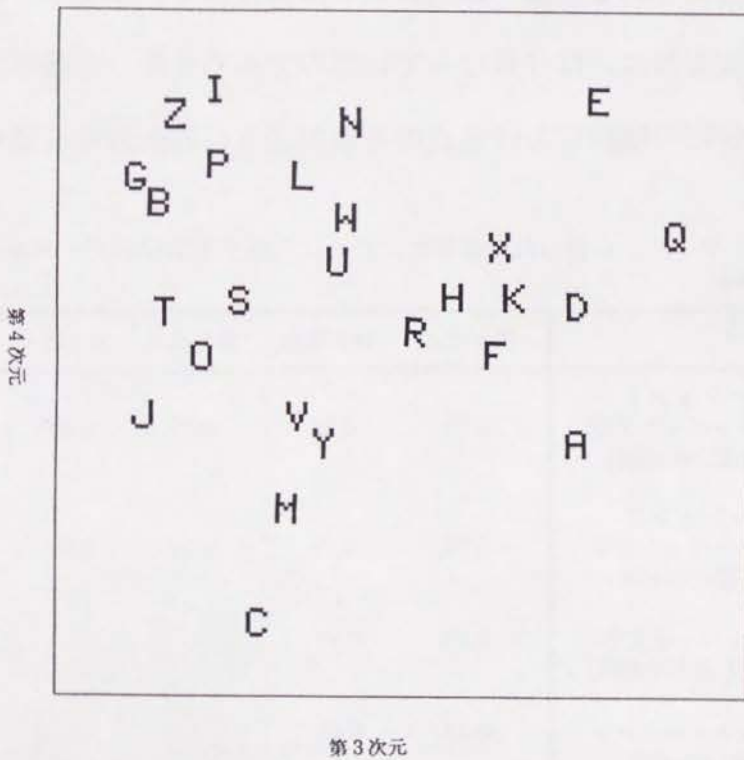
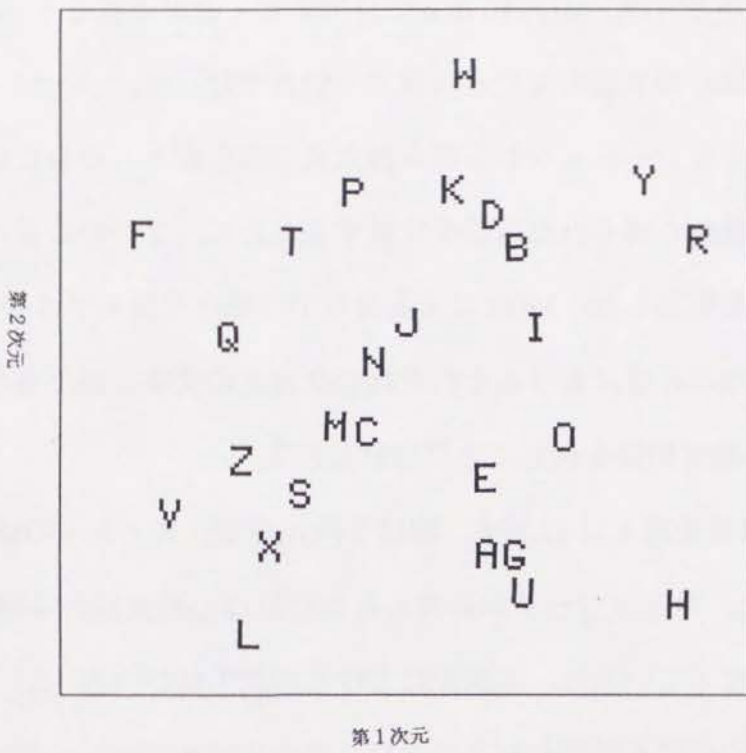


図 1.11 Gilmore et al. (1979)のフォントの分類を学習したニューラルネットワークの内部表現から得られたアルファベット文字の布置。(上)第1・第2次元。(下)第3・第4次元。

同行列の場合と同様，MDSによって分析することができる．MDSによる分析は心理実験の場合と同様な手続きで行った．

次に，シミュレーションから得られた文字の布置と，それに対応する心理実験から得られた布置を比較するために，2つの布置の間の相関係数を算出した．MDSによる分析では順序尺度モデルを用いたので，2つの布置におけるそれぞれの次元上の文字の順序を求め，Spearmanの順位相関を次元ごとに計算した⁷⁾．

MDSの結果を図1.11に示す．解は5次元解で，ストレスの値は.11であった．ネットワークから得られた布置は心理実験から得られた布置と異なっており，各次元における相関も低い(表1.2)．入力パターンの分類を学習したネットワークの内部表現は，心理実験の結果から推測される人間の脳内表現とは異なっているようである．

両者の内部表現は，なぜ異なっていたのであろうか．分類学習の問題点は，特定の場所における点の有無のような局所的な特徴に注

表1.2 ネットワークから得られた布置と，それに対応する実験のデータから得られた布置の相関．

	第1次元	第2次元	第3次元	第4次元
シミュレーション1 26個のアルファベット文字 (Gilmoreの布置との相関)	0.21	0.22	0.29	0.46
シミュレーション2 26個のアルファベット文字 (Gilmoreの布置との相関)	0.79	0.73	0.69	0.39
6個のアルファベット文字 (Rumelhartの布置との相関)	0.94	0.94		
ランダムドット・パターン (Bloughの布置との相関)	0.25	0.44		
シミュレーション3 ランダムドット・パターン (Bloughの布置との相関)	0.76	0.66		

目するだけで正しい分類が可能になってしまうところにある。ネットワークはそのような特徴についてのみ学習するだけで出力層における誤差を十分に小さくできるので、それ以上の学習を行わない。その結果、ネットワークは大局的な特徴を用いなかったのではないかと考えられる。入力層から中間層への結合パターンを実際に調べた結果も、この見方を裏付けるものであった。

2. シミュレーション2：恒等変換

シミュレーション2では、ネットワークに入力パターンの恒等変換を学習させた。すなわち、入力されたものと同一のパターンを再構成するようにネットワークを訓練した（図 1.12）。

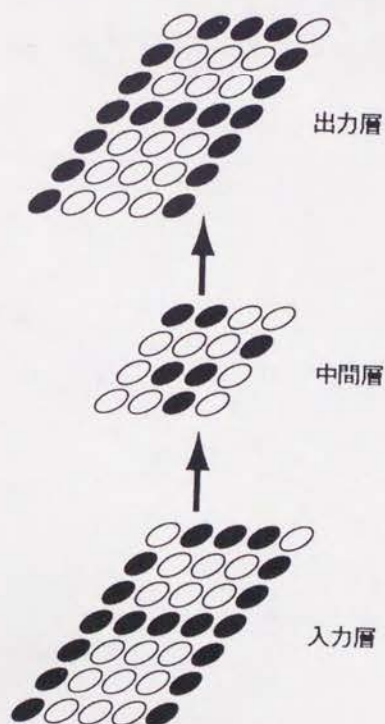


図 1.12 刺激パターンの恒等変換を行うニューラルネットワーク。

このような学習を行う場合、中間層のユニットの数を入力層や出力層より少なくすると、中間層において情報が圧縮されることになる。この種のネットワークは砂時計型のネットワークと呼ばれる。このようなネットワークが行う変換は、主成分分析とよく似た性質をもつことが知られている (Cottrell, Munro, & Zipser, 1988)。

主成分分析は、データのもっている情報をなるべく損なわないように、より低い次元の空間にデータを投影する手法である。図 1.13 のような 2 変数のデータを例にとりて考えてみよう。最初に、データ全体の分散が最大であるような方向を示す直線を決める（この直線は、データの各点から垂直に下ろした線分の長さの平方和が最小になるような直線として定義される）。この直線が第 1 主成分である。このような直線上に投影した点は、2 次元空間におけるもとのデータの散布の仕方をもっとも良く保存する。第 2 主成分は第 1 主成分に直交する直線である。

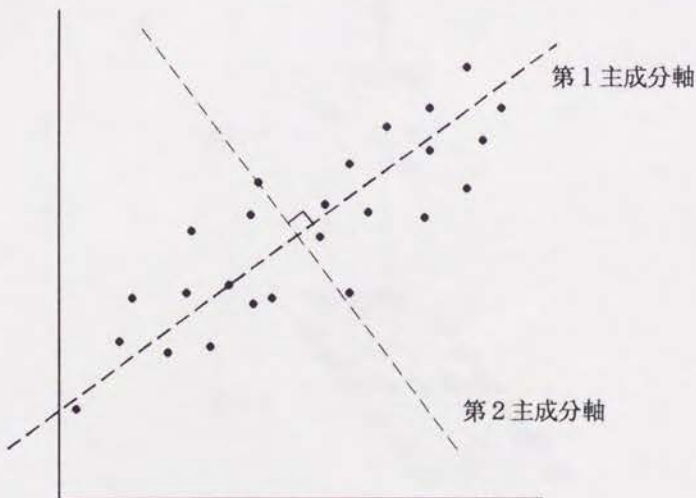


図 1.13 もっとも単純な主成分分析の例。

もとのデータが3つ以上の変数をもつ場合には、上と同じ方法で第1主成分を決めた後、第1主成分に直交する直線の中でデータ全体の分散が最大であるような方向を示すものを、第2主成分として選択する。このような手続きを繰り返すことによって、 p 個の変数 x_1, x_2, \dots, x_p を、 m 個の変数 y_1, y_2, \dots, y_m に要約することができる ($m \leq p$)。変換後の変数 y_j は、

$$\begin{aligned} y_1 &= h_{11}x_1 + h_{12}x_2 + \dots + h_{1p}x_p \\ y_2 &= h_{21}x_1 + h_{22}x_2 + \dots + h_{2p}x_p \\ &\dots \\ y_j &= h_{j1}x_1 + h_{j2}x_2 + \dots + h_{jp}x_p \\ &\dots \\ y_m &= h_{m1}x_1 + h_{m2}x_2 + \dots + h_{mp}x_p \end{aligned}$$

と表される。 y_1, y_2, \dots がそれぞれ第1主成分、第2主成分・・・である。各主成分の分散は固有値と呼ばれ、第1主成分、第2主成分・・・の順に大きい。主成分 y_j の係数の列 $h_{j1}, h_{j2}, \dots, h_{jp}$ は固有ベクトルと呼ばれる。固有値の値が大きいほど、その主成分はもとのデータを説明する力が大きい。

砂時計型のニューラルネットワークの場合、中間層のユニットの個数が入力層・出力層より少ないため、入力パターンはより低い次元に投影されることになる。このとき、入力層から中間層への結合は固有ベクトルに相当する働きをしている^⑧。中間層から出力層への結合はこれと逆の働きをしており、もとの空間に入力パターンを復元する。シミュレーション2でこのようなネットワークを用いるのは、脳内には形を再現するのに十分な情報が表現されていると考えられるからである。人間は形を弁別するだけでなく、想起するこ

ともできる。これと同様に、ネットワークも形を再構成することができるような情報を、その内部に保持している必要があるのではないだろうか。

恒等変換を行うネットワークの出力は入力パターンそのものなので、出力と実験データとを比較することはできない。ここでの目的は、ネットワークの内部表現の性質を調べ、それを心理実験から推測される人間の脳内表現と比較することである。

2.1 26 個のアルファベット文字に関するシミュレーション

ネットワーク 入力パターンにはシミュレーション1と同じものを使用した。ここでは恒等変換を学習させたため、教師パターンは入力パターンと同一である。その結果、ネットワークの入力層と出力層は共に 35 個ずつのユニットから構成されることになった。中間層のユニットの個数は 16 個であった。学習手続きやパラメータはシミュレーション1と同一で、学習回数は 20 万回であった。

結果と考察 シミュレーション1と同様な方法で中間層のユニットの活動パターンを分析した。MDSの結果を図 1.14 に示す。採用した解は 5 次元解で、ストレスの値は .05 であった。シミュレーション1の分類学習の場合とは異なり、ネットワークから得られた布置は心理実験から得られた布置とよく似ている。第一次元は文字の幅に、第二次元は文字の丸みに対応しており、これは Gilmore et al. (1979) のデータから導出された布置と同様である。第三次元に関しても、両者の布置はよく似ている。軸ごとの順位相関も、表 1.2 に示すように高い。この結果は、ネットワークが人間の場合と類似した内部表現を獲得したことを示唆している。

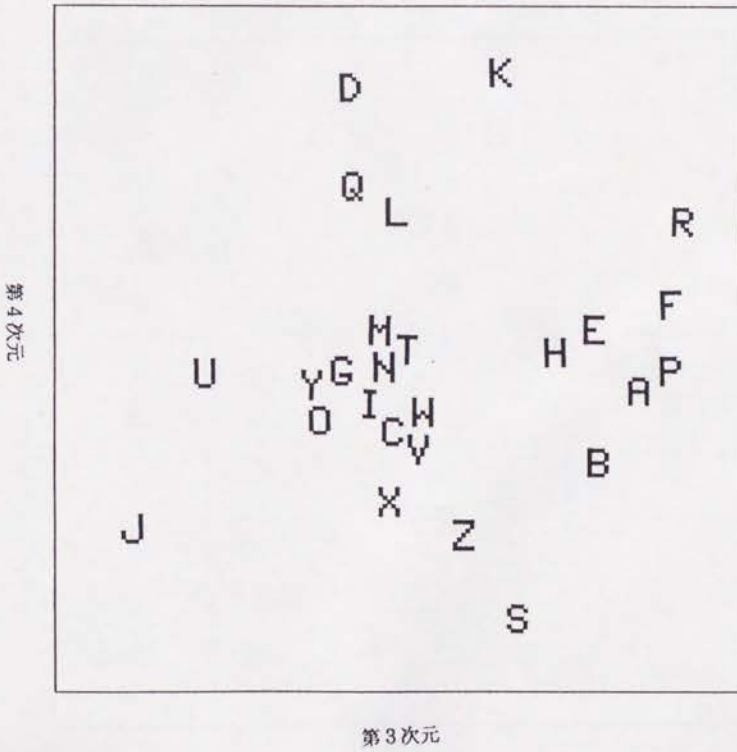
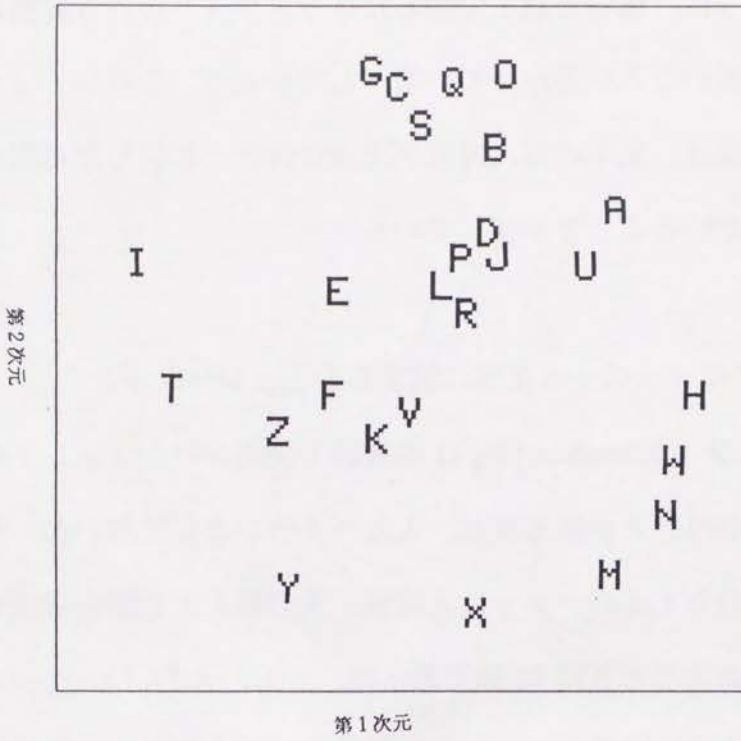


図 1.14 Gilmore et al. (1979)のフォントの恒等変換を学習したニューラルネットワークの内部表現から得られたアルファベット文字の布置. (上) 第1・第2次元. (下) 第3・第4次元.

前述のように、恒等変換を学習したネットワークが行う変換は主成分分析に近いことが分かっている。したがって、このシミュレーションの結果は、文字の脳内表現が主成分分析と類似した過程によって生成されていることを示唆している。

2.2 6個のアルファベット文字に関するシミュレーション

ネットワーク Rumelhart (1971) の実験で使用されたフォントを 35次元の 2 値のベクトルに変換し、入力パターンとして用いた。そのため、前回のシミュレーションと同様、入力層と出力層を構成するユニットの個数はともに 35 個であった。一方、入力パターンの数が少ないために、中間層のユニットの個数は 3 個となった。学習手

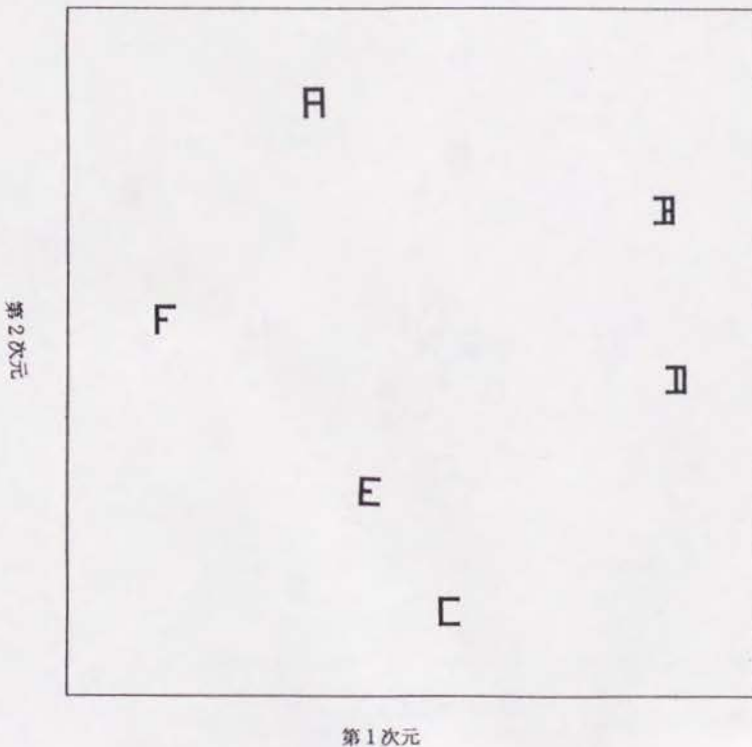


図 1.15 Rumelhart (1971) のフォントの恒等変換を学習したニューラルネットワークの内部表現から得られたアルファベット文字の布置。

続きやパラメータは前回のシミュレーションと同一であり、学習回数は 20 万回であった。

結果と考察 MDSの結果を図 1.15 に示す。2次元解が採用され、ストレスの値は .0031であった。ネットワークから導出された布置は Rumelhart (1971) のデータから導出されたものとよく似ている。また、表 1.2 に示されるように、軸ごとの順位相関も高い。Gilmore et al. (1979) の場合と比べて刺激の種類が少なく、異なったフォントを使用している場合においても、ネットワークの内部表現は心理実験から推測される人間の脳内表現と類似していることが分かった。

2.3 ランダムドット・パターンに関するシミュレーション

ネットワーク Blough (1985) の使用したランダムドット・パターンにもとづいて入力パターンを作成した。このランダムドット・パターンは 8×7 の升目上で定義されているが、ここでは各パターンの重心を同じ位置にそろえるために、 9×9 の升目上に移し変えた。その結果、各ランダムドット・パターンは 81次元の 2 値ベクトルに変換された。入力層と出力層を構成するユニットの個数はともに 81 個で、中間層のユニットの個数は 10 個であった。学習手続きやパラメータは前回のシミュレーションと同一であり、学習回数は 20 万回であった。

結果と考察 MDSの結果を図 1.16 に示す。心理実験の分析と同じく 4次元解が採用され、ストレスの値は .06 であった。図に示されるように、ネットワークから導出された布置は Blough (1985) のデータから導出されたものと異なっている。さらに、軸ごとの順位相関も低い(表 1.2)。

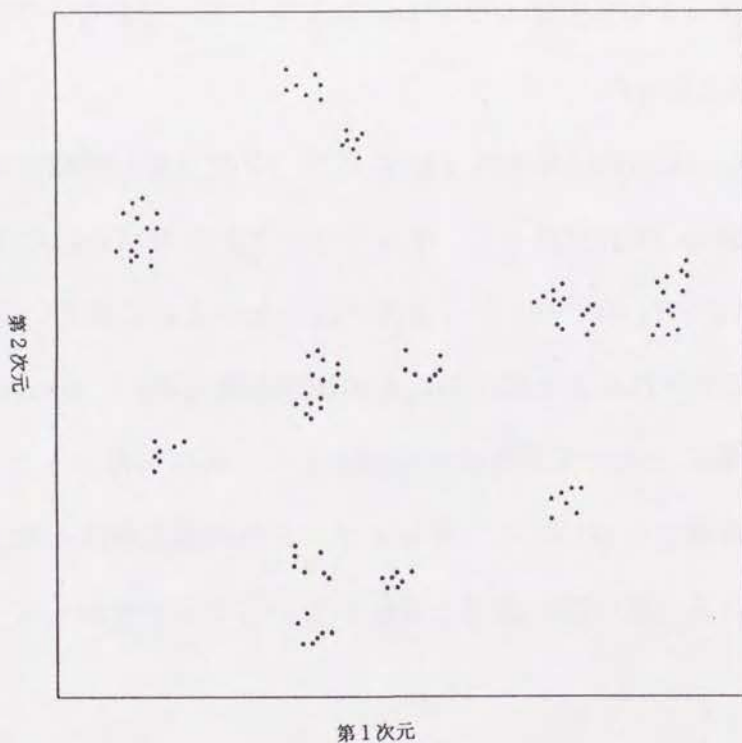


図 1.16 Blough (1985) のランダムドット・パターンの恒等変換を学習したニューラルネットワークの内部表現から得られた布置.

アルファベット文字の場合とは異なり、ネットワークから得られた布置が心理実験から得られた布置と類似しなかったのはなぜだろうか。これまでに述べたシミュレーションで用いた入力表現は、アルファベット文字やランダムドット・パターンを構成する点をそのまま2値化したものであった。このような入力パターンの中には、どの点とどの点が隣り合っているのかに関する情報が含まれておらず、点同士の2次元的位置関係に関する情報が失われてしまっている。その結果、人間とは異なる内部表現を獲得することになったのではないだろうか。

それでは、アルファベット文字に関するシミュレーションでは心理実験の場合と類似した布置が得られたのはなぜだろうか。アルファ

ベットの場Ⓐ、ある文字を構成する線分は他の文字と共通する位置を占めていることが多い。たとえば、"E" と "F" に含まれる垂直な線分は共に左端に位置し、同一の点から構成されている。その他にもこれと同じ線分を含む文字が多数存在するため、ネットワークが学習を行っていたとき、これらの点に対応する入力ユニットの活動は同じように変動する確率が高かったはずである。その結果、点同士の2次元的位置関係が明示的に表現されていなくても、ネットワークはそれらの共変関係を学習することができたのではないかと考えられる。これに対して、ランダムドット・パターンの場合には点の密度が低いため、あるパターンを構成する点は他のパターンとほとんど重複していない。そのため、ネットワークは点同士の関係を学習することができなかつたのではないかと考えられる。ランダムドット・パターンに関しても人間の場合と同様な内部表現を獲得させるためには、入力表現を改良する必要がある。

3. シミュレーション3：複雑型細胞を用いた入力表現

シミュレーション3では、一次視覚野のハイパーコラムに類似した入力表現を使用した。一次視覚野(V1)においては、各方位の線分に対して選択的に反応する細胞が、1つの柱状の構造物の中に整然と並んでいる(Hubel & Wiesel, 1977)。このような構造物はハイパーコラムと呼ばれている。一次視覚野には数千個のハイパーコラムが存在し、それぞれのハイパーコラムが視野の一部分に関する情報を表現していると考えられている。ハイパーコラムを構成する線分検出器には、単純型細胞と複雑型細胞がある。これらの細胞は特定の方位を持つ線分に対して選択的に反応するので、点同士の2次元的

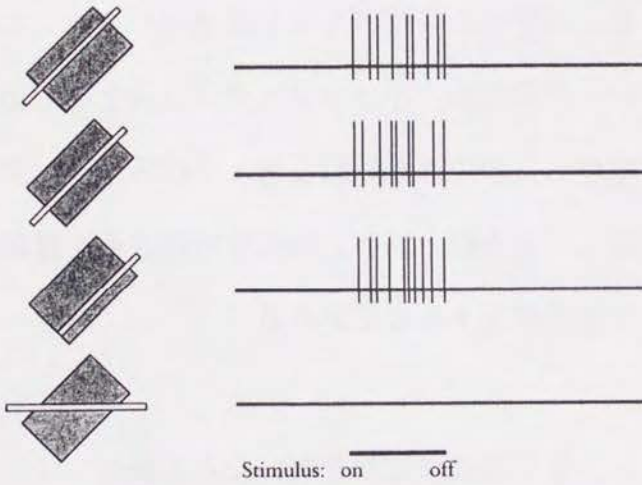
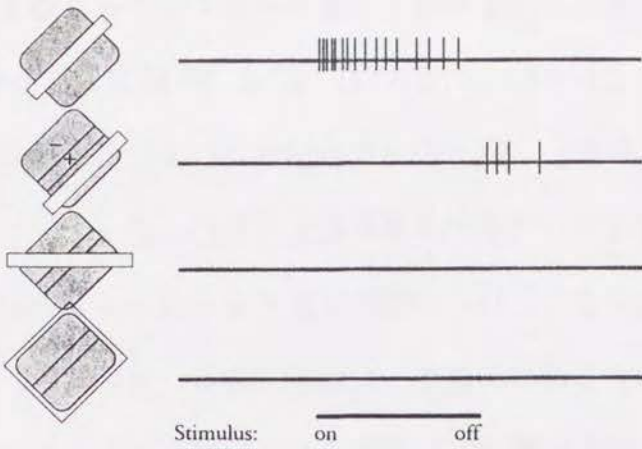


図 1.17 (上) 単純型細胞の反応の仕方. 図の右側の櫛状に並んだ垂直な線分は, 細胞の反応の度合いを示す. 適切な位置に適切な方位をもつ線分が呈示されると, 細胞は反応する. しかし, 位置がずれていたたり, 方位がずれていたたり, 刺激が大きすぎたりすると反応しない. (下) 複雑型細胞の反応の仕方. 複雑型細胞は, 線分が適切な方位をもっていれば, 位置がずれても同じように反応する. (図はHubel(1988)より抜粋.)

な位置関係に関する情報を抽出することができる.

図 1.17は, 単純型細胞と複雑型細胞の反応の仕方を示している. 単純型細胞(上)は線分の方位だけでなく, 位置に対しても選択性をもっている. すなわち, 線分の位置が受容野の中心部分をはずれ

ると、適切な方位をもっていても単純型細胞は反応を示さなくなる。これに対して複雑型細胞（下）は、一定の範囲内であれば、線分の位置が変化しても同じような反応を示す。

予備的な研究においては、単純型細胞に類似した線分検出器を用いて、ランダムドット・パターンに関するシミュレーションを行った。その結果、ネットワークと心理データから得られた布置同士の相関は増大したが、アルファベット文字の場合と比較すると低い値にとどまっていた。このような方法では、ランダムドット・パターン同士が共有する特徴を十分に抽出することができないようである。これはなぜであろうか。たとえば、図 1.18 に示すような 2 つのランダムドット・パターンは、左端に縦に並んだ 2 つの点を含んでいる。人間の眼には、これらは共に垂直な線分を構成するように見える。しかし、これらの「仮想的な線分」は異なった位置に存在するため、単純型細胞は異なった反応をすることになる。2 つのパターンが共有する特徴を抽出するためには、位置のずれをある程度許容するような入力表現が必要である。そこで、シミュレーション 3 では複雑型細胞に類似した特徴検出器を用いて入力パターンを作成した。

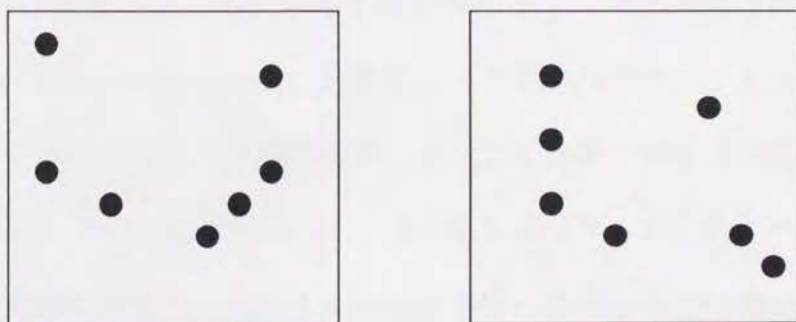


図 1.18 Blough (1985) の実験で使用された 2 つのランダムドット・パターン。

3.1 ランダムドット・パターンに関するシミュレーション

ネットワーク 以下のような手続きで入力パターンを作成した。縦3×横3個の「ハイパーコラム」を、シミュレーション2で使用した入力パターンをちょうど覆うように配置した。個々のハイパーコラムは8種類の線分検出器から構成されており、それぞれの検出器は0, 22.5, 45, 67.5, 90, 112.5, 135, 157.5度の傾きをもつ線分に対して最も良く反応した。これらの線分検出器は、刺激の位置のずれを3×3ドットの範囲内で許容するように作られていた。ハイパーコラムの受容野の中にこれらの線分検出器が反応する特徴が1つでも存在した場合には、コラム内のその特徴に対応する入力パターンの値を1に、存在しない場合は0に定めた（補遺2参照）。

以上のような手続きで作成した入力パターンの恒等変換を、ネットワークに学習させた。入力層と出力層のユニットの個数はそれぞれ72個（ハイパーコラムの個数×特徴の種類）であった。また、中間層のユニットの個数は16個であった。学習手続きやパラメータは前回のシミュレーションと同一で、学習回数は20万回であった。

結果と考察 MDSの結果を図1.19に示す。心理実験の分析と同じく4次元解が採用され、ストレスの値は.089であった。予想したとおり、ネットワークから導出された布置は、Blough (1985)のデータから導出されたものと類似していた。相関係数も、シミュレーション2と比べて高くなっている（表1.2）。この結果は、シミュレーション2においてネットワークがランダムドット・パターンに関する適切な内部表現を獲得できなかったことの原因が入力表現にあるという考えを裏付けるものである。

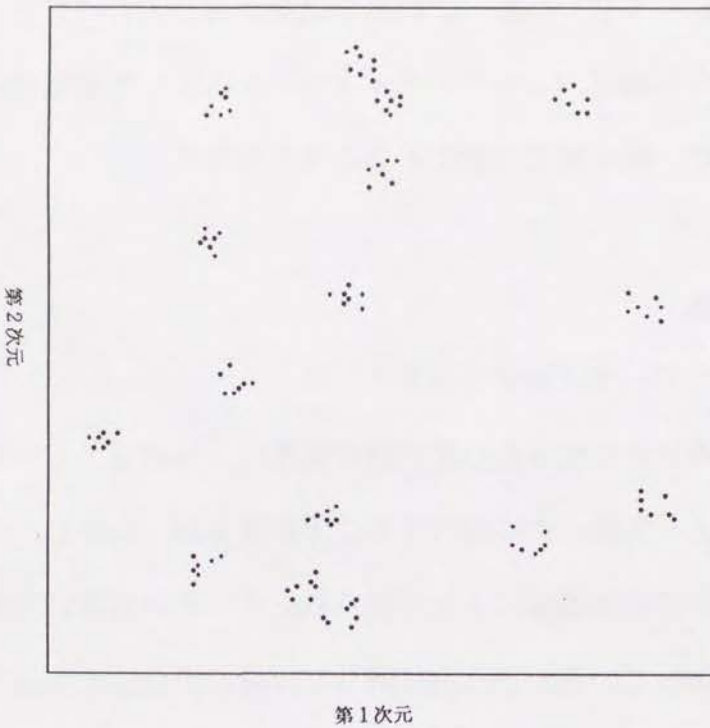


図 1.19 Blough (1985) のランダムドット・パターンの恒等変換を学習したニューラルネットワークの内部表現から得られた布置. 入力パターンは複雑型細胞に類似した特徴検出器から構成される「ハイパーコラム」によって生成された.

4. シミュレーションのまとめ

アルファベット文字の分類を学習したニューラルネットワークの内部表現は、心理実験の結果から推測される人間の脳内表現とは似ていなかった（シミュレーション1）。しかし、アルファベット文字の恒等変換を学習したネットワークは、心理実験の結果から推測されるものと類似した内部表現を獲得することができた（シミュレーション2）。ところが、ランダムドット・パターンを入力パターンとして用いた場合には、ネットワークの内部表現は心理実験の結果から推測される脳内表現とは似ていなかった（シミュレーション2）。この矛盾は、入力表現を改良することで解消することができ

た（シミュレーション3）。これらの結果から、入力パターンの恒等変換を行う単純なニューラルネットワークによって複数の心理実験の結果を統一的に説明可能であると考えられる。

第4節 考察

1. 心理実験のデータに関する再考

第2節で再分析を行った心理実験の結果は、単純なニューラルネットワークによって統一的に説明することができた。しかし、瞬間呈示された文字の同定課題によって得られたデータが通常の文字認知の過程を反映しているかどうかに関しては議論の余地がある。

Gilmore et al. (1979) の実験においては、平均正答率が50%程度になるように刺激の呈示時間が調節された。そのため、被験者は非常に困難な状況下で課題を行ったことになる。そのような状況で用いられた刺激の属性は、通常の文字認知で用いられるものとは異なっている可能性がある。Rumelhart (1971) の実験に関しても同様なことがいえる。

しかし、Blough (1985) の実験では、比較的自然な状況下でランダムドット・パターン間の類似度が測定されている。被験者の課題は一对のランダムドット・パターン間の類似度を評定することであった。この実験では被験者が反応するまで刺激が呈示され続けていたため、被験者には刺激の認知のための時間が十分にあったはずである。したがって、評定の際に使用された刺激の属性は通常の認知過程で用いられるものとそれほど異なっていなかったのではないかと考えられる。シミュレーション3で用いたネットワークがBlough (1985) の結果を説明できたことは、本論における説明の枠組みが瞬間呈示

条件に限定されたものではないことを意味している。

この見方を支持するような心理実験のデータは、他にも得られている。森崎と乾(1995)は類似度評定課題や並べ換え課題を用いて顔写真の間の類似度を測定した。さらに、彼女らは顔パターンの恒等変換をニューラルネットワークに学習させた。本論におけるシミュレーションと同様、ネットワークの獲得した内部表現は心理実験の分析から推測される脳内表現と類似していた。この実験で用いられた課題も瞬間呈示を用いたものではなく、被験者は通常の認知過程で使用されるのと同様な刺激の属性を用いて課題を遂行していたのではないかと思われる。

2. 主成分分析との関係

前述のように、恒等変換を学習したネットワークが行う変換は主成分分析と類似した性質をもっている。このネットワークは、入力パターンを中間層のより低い次元の空間に投影し、出力層において再構成する。ネットワークの内部表現から導出された布置が心理実験の分析から得られたものと類似していたことから、文字やランダムドット・パターンのような2次元パターンの脳内表現は、主成分分析と同様な過程によって生成されているのではないかと考えられる。この考えを検証するため、入力パターンそのものに対して主成分分析を行った。シミュレーション1・2で用いた入力パターン

(Gilmore et al. (1979)の使用したフォントから作成されたもの)に対して主成分分析を行った結果、第1主成分と第2主成分から形成される布置は、心理実験から得られた布置(図1.3)及びネットワークから得られた布置(図1.14)と類似していた。これは上の考えを

裏付ける結果である。

主成分分析は2つの優れた特性をもっている。第一に、主成分分析はもとのパターンがもつ情報を最大限に保ちながら、入力パターンをより低い次元の空間に投影する。このため、もとのパターンを再構成するために必要なデータの大きさが最小化される。第二に、投影後の空間において分散を最大化するように軸の方向を決定するため、弁別能力が最大化される。脳が主成分分析と同様な性質をもつ過程を実現させているのではないかという考え自体は比較的古くからあり、Barlowが1960年代に既に指摘している (Barlow, 1995)。

ところで、ネットワークの内部においては、中間層の個々のユニットがそのまま主成分に対応しているわけではない。Cottrell et al.

(1988) は、入力パターンの恒等変換を学習したネットワークの内部表現を分析し、中間層の個々のユニットが出力パターンの分散をほぼ均等に説明していることを見出した。もしネットワークが主成分分析と全く同一の変換を行っているのであれば、この分散は主成分分析における固有値のように、異なった大きさをもつはずである。彼らは、このような違いはネットワークを構成するユニットのダイナミックレンジが固定されているために生じると説明している。主成分は個々のユニットに対応するのではなく、複数のユニットの活動パターンに対応していると考えられる。

3. 学習規則について

今回のシミュレーションで用いたバックプロパゲーション法は、教師あり学習規則の一種であり、ネットワークに対して外部から教師パターンを与えてやらなければならない。心理実験との比較を行

う際には、この教師パターンの中に過剰な情報が含まれていないかどうかについて注意する必要がある。教師パターンが心理実験の結果に関する情報を何らかの形で含んでいる場合には、シミュレーションの妥当性が失われてしまうからである。シミュレーション1における教師パターンは、活性化すべきユニットの値を1と定め、その他のユニットの値を0と定めたものであった。このパターンは正しい出力ユニットを指定するだけの情報しかもっておらず、人間が文字を学ぶ際に与えられる情報と等価であると考えられる。一方、シミュレーション2と3においては、教師パターンは入力パターンと同一であった。この場合も、教師パターンは余分な情報を含んでいない。

一方、バックプロパゲーション法は誤差を計算する際にユニット間の非局所的な相互作用を必要とするため、神経生理学的な見地からすると妥当性が低いと言われている。そこで、局所的な相互作用のみを用いて主成分分析と同様な変換を学習する方法も提案されている (e.g. Oja, 1989)。脳のモデルとしては、このような規則を用いた方が妥当であろう。しかし、本研究の目的はパターン認知のモデルをつくることではなく、文字の内部表現の性質を探ることである。そのため、ネットワーク自体の妥当性ではなく、ネットワークが行う変換そのものが問題となる。先に述べたように恒等変換を行うネットワークは主成分分析に近い変換を行うことが明らかにされており、学習規則そのものはあまり問題ではないと考えられる。

4. 神経生理学的な研究との関係

本研究の結果は、刺激の類似度空間の構造が大局的な特徴にもと

づいているという点で、神経生理学における知見と一致している。下部側頭葉にみられるような大局的な特徴に反応する細胞が、どのようにその受容野を形成していくのかはまだ分かっていない。本研究の結果は、それが主成分分析に類似した過程によって行われていることを示唆している。

一方、Hancock, Baddeley & Smith (1992) は、自然画像に対して主成分分析を行うことによって得られる主成分が、一次視覚野の単純型細胞の受容野に類似していることを示した。これは本研究の結果とどのような関係にあるのだろうか。シミュレーション3で用いた入力パターンは、一次視覚野の複雑型細胞に類似した特徴検出器の出力にもとづいて作成された。したがって、この入力パターンは一次視覚野からの出力に相当すると考えられる。一方、シミュレーション1と2で用いた入力パターンは、心理実験で使用されたフォントを単純に2値化したものであった。しかし、この入力パターンにおいては、個々のパターンの大きさや位置があらかじめ統一されていた。実際の視覚系の中では、大きさや位置が一定であるような入力表現は、それらの変化に対する不変性を実現するような過程を経なければ実現できないはずである。したがって、シミュレーション1・2の入力パターンも一次視覚野以降の部位に対応すると考えられる。以上のような議論から、本研究で用いたネットワークの中間層は一次視覚野より高次の視覚野に対応していると考えられる。この見方は Hancock et al. (1992) の結果と矛盾しない。一次視覚野から高次視覚野までの情報処理過程は、主成分分析という観点から統一的に説明できるかもしれない。

最近、Miyashita, Date, & Okuno (1993) は、サルの下部側頭葉にお

ける神経細胞の刺激選択性が、記憶課題を行うことによって変化する
ことを見出した。これらの神経細胞の受容野は、繰り返し課題を
行うことによって、実験で使用された刺激の形に近づく方向に変化
した。これは、形の脳内表現が学習によって動的に変化することを
示している。この結果も、ネットワークが入力パターンを学習する
ことによって形の内部表現が形成されるという本研究の枠組みと一
致している。ネットワークの学習過程を人間のそれと比較すること
によって、形の内部表現に関してより深い理解を得ることができる
のではないかと考えられる。

第2章 単語認知過程のモデル化

第1節 文字認知から単語認知へ

1. 単語認知と文字認知との関係

第1章では、文字の内部表現の性質を主成分分析という枠組みによって理解することができることが示された。単語の内部表現に関してもこれと同じ枠組みを適用することができるだろうか。これを検討するためには、まず単語認知過程の入力表現について議論する必要がある。

第1章のシミュレーション1・2では、文字を構成する点をそのまま2値化した入力表現を用いた。これと同じ方法で単語に関する入力パターンを作成した場合、1個の単語は（文字を構成する点の数）×（単語を構成する文字の数）の次元をもつベクトルによって表現されることになる。このような入力表現は、同一の文字であっても位置が異なれば全く違うパターンによって表現されるという問題点をもっている。たとえば、“ON”と“NO”は同じ文字によって構成されているが、左側の“O”と右側の“O”は異なったユニットによって表現されることになる（図2.1）。

この場合、“ON”と“NO”の類似性の度合いは、それぞれの位置で文字が重なり合うドットの数、すなわち“O”と“N”が共有するドットの数に依存することになってしまう。その結果、“ON”と“NO”の類似度は非常に低くなる。人間の内部表現のモデルとしては、これは明らかに不適切であろう。

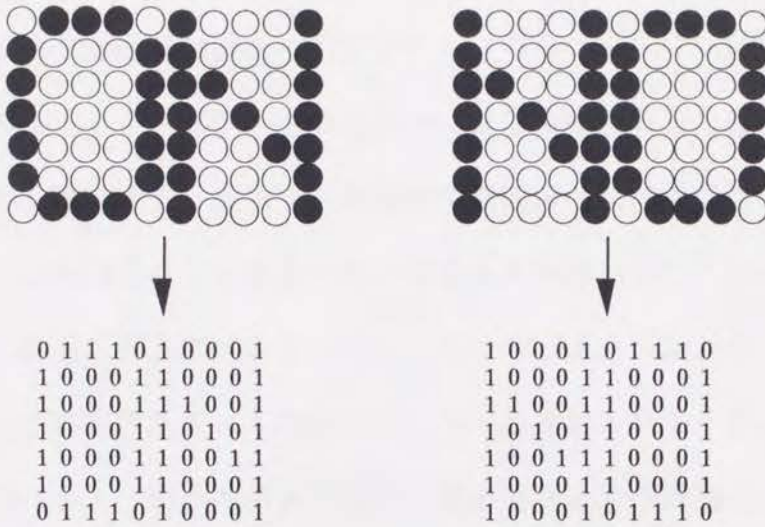


図2.1 "ON" と "NO" を構成するドットを2値化したもの。

したがって、適切な単語の内部表現を生成するためには、文字の位置の変化に対してある程度の不変性をもつような入力表現を用いる必要がある。ところが、単語同士を弁別するためには、文字の位置に関する情報を完全に失ってしまってはならない。単語を構成する文字の位置が異なれば、同じ文字から構成されていても異なった単語として認知しなければならないからである。"ON" と "NO" を弁別するためには、2つの文字の位置に関する情報を何らかの形で残しておかなければならない。

以上のような考察から、単語の内部表現について検討するためには、入力パターンの中で文字の位置をどのように表現するのかという問題を先に解決しなければならないことが分かる。第一に、文字の位置が変化しても、それがどの文字であるのかに関する情報を保存するような入力表現を考える必要がある。第二に、単語の意味はそれを構成する文字の順序によって決まるので、文字同士の相対的

位置関係も入力空間に反映されていなければならない。

それでは、シミュレーション3で使用した複雑型細胞に類似した特徴検出器による入力表現を用いた場合はどうであろうか。この種の特徴検出器は、特徴の位置がある範囲内で変化しても、反応の仕方を変えない。この範囲の大きさが1個の文字の大きさに対して十分小さい場合には、シミュレーション1・2の入力表現と同様、位置の異なる同一の文字は異なったパターンによって表現されることになる。この範囲の大きさが複数の文字にまたがるほど大きい場合には、異なる位置に存在する同一の文字に対する反応は類似することになるが、今度はどの特徴がどの位置の文字に含まれているのかに関する情報が失われることになる。その結果、正確な単語認知が不可能になってしまう。後述するように、このような矛盾を解決するためには特徴の位置のずれを徐々に許容していくような階層的ネットワークを用いる必要がある (Fukushima, 1975; Mozer, 1991)。

以下の部分では、単語認知過程に関する生理学・心理学的な知見について概説し、これまでに提案されたモデルの妥当性について検討する。単語の内部表現と主成分分析との関係に関しては、第4章であらためて取り上げることにする。

2. 失読と単語認知

ここでは、脳損傷患者の症例から得られた知見について概説する。単語の認知と関係の深い脳損傷として、純粹失読 (pure alexia) がある。純粹失読の患者は文章を読む能力が著しく損なわれているが、その他の視覚的認知能力は比較的保たれている。また、書字や会話の理解などの、言語に関連する能力も損なわれていない。これらの患者

が文章を読む際には、単語を一文字ずつ音読することが多い。音読をしない場合であっても、1個の単語を読むのに要する時間は単語の長さに比例して増大する (Bub, Black, & Howell, 1990)。したがって、これらの患者は主として単語全体を同時に読む能力が損なわれているのではないかと考えられる。

このような症例は、単語の認知を行うために特殊化した部位が脳内に存在することを意味しているのだろうか。前述のように、書き言葉は人類の歴史全体から見れば比較的最近に発明されたものであり、現在でも書き言葉を用いないで生活している人々が多い。したがって、単語認知を行うための生得的な機構が進化によって形成されたとは考えにくい。純粹失読の患者は、脳内の特定の部位が損傷を受けることによって、単語認知に必要な何らかの過程が損なわれているのではないかと考えられる。

純粹失読は、視覚的失認 (visual agnosia) の一種である。視覚的失認とは、光点の検出のような比較的低次の知覚に関する能力はほぼ正常であるが、対象の認知に関しては重度の障害を示すような症例を指す。視覚的失認には様々な種類があり、研究者によってまちまちな分類が行われてきた。Farah (1990) は多数の症例を系統的に分析することによって、図 2.2 のように視覚的失認を分類した。彼女の分類の枠組みを、以下に説明する。

視覚的失認は、大きく分けて統覚失認 (apperceptive agnosia) と連想失認 (associative agnosia) に分類することができる。前者は比較的低次の認知過程の障害であり、後者は比較的高次の過程の障害であると考えられている。統覚失認には、狭義の統覚失認・腹側性同時失認 (ventral simultagnosia) ・背側性同時失認 (dorsal simultagnosia) が含

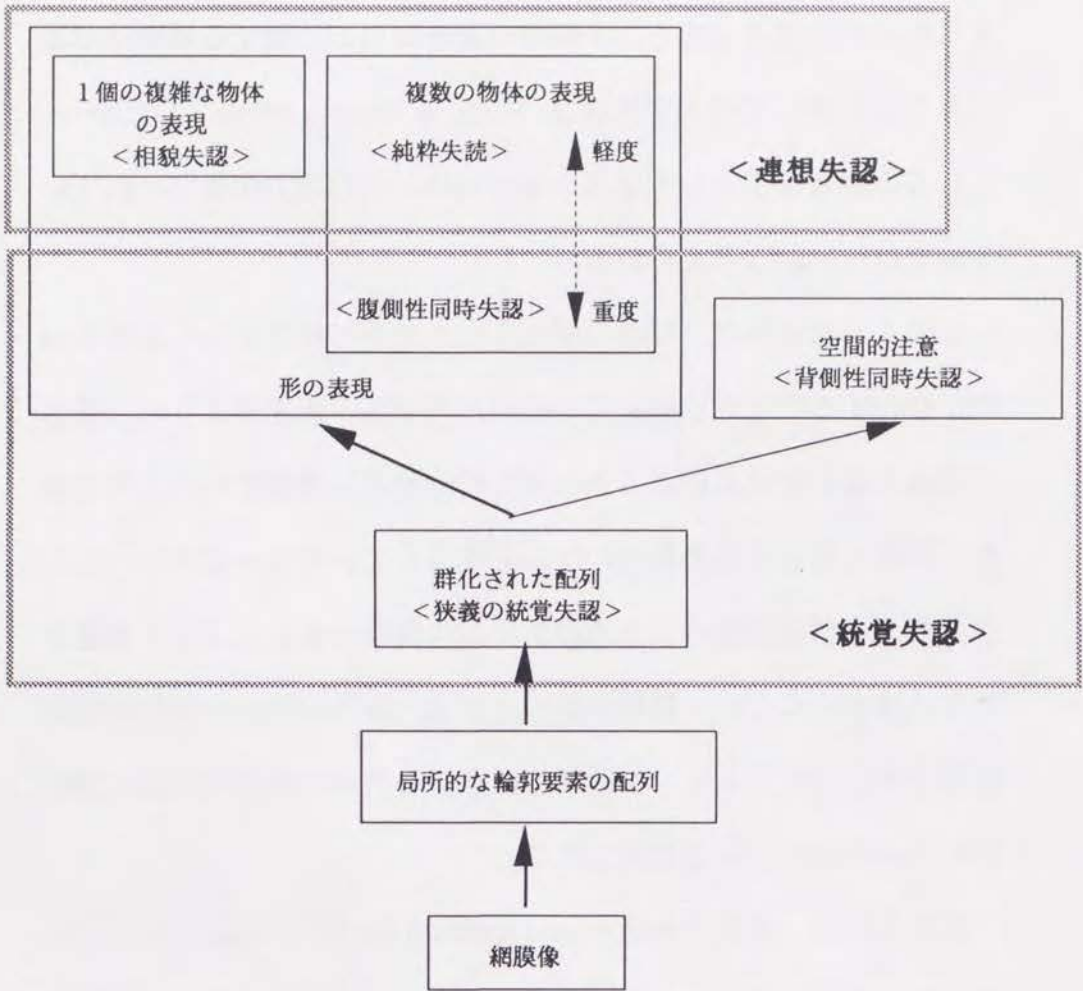


図 2.2 Farah (1990) による視覚的失認の分類.

まれる。狭義の統覚失認は、局所的な輪郭要素を結合させて、群化された配列 (grouped array) を作り出す過程が損なわれることによる障害であると考えられている。典型的な患者は後頭葉に広範な損傷がみられ、視野・視力・光の波長の弁別等は正常であるが、対象の認知がほとんどできない。背側性同時失認は、空間的注意を司るシステムが損傷を受けたために、一定時間内に注意を向けることができる量が限定されることによる障害であると考えられている。典型的な患者は後頭葉から頭頂葉にかけての両側性の損傷がみられ、1

個の対象の認知を行うことはできるが、複数の対象を同時に認知することができない。一方、腹側性同時失認は、対象の形を処理するシステムが損傷を受けたために、一定時間内に認知することができる量が限定されることによる障害であると考えられている。典型的な患者には後頭葉から左側頭葉下部にかけての損傷がみられる。患者の症状は背側性同時失認の場合と似ており、1個の対象の認知を行うことはできるが、複数の対象を同時に認知することはできない。しかし、複数の点の個数を数えることができる・対象を操作できる・障害物にぶつからずに歩き回ることができる・十分に時間をかければ複数の対象を認知することもできる、といった点で背側性同時失認と異なっている。

連想失認は知覚的な過程の障害ではなく言語系との結合の障害であると考えられることが多かったが、Farah (1990) はこれも対象の形の分析過程における障害としてとらえた方が適切であることを示した。彼女の枠組みにおいては、連想失認は従来考えられていたような視覚と他の感覚との間の連想に関する障害ではなく、視覚的認知過程の比較的高次な部分における障害とみなされる⁽⁹⁾。図 2.2 に示されるように、対象の形の内部表現は、複雑な対象を1個ずつ処理するサブシステムと、複数の対象を同時に処理するサブシステムの2つに分けられる。彼女は、純粹失読の原因が複数の対象を処理するサブシステムが損なわれることにあると主張している。一方、複雑な対象を処理するサブシステムが損傷を受けた場合には、相貌失認 (prosopagnosia) が生じるとされる。

Farah (1990) は、患者を純粹失読と診断する際に用いられてきた課題について検討を加えた。その結果、たとえば Warrington & Shallice

(1980) の用いた課題の多くは単純な情景についての回答を求めるものであり、物体間の関係を把握することが必要な課題はごく一部しかなかったことが分かった。このような課題は同時失認の症状を見出すのには不適切であると考えられる。また、Friedman & Alexander (1984) は純粹失読の患者に対して言語的な刺激と非言語的な刺激を瞬間呈示し、成績を比較した。その結果、線分の向きの弁別や単独呈示された文字の再認のような単純な課題においては、患者の成績は健常者と変わらなかった。一方、単語や線画の認知のような比較的困難な課題においては、患者の閾値は健常者と比べて上昇した。ところが、言語刺激と非言語刺激の間に差は認められなかった。このような結果は、純粹失読が形の分析過程の障害であることを示している。Farah (1990) は、純粹失読が腹側性同時失認の徴候の1つなのではないかと主張している。腹側性同時失認による障害が比較的軽微であるとき、複数の対象の同時認知を要する課題の中でも特に困難な、単語や文の認知に関する障害が顕著に表出するのではないかと考えられる。

なぜ、複数の対象を同時に認知するために特別なサブシステムが必要とされるのだろうか。複数の対象を同時に処理しようとする際には、同じ種類の対象が複数存在する場合に、それらの対象をどのように表現するのか、という点が問題になる。“LETTER”という単語を認知する場合について考えてみよう。このとき、“LETTER”に含まれる2つの“T”はどのように表現されているのであろうか。“T”の表現は1つしかなく、何らかの方法でそれが2個存在することが表現されているのであろうか。それとも、個々の“T”に対する表現が複数存在するのであろうか。このような問題は「タイプ-トークン

問題 (type-token problem) 」と呼ばれる(e.g. Mozer, 1989). タイプは対象の種類の変現に相当し, トークンは個々の対象の変現に相当する. 複数の対象の処理に特別なサブシステムが必要とされる理由の一つは, このタイプ-トークン問題にあるのではないかと考えられる.

タイプ-トークン問題は, 単語認知過程の入力変現に関する問題と密接に関連している. 異なる位置の文字に対して別々のトークンを割り当てた場合には, それらのトークンが同じ文字を表現していることを示す情報が失われてしまう. 一方, 異なる位置の文字を1個のトークンによって表現する場合には, 文字同士の位置関係に関する情報が失われることになる. 単語認知過程のモデルは, この矛盾を適切に解決するようなメカニズムをもっている必要がある. このような問題を容易に解決できるようなメカニズムを人間があらかじめ備えているからこそ, 文字と単語という表記体系が生まれたのではないかと思われる.

3. 注目すべき心理現象

単語認知過程は認知心理学における主要なトピックの1つであり, これまでに数多くの研究が行われてきた. ここでは, 単語優位効果と文字の混合という2つの心理現象に注目する. これらの現象は単語認知過程のメカニズムと密接に関連していると考えられる.

3.1 単語優位効果

Reicher (1969) は, 単語を構成する文字は非単語 (nonword) の中に含まれる文字と比べて認知が促進されることを見出した (非単語とは, 単語として使用されることのない文字列の総称である). 彼は単語の部分報告課題を用いて実験を行った. まず1個の単語が瞬間

的に呈示され、続いてマスク刺激と共に選択肢が呈示された。この
選択肢は、呈示された単語の中の、特定の位置の文字を答えさせる
ものであった。たとえば、単語 "WORD" が呈示された後には
" $-\text{---}\frac{D}{K}$ " という選択肢が呈示された（この選択肢は被験者に右端の
文字を選択するように要求している）。"D" と "K" の2つの文字は、
呈示された単語の選択肢以外の3文字と組み合わせたときに、使用
頻度のほぼ等しい単語（"WORD" と "WORK"）を構成する。これは、
被験者が他の3つの文字からの推測によって反応することが結果に
与える影響を排除するためである。実験の結果、単語刺激を用いた
場合の方が、"ORWD" のような非単語刺激の場合より正答率が高い
ことが分かった。このような現象は単語優位効果 (word superiority
effect) と呼ばれている。単語に関する知識は、文字の認知を促進す
る役割を果たしていると考えられる。

さらに、発音可能な非単語に含まれる文字は、ランダムに生成さ
れた非単語に含まれる文字と比べて認知が促進されることも知られ
ている (e.g. Aderman & Smith, 1971)。発音可能な非単語は疑似単語
(pseudoword) と呼ばれ、英語の正書法的な規則に従って文字が配置
されている。たとえば "MAVE" のような疑似単語の中に含まれる文
字は、"XPQJ" のようなランダムな非単語に含まれる文字と比べて認
知が促進される。

3.2 文字の混合エラー

複数の文字を瞬間呈示した場合には、単語間で文字が入り交じっ
て見えることがある (Allport, 1977)。このような現象は「文字の混合
エラー (letter migration error)」と呼ばれる。たとえば、"SAND" と
"LANE" という2つの単語を同時に瞬間呈示し、一方の単語を被験

者に報告させると、"SANE" または "LAND" というエラーがかなりの頻度で見られる。

文字の混合エラーの頻度は、同時に呈示する2つの単語間の距離を大きくすると減少する (Mozer, 1983)。これは、文字の混合エラーが2つの単語間における何らかの相互作用に起因することを示している。被験者がターゲット単語全体を知覚できなかったときに、識別できた文字から残りの部分を推測することによって文字の混合が生じるのであれば、このような効果はみられないはずである。さらに、Treisman & Souther (1986) は、探索課題を用いた場合にも文字の混合が生じることを確かめた。被験者に与えられた課題は、瞬間呈示された4個の単語（プローブ）⁽⁴⁰⁾の中にターゲット単語があったかどうかを答えることであった。ターゲットが実際にはプローブの中に含まれていなかったときに被験者が「あった」と答えた場合

（フォールス・アラーム事態）が、文字の混合エラーとみなされる。ターゲットには3文字の単語が使用された。結合条件のプローブには、この3文字がすべて含まれていた（図2.3左）。これに対して、コントロール条件のプローブには、3文字のうちの2文字しか含まれていなかった（図2.3右）。

被験者が単語を構成する3文字のうちの2文字あるいは1文字しか認知できなかったときに、認知できた部分から残りの部分を推測することによって文字の混合が生じるのであれば、コントロール条件においても結合条件と同じ頻度でフォールスアラーム反応がみられると予想される。たとえば、結合条件のプローブに含まれる単語 "say" の中の2文字 "ay" からターゲット "day" を推測する確率は、コントロール条件のプローブに含まれる単語 "bus" の中の2文字 "bu"

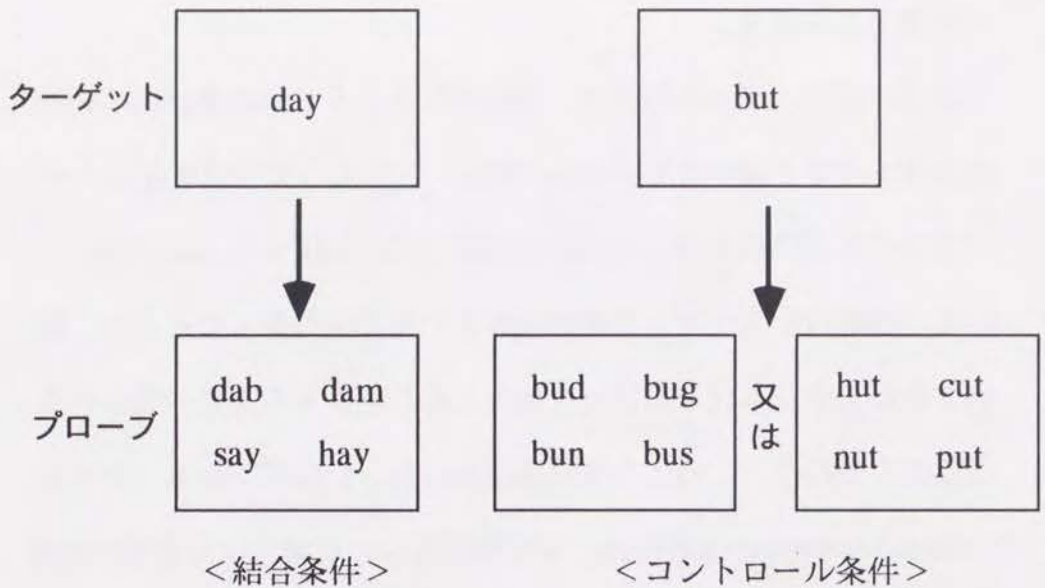


図 2.3 Treisman & Souther (1986) の使用した刺激の例.

からターゲット "but" を推測する確率と等しいはずである。これに対して、他の単語から文字が混入して見えることによって文字の混合が生じるのであれば、結合条件の方が文字の混合エラーの頻度が高くなるはずである。実験の結果、結合条件ではコントロール条件より有意に多いフォールス・アラーム反応がみられた。この結果は、実際に単語間で文字の混合が生じていることを意味する。さらに、探索課題においても文字の混合がみられたことは、この現象が記憶に対する負荷に起因する誤りではないことを示している。

Treisman & Souther (1986) は、特徴統合理論 (Treisman & Gelade, 1980) によって文字の混合を説明することができるのではないかと主張した。特徴統合理論は、視覚的特徴の統合過程に関する一般的な理論である。この理論は、線分の方向・色・運動などの視覚的特徴がそれぞれ独立したモジュールによって処理されると仮定している。対象の認知を行う際には、対象に対して視覚的注意 (visual attention)

が向けられる。注意の働きによって各次元の特徴に関する対象の情報が位置に関する情報と結びつけられ、対象の認知が成立する。これと関連して、Treisman & Schmit (1982) は注意が阻害された場合に視覚的特徴が誤って結合して見えることがあることを発見した。彼女らは、青い”X”・緑の”T”・赤い”O”を含む刺激を瞬間呈示し、注意を妨害するためにこれと同時に数字を呈示した。数字を先に被験者に報告させ、次に文字について報告させると、赤い”X”や緑の”O”などのエラーが頻繁にみられた。この現象は「結合錯誤 (illusory conjunction)」と呼ばれる。Treisman & Schmit (1982) は、注意が阻害されることによって形や色に関する情報が誤った位置情報と結びつき、その結果結合錯誤が生じるのではないかと主張している。

Treisman & Souther (1986) は、文字の混合が結合錯誤と同様な原因で生じると考えた。複数の単語を同時に瞬間呈示した場合には、被験者は刺激に対して十分な注意を向けることができない。その結果、文字位置に関する情報が本来その位置にあった文字以外と結び付いてしまうことによって、文字の混合が生じるのである、と彼女らは主張している。

ところが、文字の混合の原因は注意が阻害されることだけではなく、単語同士の類似度も影響を与えていることが示されている。

Shallice & McGill (1978) は、プローブを構成する単語同士の類似度が高いほど文字の混合が起りやすくなることを見出した。さらに Mozer (1983) はより厳密な手続きでこれを確認した。彼の実験では横に並んだ2個の単語がプローブとして用いられた。被験者の課題は、瞬間呈示されるプローブを見てキューによって指定された方の単語を報告することであった。実験の結果、”CAPE CONE”のよう

な類似した単語から構成されたプローブに対しては、"CANE" のような文字の混合によるエラーが有意に認められた ("CANE" という反応は、"CAPE CONE" の左右どちらを答えさせた場合にも生じ得る)。しかし、"CAPE MONK" のような類似度の低いプローブの場合には "CANE" のようなエラーの比率は有意ではなかった。これらの結果から、文字の混合は個々の文字の位置情報が失われることだけに起因するのではなく、周囲の文字に関する情報の影響を受けていると考えられる。

さらに、文字の混合エラーが単語に関する知識の影響を受けるかどうかを検討されている。Treisman & Souther (1986) は、前述の実験の中で、ターゲットに単語と疑似単語を用い、両者の間で文字の混合エラーの頻度に差が生じるかどうかを調べた。その結果、ターゲットが単語であるかどうかによる有意な効果はみられなかった。ところが、McClelland & Mozer (1986) の実験では、単語と疑似単語の間で有意な差がみられた。彼らの実験では、Mozer (1983) と同様、2個の4文字単語がプローブとして用いられた。たとえば、プローブが "WAKE WOVE" である場合には、文字の混合によって "WAVE" または "WOKE" という単語が生成され得る。これに対して、"WAKE WISE" というプローブの場合には、"WASE" または "WIKE" という疑似単語が生成されることになる。Mozer (1983) と同様な手続きによる実験の結果、前者の方が後者よりも文字の混合の比率が高いことが示された。これは、単語に関する知識が文字の混合に影響を与えることを意味している。さらに、van der Velde, van der Heijden, & Schreuder (1989) は、文字の混合によって生成される単語と意味的に関連した単語を事前に呈示すると、文字の混合が増加

することを示した。この結果も、文字の混合が単語に関する知識の影響を受けているという見方を支持している。van der Velde (1992) は、Treisman & Souther (1986) の実験では3文字の刺激が使われ、しかも単語の一部でもあるような疑似単語 (e.g. "sen") が使用されたために、単語と疑似単語の差が生じにくかったのではないかと主張している。

一方、日本語を刺激として用いた場合も、単語間で文字の混合が生じることが示されている (伊藤, 1989)。さらに、複数の漢字を同時に瞬間視した場合には、漢字を構成するへんやつくりの混合が生じる (Fang & Wu, 1989; 牧岡, 1989)。これらの結果は、文字の混合がアルファベット文字に固有な現象ではないことを示している。

第2節 単語認知過程のモデルの発展

この節では、これまでに提案された単語認知過程のニューラルネットワーク・モデルの中から代表的な3つのモデルを取り上げ、特に入力表現に注目しながら解説する。

1. Interactive Activation Model (IA)

McClelland & Rumelhart (1981) は、Interactive Activation Model (IA) と呼ばれる単語認知のニューラルネットワーク・モデルを提案した。このモデルは、4文字の単語を1度に1個ずつ認知することができる。ネットワークは特徴レベル・文字レベル・単語レベルの3層に分かれている (図 2.4)。ユニットの活動の程度は活性度 (activation) という量で表される。あるユニットの活性度は、それと結合している他のユニットの活性度と結合強度との積和によって決まる。各層のユニットはそれぞれ特徴ユニット・文字ユニット・単語ユニット

と呼ばれる。ユニット間の結合の強度はあらかじめ定められていて、学習は行われない。

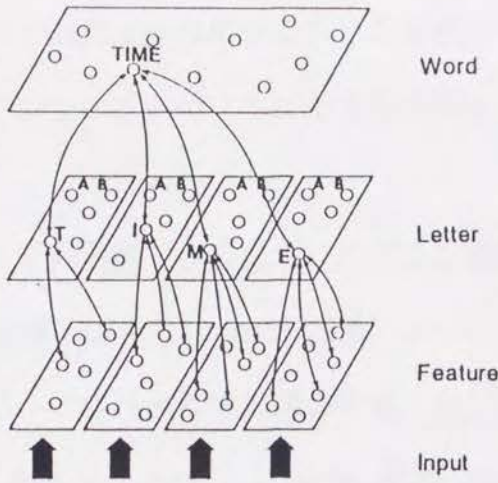


図 2.4 McClelland & Rumelhart (1981) の Interactive Activation Model (IA). 単語 "TIME" に関する結合だけが描かれている (Rumelhart, McClelland & the PDP research group; 1986 より抜粋)。

外部からの入力はず特徴レベルに入る。単語が呈示されると、その単語に含まれる特徴を表現する特徴ユニットが活性化する。たとえば "TIME" という単語が呈示されると、左端の文字に含まれる縦向きの線分に反応する特徴ユニットが活性化する。特徴ユニットは、その特徴を含む文字を表現する文字ユニットとは興奮性の結合で、そうでない文字ユニットとは抑制性の結合で結ばれている。たとえば、文字の中央に位置する縦向きの線分を表現する特徴ユニットは、"T" を表現する文字ユニットと興奮性の結合で結ばれている。これと同じ特徴ユニットは、"D" を表現する文字ユニットとは抑制性の結合で結ばれている。その結果、呈示された単語に含まれる文字を表現する文字ユニットが最も強く活性化することになる。

文字ユニットは、その文字を適切な位置に含むような単語を表現する単語ユニットと興奮性の結合で結ばれ、そうでない単語ユニットとは抑制性の結合で結ばれている。たとえば、左端の”T”を表現する文字ユニットは”TIME”を表現する単語ユニットと興奮性の結合で結ばれている。一方、これと同じ文字ユニットは”DIME”を表現する文字ユニットとは抑制性の結合で結ばれている。さらに、ある文字ユニットと単語ユニットが興奮性の結合で結ばれている場合、単語ユニットから文字ユニットへ向かう興奮性の結合も存在する。これらの結合は正のフィードバック・ループを形成するため、単語ユニットが活性化すると、その単語に含まれる文字を表現する文字ユニットはますます活性化し、その結果単語ユニットの活性度も高まることになる。

単語ユニット同士は相互に抑制性の結合で結ばれている。そのため、活性度が内部で循環するにつれて、最も強く活性化した単語ユニットだけが生き残り、その他の単語ユニットの活性度は0に近づいてゆく。このような競合の過程はIAの出力を明確にする働きがある。IAの最終的な出力は確率的に決定される。ある単語ユニットが出力として選択される確率は、その単語ユニットの活性度を、その他すべての単語ユニットの活性度の和で割ったものである（Luceの公式）。その確率が十分に大きくなったとき、すなわち、入力された単語を表現する単語ユニットが他の単語ユニットと比べて十分に大きな活性度を獲得したとき、IAは単語の認知に成功したとみなされる。

結局のところ、IAを構成する個々のユニットは、自分が表現する特徴（あるいは文字・単語）と整合性をもつユニットを興奮させ、

矛盾するユニットを抑制する、という役割を担っている。個々のユニットが行う処理は単純であるが、多数のユニットが並列に働くことによって単語認知の過程をうまく再現することができる。

IA は単語優位効果を定量的に説明することに成功した。IA が単語優位効果を再現する過程は、次のようなものである。単語が IA に呈示されると、これまでに述べたような過程を経て、その単語を表現する単語ユニットが活性化し始める。すると、その単語に含まれる文字を表現する文字ユニットは単語レベルからの正のフィードバックを受けることになり、活性度を増大させる。これに対して、ランダムな文字列が呈示された場合には、それを表現する単語ユニットが存在しないためにフィードバックが生じない。その結果、文字ユニットの活性度は単語が呈示された場合と比べて小さくなる。このような活性度の差が単語優位効果の原因となる。一方、疑似単語の優位効果は、それらが単語と共通する文字列から構成されていることが原因で生じる。たとえば疑似単語 "MAVE" は、"MA" と "VE" という他の多くの単語と共通する文字列からなる。"MAVE" が呈示されると、"MAKE" や "HAVE" など表現する単語ユニットがある程度活性化し、それらのユニットからのフィードバックによって文字ユニットの活性度が増加する。このような過程によって、疑似単語に含まれる文字の認知が促進される。疑似単語の促進効果は「発音可能であること」ではなく、「単語と共通する文字列を含むこと」によって説明されることになる。とすると、単語と共通する文字列を含んでいれば、子音のみで構成された発音不能な非単語に関して優位効果が生じるのではないかと考えられる。そこで、Rumelhart & McClelland (1982) は、単語の中に出現する頻度が高い子音の文字

列を用いて非単語 (e.g. "SPCT") を作成し、実験を行った。その結果、このような刺激に関する優位効果が実際に存在することが確かめられた。さらに彼らは IA によるシミュレーションを行い、IA がこの実験の結果をうまく再現できることを示した。

このように IA は複数の心理実験の結果をうまく説明することができ、読みのモデルとしては完全ではない。なぜなら、位置の定まった1個の4文字単語を認知することしかできないからである。日常的な読みの事態では、単語の長さ（文字数）や位置は様々に変化し、しかも一度に複数の単語が視野に入る。より現実的な単語認知のモデルは、視野内の文字列の中から任意の長さの単語を抽出し、正しく認知することができなければならない。

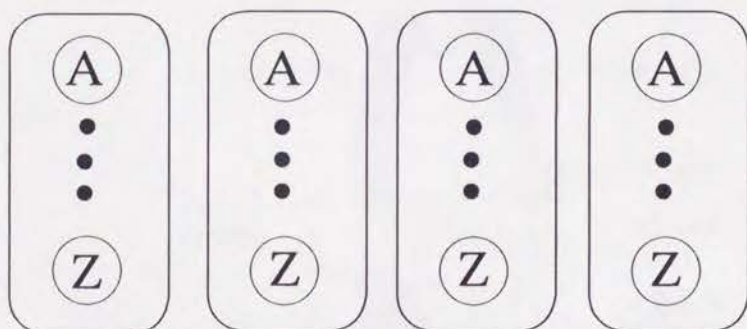


図 2.5 「スロット」型の入力表現。

さらに、入力表現の妥当性という見地からも、IA には問題がある。IA の文字レベルでは、文字は位置ごとに異なったユニットによって表現されている（図 2.5）。この場合、1種類の文字に対して複数のトークンを割り当てることになる。このような入力表現を、位置によって分けられているという意味で「スロット」型の入力表現と呼ぶことにしよう。スロット型の入力表現の問題点は、単語を呈

示す位置がずれると入力パターンが大きく変化することである。たとえば, "CAT" という単語を IA の入力層の左端に呈示した場合と, 右端に呈示した場合とでは, 文字レベルにおける活動パターンはまったく異なったものになる。このような表現を用いても IA が心理実験の結果を正しく予測できるのは, 状況を 1 個の 4 文字の単語の認知に限定しているからである。IA は説明の対象を限定することによって, 入力表現に関する問題を回避しているわけである。

2. Programmable Blackboard Model (PABLO)

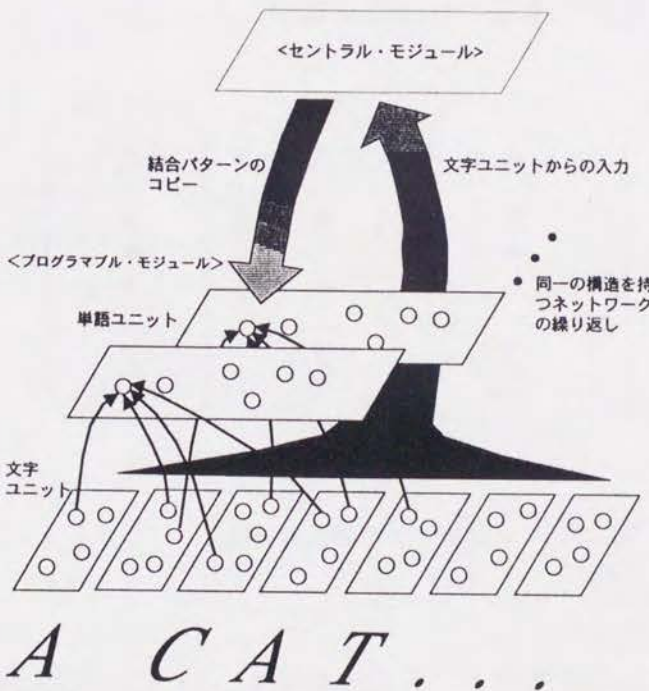


図 2.6 McClelland (1986) の Programmable Blackboard Model (PABLO).

McClelland (1986) は, 様々な長さの単語を同時に複数認知することができるよう IA を拡張した。彼がとった方略は, IA と同様な

構造をもつネットワークをいくつか複製して、同時に働かせることであった。これによって複数の単語を同時に認知することが可能になる。しかし、単にネットワークを複製しただけでは、同じ単語に関する知識が重複して存在することになってしまう。そこで彼は、文字-単語レベル間の結合に関する情報をセントラル・モジュールに貯蔵し、必要に応じてコピーさせることにした。これがProgrammable Blackboard Model (PABLO) である (図 2.6)。

PABLO には特徴レベルはなく、文字レベル以降の過程のみをモデル化している。文字レベルには20個の「スロット」があり、それぞれのスロット中に各文字を表現するユニットが入っている。図のように、PABLO は1つずつスロットをずらしながら IA をいくつも並べたような構造をもっている。そのため、単語が20個のスロットの中のどこに呈示されても、それに対応する単語ユニットが存在することになる。このような構造をもつことによって、単語の位置や長さ(文字数)に関わらず、単語を認知することができる。これらの構造はプログラマブル・モジュールと呼ばれる。単語に関する知識、すなわち文字ユニットと単語ユニットの結合パターンは、セントラル・モジュールに貯蔵されている。プログラマブル・モジュールのユニット間の結合は、その名の通り「プログラム可能」であり、セントラル・モジュールから転送された結合パターンをそのまま再現する。

PABLO は次のような過程を経て単語の認知を行う。単語が呈示されると、プログラマブル・モジュールにおいて、その単語を構成する文字に対応するユニットが活性化する。これらの文字ユニットはセントラル・モジュールと結合している。20個の「スロット」に含

まれるすべてのユニットからの結合は、セントラル・モジュールの1つの単語ユニットに収束している。そのため、単語がどのスロットに呈示されても、呈示された文字をすべて含むようなセントラル・モジュールの単語ユニットが強く活性化する。活性化した単語ユニットは、プログラマブル・モジュール内の文字ユニットと単語ユニットの間のプログラム可能な結合に対して信号を送り、その単語に関する結合を作り出す。その結果、プログラマブル・モジュールの文字ユニットと単語ユニットの間にIAの場合と同様な相互作用が生じ、最終的には呈示された単語に対応する単語ユニットが強く活性化することになる。そして、プログラマブル・モジュールの単語ユニットの中で最も強く活性化したものが、出力として選択される。

ところが、セントラル・モジュールはすべてのスロットの文字ユニットからの情報を同時に受け取るため、文字の位置に関する「混線」が生じてしまう。たとえば、"CAT" と "ACT" という単語は同じ文字から構成されているので、IAと同じような文字ユニットを用いた場合、セントラル・モジュールはこの2つを区別することができない。これを避けるために、プログラマブル・モジュールの文字ユニットは、文字の種類と相対的位置を同時に符号化するようにつくられている。たとえば "CAT" という単語を構成する文字は、{*C, C, _A, A, _T, T*} という文字ユニットによって表現される。{}内のユニットは、すべて同時に活性化する。"*" は空白を表し、"_ " は任意の文字を表す。たとえば、*C というユニットは「左側に空白がある C」を、A_ というユニットは「右側が空白でない（任意の文字がある）A」を表現する（文字や単語を表す場合と区別するために、これらのユニットは引用符を付けずに表記する）。これ

に対して、"ACT" は { *A, A_, _C, C_, _T, T* } という文字ユニットによって表現される。両者のパターンは異なっているので、セントラル・モジュールはこれらの単語を識別することができる。このようにして、単語がどこに呈示されても、また同時に複数の単語が呈示されても、PABLO はそれを認知することができる。

しかし、まぎらわしい単語を同時に呈示した場合には、誤りが生じやすくなる。たとえば "SAND LANE" という2つの単語を含む文字列を呈示すると、"SANE" や "LAND" を表現するセントラル・モジュールの単語ユニットも活性化してしまう。そのため、処理のための時間が十分に与えられない場合には、"SANE" や "LAND" という誤った出力が選択される確率が、ある程度生じる。しかし、十分な時間がある場合には、プログラマブル・モジュール内での単語レベルと文字レベルとの相互作用によって、実際に呈示された単語を表現するユニットの活性度が高まり、誤った単語が選択される確率は0に近くなる。これは、文字レベルにおいては常に正しいユニットが活性化するためである。このような「混線」によるエラーは、心理実験における文字の混合エラーに対応すると考えられる。冗長性を減らすために導入したセントラル・モジュールが文字の混合エラーの原因となることは興味深い。

しかし、入力表現の妥当性という点ではどうであろうか。IA の場合と同様、単語を呈示する位置をずらすと、文字レベルの活性度パターンは大きく変化する。一方、すべての文字ユニットからの結合はセントラル・モジュールに収束しているので、セントラル・モジュールにおいては呈示位置に関する不変性が実現している。このような手法によって入力表現に関する問題は解決できたのであろう

か。

PABLO の構造を詳しく検討してみると、この問題は実際には解決されていないことが分かる。PABLO の最終的な出力はプログラマブル・モジュールの単語ユニットであるが、このユニットが位置に依存した表現となっているからである。前述のように、このユニットは位置をずらしながら複製されている。そのため、同一の単語を表現するユニットが複数存在することになる。ここで問題になるのは、それら同一の単語を表現するユニットが「同じものを指している」ことを示す情報が存在しないことである。たとえば、左端のスロットから始まる位置に "CAT" が入力された場合と、そこからスロット 1 個分右にずらした位置に "CAT" が入力された場合とでは、異なる単語ユニットが活性化する。これらのユニットはあらかじめ単語 "CAT" を表現するユニットとしてつくられたものであるが、PABLO の外部のシステムにとっては、これらが同じものを指していることを知る方法は存在しない。単語の表現を複製するという方略をとったために、PABLO では単語に関するタイプ-トークン問題が生じてしまった、ともいえる。一方、セントラル・モジュールの中には位置に依存しない単語の表現が存在するが、複数の単語が入力された際の混線はこのモジュールの内部では解決できないため、出力として使用することは不可能である。

さらに、結合パターンのコピーを行うという点も問題の 1 つである。そのようなメカニズムが脳内に存在するという神経生理学的な証拠は、現在のところ見つかっていない。また、ユニット間の結合をどこへコピーすべきかを学習するための方法も不明である。

PABLO は呈示位置に依存しない単語の認知・複数の単語の認知・

文字の混合エラーの説明を実現することができた。しかし、脳内に存在するメカニズムのモデルとして考えた場合には、不十分な点が多いといえる。

3. BLIRNETとMORSEL

より妥当性の高いモデルを構築するためには、入力表現に関する十分な検討を行う必要がある。呈示された文字の位置が変化してもそれが同じ文字であることを示す情報を保持し、かつ文字同士の位置関係の変化に関する情報を保つことができるような入力表現が望ましい。

神経生理学的な研究からは、生体が位置に関する不変性を徐々に実現していくことによってこの問題を解決していることを示す証拠が得られている。視覚野の側頭葉に向かう経路上の神経細胞は、処理の段階が進むにつれてより複雑な刺激に反応するようになる。そしてその一方で受容野の大きさは拡大していく (e.g. Rolls, 1994)。たとえば、サルの下部側頭葉における神経細胞はある程度複雑な刺激に対して選択的に応答し、直径20~30度という非常に広い受容野をもつ (田中, 1994)。このような側頭葉に向かう経路 (腹側系) は、対象が「何」であるのかに関する処理を行うと考えられている。一方、頭頂葉に向かう経路 (背側系) は対象が「どこ」にあるのかに関する処理を行うと考えられている (Ungerleider & Mishkin, 1982)。対象の形に関する情報は、位置に関する情報と切り離されて表現されているようである。このような知見は、視覚的失認に関する Farah (1990) の分類の枠組みと一致している。

「何」と「どこ」を分離して表現することは、単語の認知だけで

なく、対象の認知一般にとって有効である。我々は対象が移動してもそれを同一のものとして認知しなければならないからである。もし視覚系がこのような性質を備えていなかったとしたら、対象の位置が変化する度に、我々はそれが何であるのかを新たに学び直さなければならないになってしまうであろう。一方、これまでに議論したように、単語認知を行うためには文字同士の相対的な位置関係に関する情報も不可欠である。この情報は、対象が何であるのかに関する情報と共に、腹側系で表現されていると考えられる。

BLIRNET

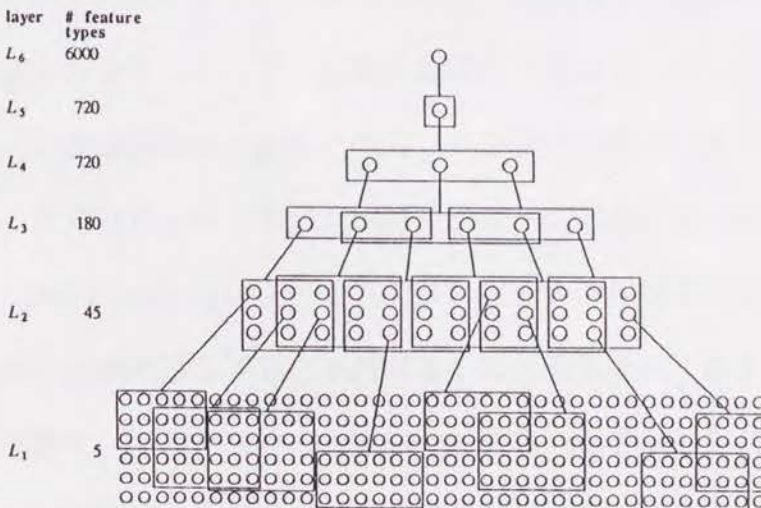


図 2.7 BLIRNETの構造. 単純化のために、各層につき 1 種類の特徴を表現するユニットだけが描かれている。ユニットを囲む長方形は、直線で結ばれたユニットの受容野を表す (Mozer, 1991 より抜粋)。

Mozer (1991) は、多層のニューラルネットを用いることによって、位置に関する不変性の実現と相対的位置関係に関する情報の保持を両立するような表現を作り出すことができることを示した。このモデルは BLIRNET (it Builds Location Invariant Representation of multiple

letter strings) と呼ばれる。BLIRNET は 6 層からなるニューラルネットワークで (図 2.7), 複数の単語を視野内での位置に関わらず認知することができる。

BLIRNET の入力層は一次視覚野のハイパーコラムに類似した構造を持ち, 縦 6 × 横 36 個のコラム内にそれぞれ 5 種類ずつの特徴検出器が配置されている。一方, 出力層は文字クラスターユニット (letter-cluster unit) と呼ばれるユニットから構成される。文字クラスターユニットは, 3 つの文字と, それらの相対的位置関係を表現する。

BLIRNET は文字クラスターユニットの活性度パターンによって単語を表現する。たとえば "CAT" という単語は, {**C,**_A,*CA,*_AT,*C_T,CAT,C_T*,AT*,A_**,T**} というユニットが同時に活性化することによって表現される (PABLO の場合と同様, "*" は空白を, "_" は任意の文字を表す)。このような表現の方法は分散表現と呼ばれる。分散表現は単一のユニットによる表現 (局所表現) より効率が良く, 項目間の類似度が反映されるという利点をもっている。文字クラスターユニットの可能な組み合わせは 56,966 種類存在するが, 英語の語彙全体の 95% を表現するためには 6,000 個の文字クラスターユニットで十分である⁽¹⁾ (実際のシミュレーションでは, 計算資源の制約から 540 個の文字クラスターユニットが使用された)。

入力層と出力層の間の 4 つの中間層は, 呈示された文字や単語の位置に関する不変性を実現しつつ, 文字間の位置関係に関する情報を保持することができるようにつくられている。入力層と同様, これらの層もハイパーコラム状の構造をもっており, コラムの中に様々な特徴に反応するユニットが配置されている。図 2.7 に示されるよ

うに、各層における一定の領域内のユニットからの結合は次の層の1個のユニットに収束している。各層のユニットは、その前の層が表現する特徴のランダムな組み合わせに対して反応する。すなわち、第 n 層のあるコラムに含まれるユニットは第 $n-1$ 層の一定の領域内のユニットとランダムに結合している。空間的な一様性を保つために、この結合パターンはコラム間で一定である。このような構造によって、1つの層に含まれるハイパーコラムの数は、出力層に近づくほど少なくなる。その一方で、1つのコラムの中に含まれるユニットの数は増加する。その結果、入力層における位置情報は徐々に失われ、それと同時に個々のユニットはより複雑な特徴を表現するようになっていく。これらのユニットが表現する特徴の組み合わせはランダムに決められたものであるが、呈示された刺激を識別するために十分な情報を含んでいる⁽¹²⁾。第5層までの結合パターンはあらかじめ定められており、すべてが順方向で、フィードバック型の結合は存在しない。第5層のユニットの活性度パターンは、文字や単語が入力層のどこに呈示されても、それが同じ文字（単語）であることを示すことができるような情報を含んでいる。

このように、第1層から第5層までのネットワークによって、BLIRNETは「呈示された文字の位置が変化してもそれが同じ文字であることを示す情報を保持し、かつ文字同士の位置関係の変化に関する情報も保つ」ことを実現している。この第5層における表現は、これまでに述べたような入力表現に関する問題を解決しているといえる。BLIRNETにおいては中間層に位置しているが、これまでの議論の流れに合わせるならば、第5層における表現を「単語認知過程に対する適切な入力表現」とみなすことができる。

第5層と出力層との間の結合は、学習によって決定される。学習の手續きは以下のようなものである。結合強度の初期値をランダムに決定する。その後、単語を入力層のランダムな位置に1個ずつ呈示し、出力層の活性度パターンを計算する。出力層では、呈示した単語に対応する文字クラスターユニットのみが活性化することが望ましい。そこで、出力層の活性度パターンが望ましい活性度パターンに近づくように、Widrow & Hoff (1960) の最小二乗規則 (LMS rule) を用いて結合強度を変化させる (補遺3)。これは教師あり連想学習の一種である。これを繰り返すことによって、BLIRNET は正しい文字クラスターユニットを活性化させることができるようになる。

第5層までのネットワークが位置に関する不変性を実現し、単語を文字クラスターユニットによって表現することで、BLIRNET はタイプ-トークン問題をうまく解決している。文字クラスターユニットの段階では単語の呈示位置に関する情報は失われているので、位置の違いに応じて複数の文字クラスターユニットを用いる必要はない。また、文字クラスターユニットは、その定義上、文字同士の相対的位置関係に関する情報を保存している。さらに、複数の文字クラスターユニットによる分散表現を用いることによって、単語を弁別するのに十分な情報を保持することができる。

BLIRNET のもう1つの優れた点は、単語認知以外の過程に対しても適用できることである。認知すべき対象を入力層に呈示し、文字クラスターユニットに代わる対象の表現を教師パターンとして学習を行えば、様々な対象の認知が可能である。このとき、単語の場合と同様、複数の対象を同時に認知することもできる。BLIRNET がこのような特性をもっていることは、純粹失読が腹側性同時失認の徴

候の1つであるとする Farah (1990) の枠組みと合致している。

MORSEL

学習後の BLIRNET は、単語がどこに呈示されてもその単語に対応する文字クラスターユニットを活性化させることができる。しかし、呈示された単語と部分的に合致する文字クラスターユニットも活性化するので、出力層の活性度パターンはあまり明確ではない。

そこで Mozer は、BLIRNET から明確な出力を取り出すための抽出ネットワーク (Pull-Out Network) と、注意のメカニズム (Attentional Mechanism) を考案した。これらを加えたネットワークの総体は、MORSEL (Multiple Object Recognition and attentional SElection) と名付けられた (図 2.8)。

抽出ネットワークは、互いに整合性をもつ文字クラスターユニッ

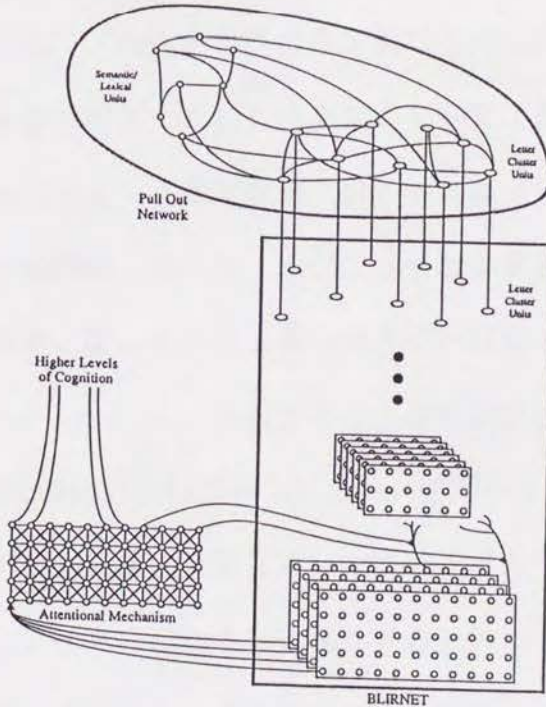


図 2.8 MORSEL の構造 (Farah & Ratcliff, 1994 より抜粋)。

トの活性度を強め、矛盾するユニットの活性度を弱める働きをする。たとえば、ユニット CAT と *CA は相互に興奮性の結合で結ばれているが、CAT と RAT は抑制性の結合で結ばれている。これによって整合性をもつ文字クラスターユニットが競合に勝ち残り、単語の一部分だけに反応していたユニットは活性度を失ってしまう。その結果、呈示された単語を表現する文字クラスターユニットだけが活性化することになり、出力が明確になる。

IA や PABLO では文字レベルと単語レベルの相互作用が単語優位効果の原因であったが、MORSEL では抽出ネットにおける相互作用が原因となる。MORSEL には文字レベルと単語レベルの区別がなく、単語と文字が共に文字クラスターユニットによって表現される。たとえば”R”という文字が単独で呈示された場合は、{**R, *R*, R**} という3個の文字クラスターユニットが活性化する。これに対して、単語”WEAR”が呈示された場合は、{WE_R, W_AR, EAR, E_R*, AR*, R**} という6個の文字クラスターユニットが活性化する。活性化するユニットの個数が多いことは競合に有利に働くので、”WEAR”が呈示された場合の方が、文字”R”を表現するユニットがより早く強い活性度を獲得することになる。このような活性度の強さの差が単語優位効果の原因となると考えられる。このような説明はIAの場合とアルゴリズムのレベルでは異なっているが、単語を表現するユニットと文字を表現するユニットとの間の相互作用が原因となっているという点では共通している。

文字の混合エラーについても同様なことがいえる。”SAND”と”LANE”を同時に呈示した場合には、抽出ネットが”SANE”あるいは”LAND”を抽出してしまうことがある。抽出ネットは文字クラス

ターユニット同士の整合性を検出するだけなので、この例のように文字の混合によって単語が生成されてしまう場合には、間違っただけのユニットを抑制することができないからである。文字の混合の原因は単語を表現するレベルで単語の位置に関する情報が失われていることであり、PABLOの場合と共通している。

実際の読みの過程においては、注意のメカニズムによって文字の混合が防止されると仮定されている。注意のメカニズムは、BLIRNETの入力層における一定の領域内の活性化度を高める働きをする。2つの単語が同時に呈示された場合、一方の単語が呈示された領域に注意を向けることができれば、その単語に対応する文字クラスターユニットが強く活性化することになり、正しいユニットが抽出される。文字の混合に関する心理実験では刺激が瞬間的に呈示されるので、個々の単語に十分な注意を向けることができず、エラーが生じると考えられる。注意のメカニズムは入力層のどの領域に「注意」しているのかに関する情報をもっているため、対象の位置に関する情報を表現しているといえる。MORSELはBLIRNETによって「何(what)」を表現し、注意のメカニズムによって「どこ(where)」を表現しているのである。注意によって位置の情報が表現されるという考え方は、Treismanの特徴統合理論と共通している。

MORSELは構造が比較的単純であり、生理学的な知見ともよく適合する。さらに、注意のメカニズムによって、単語認知過程だけでなく視覚探索課題などの説明も可能である。現在のところ、MORSELは単語認知過程に関する最も現実的かつ包括的なモデルであるといえるだろう。

とはいえ、MORSELによってすべての問題が解決できたわけでは

ない。残された課題の一つは、文字クラスターユニットという内部表現がどのように生成されるのかという問題である。Mozer (1991) は、BLIRNET の第5層からの出力と、意味や音韻の表現との連想学習を行うことによって、中間層に文字クラスターユニットのような内部表現が生成されるのではないかと主張している。このような学習過程を明らかにすることによって、各種の言語を学習し、それぞれに関する心理実験の結果を説明することができるような一般的なモデルを構築することが可能になるであろう。日本語は、漢字・ひらがな・カタカナの3種類の文字体系を併用し、しかも漢字が表意文字でひらがなとカタカナが表音文字であるというユニークな特性をもっている。これらの特性は、内部表現の性質に対してどのような影響を与えているのだろうか。意味や音韻との連想学習を行うネットワークを構築することは、このような問題を検討するための一つの手がかりになると思われる。

第二に、眼球運動に関しても考慮が必要である。MORSEL は1回の凝視 (fixation) の中の処理だけを扱っており、複数の凝視間の情報の統合がどのように行われるのかをモデル化することは行われていない。この点は、単語認知過程のモデルを文の理解のモデルと関連づける際にも問題になると思われる。

4. まとめ

IA は1個の単語の認知過程に関する定量的な説明を行うことに成功した。PABLO はこれを複数の単語に拡張したが、神経生理学的な妥当性が低い、出力が不明確である、といった問題が生じることになった。これらの問題の原因は図 2.5 のようなスロット型の入力表

現にあると考えられる。BLIRNET は位置に関する不変性の獲得と文字の相対的位置関係の保存を可能にするような表現をつくりだすことによって、これらの問題を解決した。さらに、BLIRNET に注意のメカニズム等を加えた MORSEL は、様々な心理実験の結果に対して明確な説明を与えることができた。

第3章 単語認知過程における文字位置情報の表現

第1節 文字位置情報と文字の混合エラー

1. 文字位置情報の役割

単語を同定するために必要な情報は、「単語がどんな文字から構成されているのか」と「それらの文字がどのような順序で並んでいるのか」という二種類に分けることができる。第1章では、前者の情報を抽出する過程における内部表現について検討した。一方、第2章で検討した単語認知のモデル MORSEL においては、後者の情報が文字同士の相対的位置関係によって表現されている。このような文字位置情報の表現は、人間の単語認知過程において実際に用いられているのだろうか。第3章ではこの点を心理実験によって検討する。

2. 文字の混合現象に対する文字位置の効果

Duncan (1987) は、単語の探索課題を用いて、文字の混合に対する位置の効果を検討した。被験者は1個のターゲット単語を見た後、瞬間呈示されるプローブの中にターゲット単語があったかどうかを判断した。図 3.1 のように、プローブは2つの単語または非単語から構成されていた。対象間条件のプローブは、2つの対象（単語または非単語）の間で文字の混合が生じた場合に、ターゲットが形成されるように作られていた（図 3.1 左）。一方、対象内条件のプローブでは、1つの対象の中で文字が入れ替わった場合にターゲットが形成される（図 3.1 右）。これらのプローブを呈示したときの

「あった」という反応（フォールス・アラーム反応）を比較した結

果、対象間条件の方がフォールス・アラーム反応が多くみられることが分かった。この結果は、個々の文字の位置が単語内での相対的位置によって符号化されていることを示唆している。なぜなら、ターゲットとプローブが共有する文字の単語（非単語）内での相対的位置関係は、対象間条件では同じであるのに対して、対象内条件では異なっているからである。

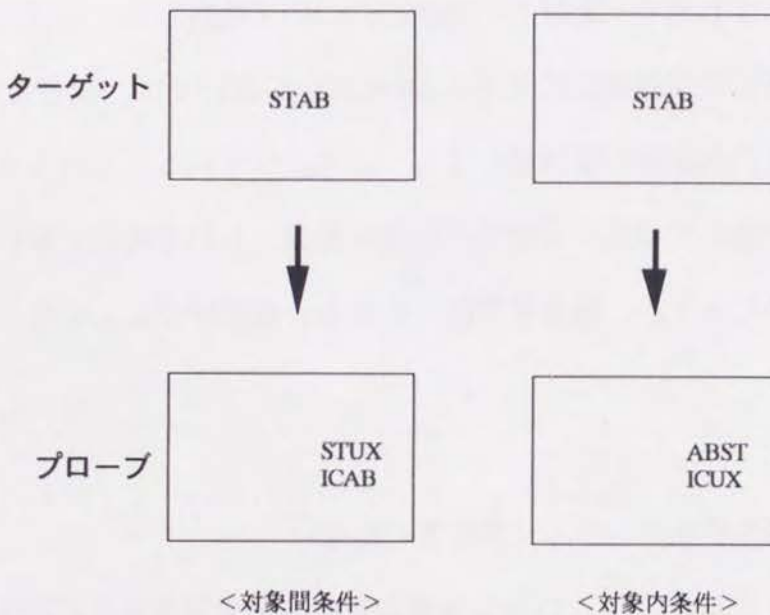


図 3.1 Duncan (1987) の使用した刺激。

しかし、この実験には、文字の位置関係が単語内と単語間で独立に操作されていないという問題点がある。対象間条件では、プローブに含まれる "ST" と "AB" という文字列の位置関係は、単語間・単語内のいずれにおいてもターゲット "STAB" と同じであった。そのため、単語間・単語内のいずれの位置関係が重要なのかははっきりしない。さらに、対象内条件と対象間条件の差が相対的位置関係の違いに起因するのではなく、同一の対象に含まれているかどうか

由来する可能性も否定できない。

そこで、実験1では単語間と単語内の文字の相対的位置関係を独立に操作し、文字の位置が単語内での相対的位置によって表現されているという仮説を厳密に検討する。実験2では、単語を空白によって区切らないという日本語の特徴的な表記法が、文字位置の表現にどのように影響しているのかを検討する。

第2節 実験

1. 実験1：単語内文字位置と単語間文字位置の効果

単語の探索課題を用いて実験を行った。被験者はターゲットである1個の単語を見た後、2個の単語から構成されたプローブを見て、プローブ中にターゲットがあったかどうかを判断した（図3.2）。

ターゲットを構成する2つの文字が、常にプローブの中に含まれるように刺激を作成した。すなわち、ターゲット単語自身がプローブの中に存在しない試行においても、2つの文字はプローブを構成する単語の中に含まれていた。このようなプローブを呈示した場合の「あった」という反応（フォールス・アラーム反応）の頻度は、文字の混合エラーの頻度を反映していると考えられる (Treisman & Souther, 1986)。ターゲットが実際にプローブの中に呈示される試行においては、プローブ中のどちらかの単語をあらかじめターゲットとして選択して呈示した。以下の部分では、ターゲットがプローブの中に呈示されない試行におけるターゲット語を「偽ターゲット」と呼ぶことにする。また、偽ターゲットとプローブが共有する2つの文字を「クリティカル文字」と呼ぶことにする。

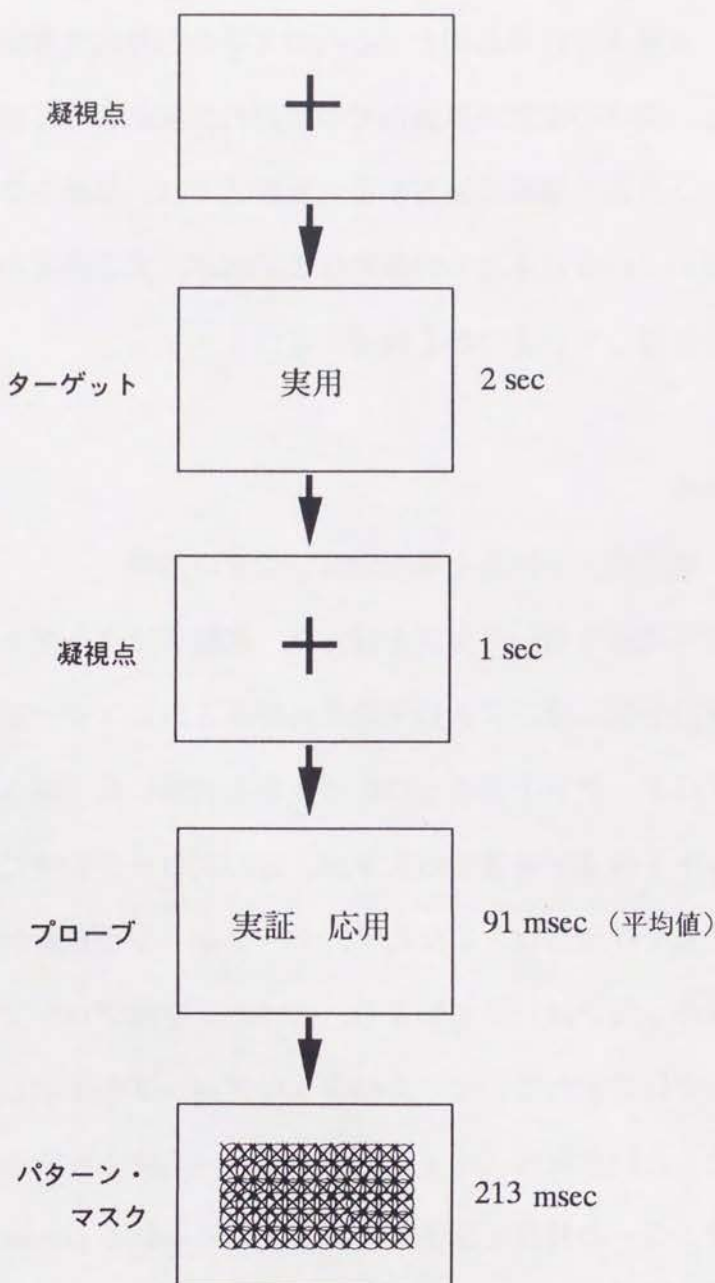


図 3.2 実験 1 で使用した刺激と呈示時間.

単語内の相対的位置によって文字位置が符号化されているのならば、クリティカル文字が単語内文字位置を保存する形でプローブに含まれている場合は、そうでない場合よりフォールス・アラーム率が高くなるはずである。たとえば偽ターゲット「実用」に対して、

プローブが「実証 応用」である場合は、プローブが「真実 用意」である場合よりもフォールス・アラーム率が高くなると予想される。これに対して、単語間の相対的文字位置はフォールス・アラーム率に対して影響を与えないと予想される。たとえば上と同じターゲット「実用」に対して、プローブ「実証 応用」と「応用 実証」のフォールス・アラーム率は変わらないと予想される。

方法

実験計画 単語内文字位置条件（同-異）× 単語間文字位置条件（同-異）の被験者内2要因計画。これ以降、実験条件を次のように略称する。

単語内文字位置：同 → WS (Within: Same)

単語内文字位置：異 → WD (Within: Different)

単語間文字位置：同 → BS (Between: Same)

単語間文字位置：異 → BD (Between: Different)

被験者 京都大学学生・大学院生12名。この他に2人の被験者が実験に参加したが、本試行全体の正答率が90%を越えたために、分析対象から除外した。

装置 パーソナル・コンピュータ(NEC: PC-9801vm21), CRT(SHARP: CU-14AG2), タイマーボード（日本アセンブラ：タイマーボードⅡ）。

刺激 漢字2字熟語を刺激として用いた。これは、刺激抽出のための条件がかなり厳しいものであったので、日本語において最も種類の豊富なタイプの単語を用いることが必要とされたためである。

表 3.1 実験 1 の刺激例.

(ターゲット「実用」に対するプローブ)

		単語間文字位置			
		同		異	
単語内文字位置	同	実証	応用	応用	実証
	異	真実	用意	用意	真実

刺激例を表 3.1 に示す. ターゲットとプローブは, 国立国語研究所(1983)の度数順 M 単位表に収録された使用頻度 7 以上の漢字 2 字熟語の中で, 旺文社 (1973) の国語辞典にも収録されている単語 2,326 個から, 以下のような手続きで作成した⁽¹³⁾.

- (1)偽ターゲット単語をランダムに 1 個選択する.
- (2)偽ターゲットを構成する 2 つの文字をそれぞれ 1 個ずつ含むような 2 つの単語をランダムに選択し, プローブとする. その際, プローブは, これら 2 つの文字の単語内での相対的位置関係がターゲットと同じもの (WS 条件) と, 逆になっているもの (WD 条件) の, 2 組が存在しなければならない.
- (3)偽ターゲットとプローブが以下の基準を満たすかどうかをチェックする. 偽ターゲット語の頻度は, プローブを構成する 2 つの単語の頻度の平均の 1/2~2 倍の範囲内. プローブを構成する文字の中の, クリティカル文字以外の 2 つの文字間の画素数 (CRT 上で 1 個の文字を構成する画素の数) の差は 20 個以内⁽¹⁴⁾. クリティカル文字の読み方は, 偽ターゲットとプローブで同じ (ただし促音は本来の読み方と区別しない). 同一の単語は重複して用いない.

以上のような手続きによって、48個の偽ターゲットと、それぞれの偽ターゲットに対応するWS・WD条件のプロープを1個ずつ作成した。単語間の文字位置は、プロープ内の単語の呈示位置を左右逆にするによって操作した。すなわち、WS・WD条件のプロープのそれぞれ半数ずつを、2つの単語の呈示位置を入れ替えることによってBS条件・BD条件に振り分けた。

96回の試行を1つのブロックとし、1人の被験者に対して2ブロック、192回の試行を行った。BS-BD条件の振り分けは、ブロック間（被験者内）でカウンターバランスがとられた。ターゲットあり試行とターゲットなし試行の回数は等しかった。その結果、1人の被験者は192回の試行の間に同一のターゲットを2回ずつ目にし、各条件のプロープをそれぞれ1回ずつ見たことになる。一方、各プロープに対するターゲットあり-なしの振り分け方は、被験者間でカウンターバランスをとった。

ブロック内の刺激の呈示順序と、ターゲットあり試行におけるターゲットの選択の仕方（プロープ中の左右どちらの単語がターゲットになるか）は、独立に、被験者ごとに乱数化された。また、各プロープに対するBS-BDの振り分けの仕方は、独立に、ターゲットあり-なしのカウンターバランスのペアとなる2人の被験者ごとに乱数化された。

刺激を構成する文字は、16×16個の画素で構成されていた。1文字の大きさは視角にして0.6°だった。プロープにおける単語間の空白の大きさは、ちょうど1文字分だった。

手続き 1回の試行の手続きは以下の通り。被験者はCRTから70cm離れたところから、あご台に顔を固定して刺激を見た。まず凝

視点が呈示され、ビーブ音の後ターゲットが2秒間呈示された。被験者はターゲットを声に出して読み上げることが求められた。ターゲットのオフセット後ふたたび凝視点が1秒間呈示され、1秒間のブランクの後、プローブが呈示された。プローブの呈示時間は、平均正答率が75%程度になるように、練習試行において被験者ごとに調節した。プローブのオフセットと同時に、パターン・マスクが213ms呈示された。続いて反応を促すキューが呈示され、被験者はプローブの中にターゲットと同じ単語があったかどうかを、「あった」・「多分あった」・「多分なかった」・「なかった」の4段階で判断することを求められた。被験者が反応用のスイッチを押すと、どのスイッチを押したのかについてのフィードバックが行われた。実験は自動制御され、試行間の間隔は被験者のペースに委ねられた。反応用スイッチの並ぶ順序は、被験者間でカウンターバランスされた。

32回の練習試行の後、192回の本試行を行った。練習試行においては、いわゆる1-up, 2-down法を用いてプローブの呈示時間を調整した。すなわち、2回連続して正答であったときには呈示時間を1フレーム(17.7ms)短くし、1回でも誤答があった場合は呈示時間を1フレーム長くし、それ以外の場合は呈示時間を変えなかった。呈示時間の初期値は114msであった。本試行においては、最初は練習試行で得られた呈示時間の値を用い、1ブロックの平均正答率が90%を越えた場合は呈示時間を1フレーム減らし、65%未満であった場合は1フレーム増やした。

結果

呈示時間 本試行におけるプローブの呈示時間の平均値は 91ms であった。

反応率に関する分析 条件ごとの各カテゴリーの反応率を表 3.2 に示す。ターゲットがプローブに含まれていなかった試行における、「あった」と「多分あった」という反応をフォールス・アラーム反応とみなした。各カテゴリーの反応率に対して、単語内文字位置条件・単語間文字位置条件・被験者をそれぞれ要因とみなした 3 要因の分散分析を行った。その結果、フォールス・アラーム率に対する単語内文字位置の主効果が、「あった」、「多分あった」の 2 つの反応カテゴリーとも有意であった（「あった」： $F(1,11)=12.19$, $p<0.01$ 、「多分あった」： $F(1,11)=26.21$, $p<0.01$ ）。単語内文字位置が同じである場合の方が、異なっている場合よりもフォールス・アラーム率が高かった。また、コレクト・リジェクション率に対する

表 3.2 実験 1 の反応率 (%)。

反応 カテゴリー	単語内文字位置							
	同				異			
	単語間文字位置				単語間文字位置			
	同		異		同		異	
	ターゲット 有	無	ターゲット 有	無	ターゲット 有	無	ターゲット 有	無
あった	44	8	44	7	40	1	47	0
多分あった	31	20	32	18	31	6	28	3
多分なかった	19	30	18	35	20	35	17	33
なかった	6	43	6	40	8	58	8	64

単語内文字位置の主効果が、「なかった」のみで有意であった

($F(1,11)=22.58, p<0.01$)。単語内文字位置が同じである場合の方が、異なっている場合よりも「なかった」というコレクト・リジェクション率が低かった。

単語間文字位置の効果や、2つの効果の交互作用は、いずれも有意ではなかった。

信号検出理論による分析 さらに、信号検出理論にもとづいた分析を行った。被験者ごとに、各条件における Goodman-Kruskal のガンマ係数を算出した。ガンマ係数は d' と同様、被験者の判断基準の変動とは独立に弁別の容易度を表す（ガンマ係数の値が小さいほど弁別が困難である）。 d' や ROC area でなくガンマ係数を用いた理由は以下の通りである(Nelson,1987)。

- (1)S事態とS+N事態における反応が、それぞれ分散の等しい正規分布をなすという仮定が不要である。
- (2)1.0や0といった累積反応率がみられた場合にも、補正值を用いる必要がない。
- (3)直線あるいは曲線のフィッティングの問題が生じない。

条件ごとのガンマ係数の値を図 3.3 に示す。値が1に近づくにつれて分散が小さくなる傾向がみられたので、ガンマ係数を角変換した値に対して分散分析を行った。その結果、単語内文字位置条件の主効果が有意にみられた ($F(1,11)=21.40, p<0.01$)。単語内文字位置が同じである場合の方が、異なっている場合よりもガンマ係数が小さかった。単語内文字位置が偽ターゲットとプローブで同じである場合の方が、ターゲット有無の判断が困難であったといえる。その他には有意な効果はみられなかった。

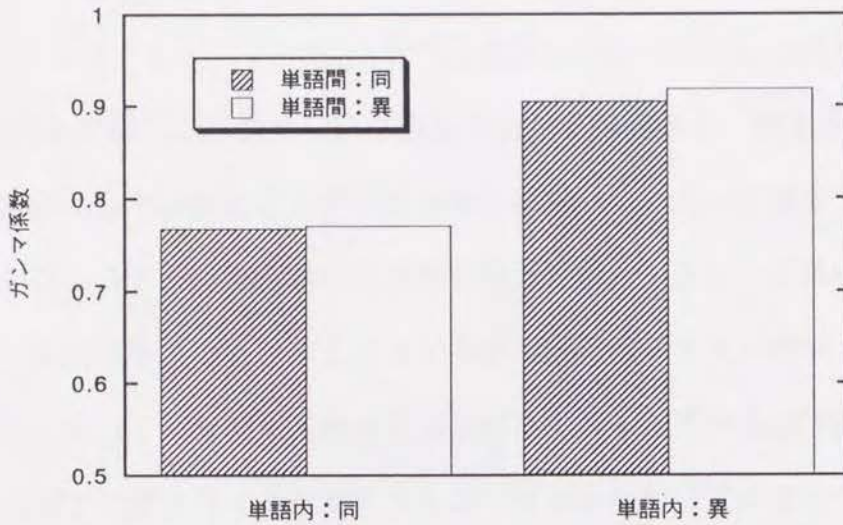


図 3.3 実験 1 のガンマ係数.

ターゲット呈示位置に関する分析 ターゲットあり試行におけるターゲットの呈示位置（左右）の、ヒット率に対する効果を分析した。いずれの反応カテゴリーにおいても、呈示位置による有意な効果はみられなかった。

考察

仮説から予想された通り、フォールス・アラーム率に対する単語内文字位置の主効果が有意にみられた。文字位置に関する情報は、単語内の相対的位置によって表現されていると考えられる。ガンマ係数を用いた分析においても反応率に関する分析と同様の結果が得られたことは、単語内文字位置の効果が、被験者の判断基準点の移動によるものではないことを示している。単語内文字位置の効果がみられたことは、Duncan (1987) の結果と一致している。

一方、単語間文字位置の主効果や、2つの効果の交互作用はみら

れなかった。これらの効果は先行研究では検討されていなかった。

2つの単語の間に空白が存在する場合には、単語間文字位置に関する情報は用いられていないと考えられる。

課題の妥当性 この実験では探索課題を用いたために、単語全体を報告させる場合と比べて、被験者の記憶に与える負荷は小さかったと考えられる。さらに、信号検出理論を用いた分析を行うことによって、反応時のバイアスの影響は除去することができた。その反面、探索課題ではターゲットを先に呈示するために、プローブに含まれるクリティカル文字はターゲットによるプライミングを受けていたことになる。しかし、その効果は各条件に対して均等に働いたと考えられ、ここで検討した仮説に直接的な影響を与えたとは考えにくい。したがって、この実験の結果から、文字の混合に対して単語内での文字の位置関係が影響を与えること、及び日本語の漢字2字熟語においても文字の混合が生じることを、厳密に検証することができたと思われる。

2. 実験2：漢字とひらがなの分節化

実験1では、文字位置に関する情報が単語内の相対的位置によって表現されているという仮説が支持された。⁹しかし、この実験で使用したプローブは、単語と単語の間が一文字分の空白によって区切られているという点で、やや不自然であった。日本語の文章は、英語等とは異なり、空白による区切りを用いないで表記されるからである。

実験2では、実験の状況をより自然な読書事態に近づけるために、単語間の空白をひらがなで置き換えたプローブを用い、実験1と同

様な手法で実験を行った。実験2で用いられたプローブは、「あ実証き応用せ」のような、プローブ中の単語に隣接する空白をランダムにひらがなで置き換えたものだった。意味的な要因が入り込まないようするために、「は」や「を」などの助詞に相当するひらがなは用いられなかった。

方法

実験計画 単語内文字位置条件（同-異）× 単語間文字位置条件（同-異）

被験者 京都大学学生・院生・助手10名。

装置 カラー AV タキストスコープ（岩通アイセルIS-701C）及びパーソナル・コンピュータ(NEC: PC-9801 VX)

刺激 刺激例を表 3.3 に示す。実験1と同一の刺激リストに、3文字のひらがなを加えることによって刺激を作成した。

すべてのひらがな（41文字）の中から、助詞としても用いられる文字を除き、さらにその中から26文字をランダムに選択した。この26文字の中から、各被験者の各試行ごとにランダムに3文字を選択

表 3.3 実験2の刺激例。

（ターゲット「実用」に対するプローブ）

		単語間文字位置	
		同	異
単語内文字位置	同	あ実証き応用せ	ひ応用ぬ実証た
	異	い真実ち用意く	み用意ゆ真実え

して使用した。

実験条件・ターゲットの有無・ターゲットあり試行におけるターゲット選択の振り分けの仕方は、実験1と同様な方法で乱数化された。ただし、実験2では48試行を1ブロックとして、1人の被験者に対して4ブロックの試行を行った。1ブロックの中には、同一の偽ターゲットに対応する4つの条件のプロープがすべて含まれていた。実験1より細かく試行をブロック化したのは、本試行におけるプロープの呈示時間の調整を、より精密に行うためである。

実験1と同様、1人の被験者に対して同じ偽ターゲットが2回ずつ呈示された。ところが、実験2では実験1より試行を細かくブロック化したために、被験者が「繰り返し呈示されるターゲットは、常にプロープ中には呈示されない」ということに気付く可能性が生じた。そこで、「同じターゲットが繰り返し呈示され、かつプロープ中にも呈示される」という事態を作り出すために、ディストラクタ刺激を作成した。実験1の刺激リストの選択に用いた2,326個の単語の中から、偽ターゲットやプロープと重複しない単語をランダムに選択し、ディストラクタ用のターゲットとプロープとして使用した。このプロープには常にターゲットが含まれていた。ディストラクタ刺激は1ブロックの中に4試行挿入された。

1人の被験者に対して4ブロック、計208回の試行を行った。ブロック内の呈示順序、及びブロックの順序は被験者ごとに乱数化された。

刺激を構成する文字は、実験1とは異なり、48×48画素で構成されていた。しかし、1文字の見えの大きさは実験1と同じく0.6°であった。

手続き 実験1 とほぼ同様な手続きが用いられたが、以下の点が異なっていた。

- (1)被験者はターゲットを声に出して読み上げることが求められなかった。実験1では被験者がターゲットを見ることを強制するためにこれを行ったが、被験者の負担を減らすために、実験2では取りやめた。
- (2)反応スイッチの並び方に関するカウンターバランスが行われなかった。これは、実験1において有意な効果がみられなかったためである。
- (3)フォールス・アラーム率が低くなりすぎないようにするために、ターゲットがプロブ中に呈示された試行において被験者が「多分なかった」あるいは「なかった」という反応をした場合（ミス事態）には、ビーブ音によるフィードバックを行った。
- (4)実験2に用いたCRTは、1フレームの長さが実験1より短く、10msであった。呈示時間の1回の調整幅は、練習試行では2フレーム、本試行では1フレームだった。また、パターン・マスクの呈示時間は200ms、練習試行における呈示時間の初期値は100msであった。

結果

呈示時間 本試行におけるプロブの呈示時間の平均値は100msであった。

反応率に関する分析 条件ごとの各カテゴリーの反応率を表3.4に示す。これらの値に対して実験1と同様な3要因の分散分析を行った。その結果、フォールス・アラーム率に対する単語内文字位置の主効果が、「あった」、「多分あった」の2つの反応カテゴリーと

も有意であった（「あった」： $F(1,9)=7.46, p<0.05$ ，「多分あった」： $F(1,9)=9.90, p<0.05$ ）。単語内文字位置が同じである場合の方が、異なっている場合よりもフォールス・アラーム率が高かった。さらに、フォールス・アラーム率に対する単語間文字位置の主効果が、「多分あった」のみで有意にみられた。単語間文字位置が同じである場合の方が、異なっている場合よりも「多分あった」というフォールス・アラーム率が高かった（ $F(1,9)=10.83, p<0.01$ ）。また、ミス率に対する単語間文字位置の主効果が、「多分なかった」のみで有意にみられた（ $F(1,9)=8.32, p<0.05$ ）。単語間文字位置が同じである場合の方が、異なっている場合よりも「多分なかった」というミス率が高かった。その他には、有意な効果はみられなかった。

表 3.4 実験 2 の反応率（％）。

反応 カテゴリー	単語内文字位置							
	同				異			
	単語間文字位置				単語間文字位置			
	同		異		同		異	
ターゲット 有	ターゲット 無	ターゲット 有	ターゲット 無	ターゲット 有	ターゲット 無	ターゲット 有	ターゲット 無	
あった	36	12	38	10	35	4	40	3
多分あった	33	29	39	25	34	25	35	14
多分なかった	22	31	18	28	22	33	19	38
なかった	8	28	5	36	9	37	7	45

信号検出理論による分析 条件ごとのガンマ係数の値を図 3.4

に示す。実験 1 と同様、ガンマ係数を角変換した値に対して分散分

析を行った。その結果、単語内文字位置条件の主効果が有意にみられた ($F(1,9)=6.05, p<0.05$)。単語内文字位置が同じである方が、異なっている場合よりもガンマ係数が小さかった。実験1と同じく、単語内文字位置が偽ターゲットとプローブで同じである場合の方が、ターゲット有無の判断が困難であったといえる。さらに、単語間文字位置条件の主効果も有意にみられた ($F(1,9)=7.85, p<0.05$)。単語間文字位置が同じである場合の方が、異なっている場合よりもガンマ係数が小さかった。単語間文字位置が偽ターゲットとプローブで同じである場合の方が、ターゲット有無の判断が困難であったといえる。その他には、有意な効果はみられなかった。

ターゲット呈示位置に関する分析 ターゲットあり試行におけるターゲットの呈示位置（左右）の、ヒット率に対する効果を分析した。実験1と同様、呈示位置による有意な効果はみられなかった。

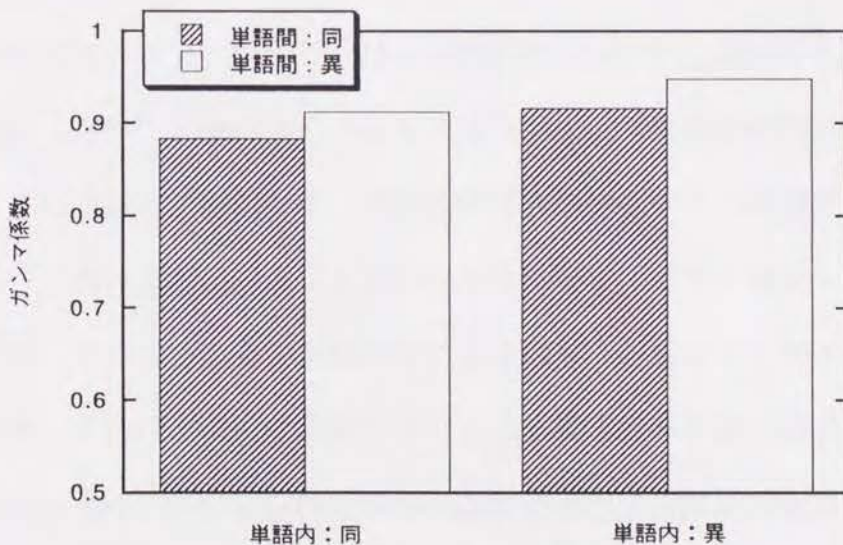


図 3.4 実験2のガンマ係数.

考察

実験2においても、フォールス・アラーム率に対する単語内文字

位置の主効果が有意にみられた。一方、実験1とは異なり、フォールス・アラーム率に対する単語間文字位置の主効果も有意にみられた。これは、単語内文字位置だけでなく単語間文字位置に関する情報もプローブの認知過程で用いられていたことを意味している。

実験2においても単語内文字位置の効果がみられたことは、漢字とひらがなの分節化が、文字位置を表現する以前に行われていることを意味する。実験2のプローブでは単語間に空白がないので、プローブを漢字とひらがななどに分節化した後でなければ単語内文字位置そのものが定義できないと考えられるからである。日本語の認知過程においては、漢字とひらがなの相対的位置関係にもとづく文字位置の表現が行われているのかもしれない。日本語はふつう漢字かな混じり文で表記されるので、このような表現は有効に働くと考えられる。文章をひらがなだけで表記すると単語の区切りが分かりにくくなるという事実も、この仮説を支持しているように思われる。

一方、単語間文字位置の効果がみられたことは、プローブ全体の大局的な特徴を検出する過程が存在することを示唆している。低空間周波数領域における特徴分析の過程が、それに相当するのかもしれない。実験1ではプローブが空白によって明確に区切られていたので、そのような過程も単語を1つずつ分析することになり、単語間文字位置の効果が見られなかったのではないかと思われる。異なる空間周波数領域における特徴の分析は並列に行われ、それらを総合して単語の認知が行われているのかもしれない。

単語間文字位置の効果について 反応率に関する分析において、フォールス・アラーム率だけでなくミス率に対しても単語間文字位置条件の効果がみられたことは、BS条件とBD条件の間で被験者の

判断基準点が移動した可能性があることを示唆している。しかし、ガンマ係数を用いた分析においてもこの効果は確認された。したがって、判断基準点の移動とは無関係に、BS条件のプロープはBD条件のプロープより弁別が困難であったといえる。

実験1との比較 実験2においては、フォールス・アラーム率が一般的に実験1より高い傾向にあった（実験1:16%, 実験2:31%, ただし統計的検定は行っていない）。これは、実験2においてのみ、ミス事態に対するフィードバックを行ったためであると考えられる。ガンマ係数を用いた分析と反応率に対する分析の結果が一致しているので、この違いは結果の解釈に対して影響を与えないと思われる。

第3節 考察

1. モデルに与える示唆

実験1で単語内文字位置の効果がみられたことは、BLIRNETの文字クラスターユニットのような相対的位置関係にもとづく文字位置の表現が、心理学的にも妥当であることを意味している。文字クラスターユニットは隣接する3つの文字の相対的位置関係を表現している。そのため、偽ターゲットとプロープのそれぞれをBLIRNETに入力したときの文字クラスターユニットの活性度パターンは、WS条件（単語内文字位置が同じ）の方が、WD条件（単語内文字位置が異なる）より類似することになる。したがって、WS条件のほうがフォールス・アラーム率が高くなることが予測される。

一方、実験2においても単語内文字位置の効果がみられたことは、文字同士の相対的位置関係を表現する以前に、漢字とひらがなの区別が行われていることを示している。BLIRNETを日本語の認知過程

に適用する際には、漢字とひらがなを区別する文字クラスターユニットを導入することが必要であろう。英語の場合には、単語と単語の間にはかならず空白が存在する。この空白を、文を単語に分節化する際の手がかりとして使用するために、BLIRNETの文字クラスターユニットは「空白」と「任意の文字」を区別して表現している。日本語の場合には、「任意の文字」をさらに「任意の漢字」と「任意のひらがな」に分けて表現するようなユニットを用いればよい。一方、日本語の文章においても、句読点の上には1文字分に近い大きさの空白が生じる。このような空白は、文章を句や文に分節化する際に使用されていると考えられる。空白・任意の漢字・任意のひらがなを区別するような文字クラスターユニットを用いれば、実験1と実験2における単語内文字位置の効果を共に再現することができるのではないと思われる。

一方、実験2における単語間文字位置の効果を説明するためには、プローブ全体の大局的な特徴を処理する過程が必要である。BLIRNETの入力層の特徴検出器は受容野の大きさが固定されているが、より大きな受容野をもつ特徴検出器を混在させることによって、そのような過程を実現できるのではないと思われる。このような拡張を施すことによってBLIRNETは複数の空間周波数チャンネルをもつことになり、より広範な種類の対象の認知が可能になるだろう。

2. より一般的なモデルに向けて

これまでの議論から、文字の順序に関する情報が文字同士の相対的位置関係によって表現されていることが明らかになった。このような知見は、より高次の認知過程とどのような関係にあるのだろうか

か。

ヒトが用いるすべての言語は「二重の階層性」（松沢, 1991）をもっている。「二重の階層性」とは、単語が文字という要素から構成され、さらに文が単語という要素から構成されていることを指す。文字から単語への階層構造は、具体的には次の2つの性質に帰着する。

- a) 単語は有限の種類 of 文字から構成される。
- b) 単語の意味は、それを構成する文字と、それらの文字の順序によって規定される。

さらに、単語と文の間にも同じような階層構造が存在する。

- c) 文は有限の種類 of 単語から構成される。
- d) 文の意味は、それを構成する単語と、それらの単語の順序によって規定される。

このように類似の階層構造が繰り返されているという意味で、ヒトの言語は「二重の階層性」をもっているのである。もちろん、これら2つの階層構造は全く同一ではない。第一に、単語の種類が有限であるのに対して、生成可能な文の種類は（単語の種類と文の長さが有限である以上、数学的には有限であるが）、ほとんど無限であるといってよい。言い換えると、我々は自らが使用する単語をあらかじめすべて記憶しているが、文に関してはそうではない。今までに経験したことがない出来事に遭遇した場合にも、我々はそれを記述する文を生成することができる。第二に、文字が単語を構成する際の制約条件である正書法的規則は、文の生成を支配する統語的規則とは明らかに異なっている。

これまでの議論から、文字クラスターユニットのような局所的な位置関係にもとづく表現を用いれば、文字の順序を明示的に表現し

なくても単語を認知することができることが分かった。これを一般的な文脈で言い換えると、記号同士の相対的位置関係を記号そのものと同時に表現することによって、記号の順序にもとづく処理を行うことができる、ということである。このような表現は、より高次の認知過程においても用いられているのだろうか。単語と文の間の階層構造と、文字と単語の間の階層構造との共通点・相違点を探ることが今後の課題である。

第4章 総合的な考察

1. 単語認知過程と主成分分析

第1章で提案した主成分分析による情報圧縮という枠組みは、単語認知過程とどのような関係にあるのだろうか。第2章で議論したように、単語認知過程に対する入力表現としては、文字の呈示位置に対する不変性と、局所的な相対的位置関係の保存を両立するような表現がふさわしい。Mozer (1991) のBLIRNET は、第5層においてそのような表現を実現している。この第5層の活性度パターンの恒等変換をネットワークに学習させることによって、文字クラスターユニットのような内部表現を生成することができないだろうか。

残念ながら、第1章と同じネットワークを用いた場合には、恒等変換を学習することによって文字クラスターユニットと同様な性質を持つ内部表現を獲得することはできないのではないと思われる。その理由は、BLIRNET の第5層の活性度パターンが、単語全体の呈示位置の変化に応じて変動することにある。この変動は単語の認知にとっては不必要であり、一種のノイズとみなすことができる。しかし、単語全体の位置に関する情報を消去することと、文字同士の位置関係に関する情報を保存することは、トレードオフの関係にある。そのため、BLIRNET は単語の呈示位置による活性度パターンの変動を、完全に消去することができない。

このような第5層の活性度パターンの恒等変換をネットワークに学習させた場合について考えてみよう。第1章で議論したように、学習後のネットワークが行う変換は主成分分析に近い性質を持つ。主成分分析は入力空間における分散を最大限に保存するように入力

パターンを投影するため、投影後の空間においても、単語全体の呈示位置の変化による変動が保存されることになる。Cottrell et al.

(1988)の分析によれば、ネットワークの中間層は、それぞれの主成分を個別のユニットによって表現するのではなく、中間層全体で分散して表現する。そのため、ネットワークの中間層のユニットは、単語の呈示位置によって異なった反応をすることになってしまう。ネットワークが文字クラスターユニットのような呈示位置に依存しない内部表現を獲得するためには、単語の呈示位置に関する情報を排除する必要がある。

そのための方法の一つとしては、ネットワークの中間層全体で主成分を分散して表現するのではなく、個々のユニットがそれぞれの主成分を表現するように学習規則を変更することが考えられる。これは、ネットワークの学習時に、中間層のユニット活性度の変動が互いに無相関になるような拘束条件を設けることによって実現できるであろう。ネットワークの内部表現を直交化させることによって、単語の呈示位置の変化による変動を、単語を構成する文字の種類・個数・位置関係の変化による変動と分離することが可能であると考えられる。呈示位置に関する情報は特定のユニットによって表現され、残りのユニットは、文字の種類・個数・位置関係を表現すると予想される。残りのユニットが表現する情報は、文字クラスターユニットが表現する内容と一致している。現在、このようなネットワークによって実際に文字クラスターユニットのような表現が生成可能かどうかを検討するためのシミュレーションを計画中である。

2. より広い意味での情報圧縮

より広い枠組みにおいては、言語によるコミュニケーション自体を情報圧縮の過程としてとらえることができる。我々がある簡単な事象を思い浮かべているときでさえ、その事象は脳内の膨大な数の神経細胞の活動によって表現されているはずである。事象の視覚的属性は視覚系における多数の神経細胞の活性度パターンによって表現され、それと同時に、その他の感覚的属性（音、においなど）や意味的に関連する事物を表現する細胞群も活性化していると思われる。ところが、その事象を言語を用いて誰かに伝達する場合には、わずか数語の文でこと足りる場合が多い。その文を聞いたり読んだりした相手は、そこから話し手（書き手）が思い浮かべていた事象を、ある程度正確に再現することができる。わずか 17 文字で風景やそれに伴う心情などまでも伝達することができる俳句は、もっとも効率的な例の 1 つであろう。これら一連の過程は、恒等変換による情報圧縮の過程と類似している（図 3.5 上）。話し手（書き手）は脳内の事象の表現を文という表現に圧縮して伝達し、相手はその文をもとに自分の脳内に事象の表現を再構成する。

もちろん、われわれは話し手（書き手）が伝えようとした事象を常に正しく再現できるわけではない。また、伝達が正しく行われた場合においても、自分自身が発話している際の事象の脳内表現と、他者の発話から再現した事象の脳内表現は、完全に同一ではないであろう（そうでなければ、自己の体験した事象と他者から聞いた事象を区別することができない）。しかし、ある言語を獲得して正確なコミュニケーションができるようになることは、この一連の過程ができるだけ恒等変換に近づくように学習することである、とみな

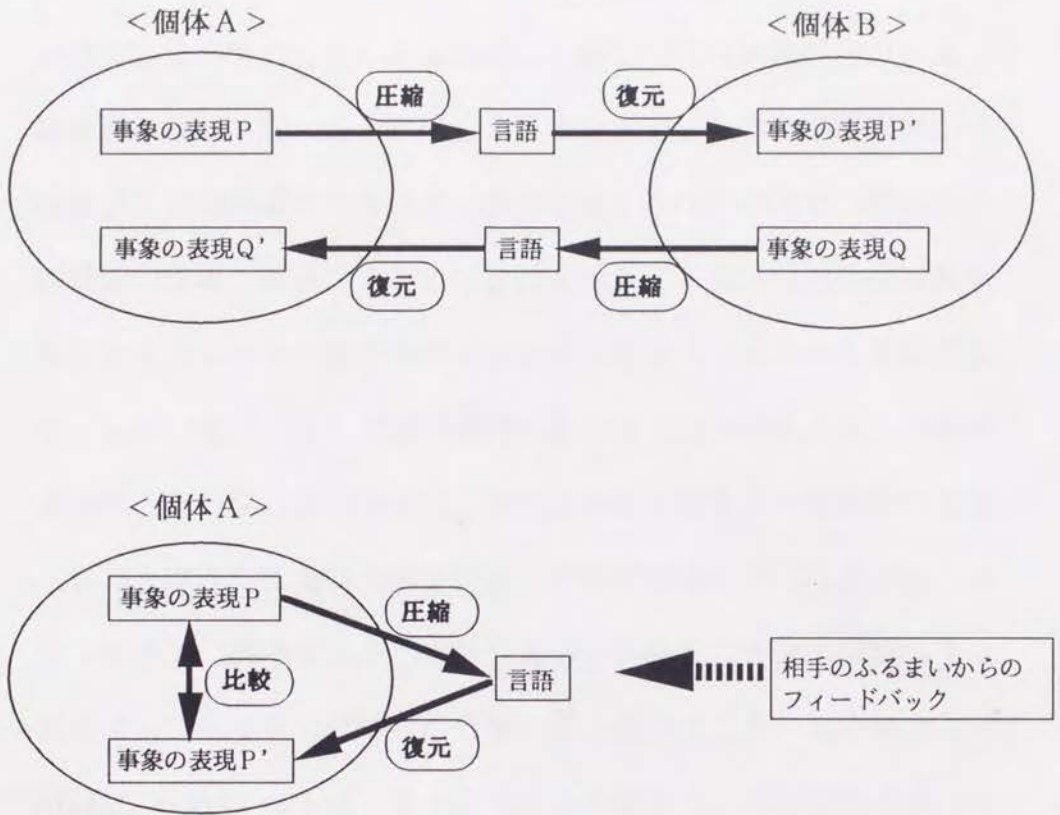


図 3.5 情報圧縮過程としての言語.

せるのではないだろうか.

このような過程が第 1 章で用いたような単純なニューラルネットワークで実現できるとは、もちろん考えられない。この過程には、事象を記述するための文の生成・発話や書字のための運動指令の生成・音素や文字などを構成する知覚的特徴の抽出・単語や文の認知・文の理解などの、一連の複雑な過程が含まれている。さらに、事象の脳内表現がどのようなものであるのかという問題自体も、まだ解明されていない。ある事象を記述するためには、そこに含まれる対象だけでなく、その対象の役割、対象間の関係に関する情報が必要である。それらの情報をうまく統合した表現を生成する試みはいく

つか行われているが、現在のところ、まだ十分な成功を収めているとは言い難い (e.g. St. John & McClelland, 1990) . 加えて、この過程は1つの個体の中で完結していないという問題がある。自分の行った発話に対して、相手がどのような事象の表現を生成したのかを正確に知ることは不可能である。学習の一部は、自己の内部で自らの発話から事象の表現を復元し、最初に意図した事象の表現と比較することによって行われているのかもしれない (図 3.5 下) . これに加えて、自己の発話に応じて相手が行った発話や行為からのフィードバックによる学習も必要であろう。

3. 進化という観点から

最後に、進化という観点から言語に関する研究の意義をとらえ直してみたい。既に多くの研究者が指摘しているように、生物の遺伝暗号の体系は人間の言語との共通点をもっている (e.g. ホフスタッター, 1985) . 遺伝子においては、3つの塩基の組み合わせが特定のアミノ酸に対応している。この関係は、単語が特定の事物を指し示すという関係に類似している。このような遺伝暗号の体系が成立したことによって、細胞というきわめて複雑な構造物を生成・維持し、さらに増殖させることが可能になった。遺伝暗号の体系は、物理・化学法則とは異なるレベルの抽象的な「論理」として、生物の構造を規定している。これと同じように、人間の脳の中に言語という記号系が生まれたことは、神経系の行う情報処理に全く新しい局面をもたらすことになった。言語は神経細胞の活動に依存しているが、細胞の活動電位やシナプス結合の強さなどに関する生理学的な法則より、明らかに抽象的な論理によって生成されている。言語によっ

て思考の内容は複雑化し、個体間のコミュニケーションの効率は格段に向上した。加えて、DNA が生物の遺伝情報を保存し、世代間の伝達を行うための媒体であるのと同じように、文字による文書は、人間の思考の内容を保存し、世代を越えた伝達をするための媒体となっている。

より広い枠組みで考えるなら、進化の過程それ自体を一種の「知性」とみなすことができる（小松, 1987）。生物は、進化を繰り返すことによって様々な環境に適応するように分化してきた。この過程は、環境と生物自体のシステムが与える条件に対して適応度を最大化させる問題解決過程としてとらえることができる。近年の遺伝的アルゴリズムに関する研究は、進化の過程を模倣することによって実際に問題解決が可能であることを示した (Holland, 1992)。コンピュータのアルゴリズムにランダムな「突然変異」を起こさせ、その適応度を特定の評価関数を用いて判定し、選別することによって、アルゴリズムの効率を向上させることができる。

実際の生物の進化は世代交代にもとづいているため、進化による「問題解決」には、数世代年以上にわたる非常に長い時間を要する。これに対して、人間が言語を媒体として行う問題解決は飛躍的に短い時間スケールで進行する⁽¹⁵⁾。人間という種の脳が言語という記号系を扱うようになったことで、生物の進化は新たな局面に入ったといえる。人間が言語を扱うことを可能にしているのは、脳内のどのような機構なのであろうか。その機構はどのように進化してきたのであろうか。これらの問いかけは、生物の進化史という点から見ても極めて興味深い。本研究では、心理実験から推測される文字と単語の脳内表現と、神経系に近い構造をもつニューラルネットワーク

の内部表現とを比較した。文字や単語の表現は、意味や文法などに関するより高次の過程に対する入力として用いられていると考えられる。その意味で、本研究は言語の起源に関するより本質的な問いかけを行うための基礎として位置づけることができるだろう。

第5章 総括

本論では、ニューラルネットワークの学習による内部表現の獲得という観点から、文字と単語の認知過程について検討した。

第1章では、文字の内部表現について検討した。まず、文字同士の混同確率やランダムドット・パターン同士の類似度を測定した先行研究のデータをMDSを用いて系統的に再分析した。MDSによって導出された心理空間の構造は、刺激の大局的な特徴によって規定されていた。この結果は、神経生理学における知見と一致している。文字やランダムドット・パターンなどの2次元的な形は、脳内において大局的特徴の組み合わせによって表現されているのではないかと考えられる。

次に、心理実験で使用されたものと同一の文字やランダムドット・パターンの恒等変換をニューラルネットワークに学習させ、獲得された内部表現を分析した。このようなネットワークが行う変換は主成分分析に近いことが知られている。シミュレーションの結果、アルファベット文字の恒等変換を学習したネットワークの内部表現は、心理実験のデータの分析によって導出された心理空間と類似した構造をもつことが分かった。ランダムドット・パターンの場合も、複雑型細胞の出力にもとづいた入力表現を用いることによって、同様な結果を得ることができた。2次元的な形の脳内表現は、主成分分析による情報圧縮という枠組みで統一的に理解することができると考えられる。

第2章では、まず単語認知過程の入力表現について理論的な考察を行った。その結果、文字の位置に関する不変性と文字同士の位置

関係に関する感受性を両立するような入力表現が必要であることが分かった。

次に、単語認知過程に関連する生理学的な研究を、主として Farah (1990) の研究にもとづいて解説した。純粹失読と呼ばれる脳損傷の患者は、文章を読む能力が顕著に損なわれている。このような脳損傷に関する研究は、単語認知過程が、複数の対象を同時に認知するためのシステムに依存していることを示唆している。さらに、単語認知過程において行われる処理の性質を検討する上で重要な、単語優位効果と文字の混合という2つの心理現象について概説した。

最後に、これまでに提案された3つのニューラルネットワーク・モデルについて解説し、検討を加えた。McClelland & Rumelhart (1981) の IA は、1個の単語の認知過程に関する定量的な説明を行うことに成功した。さらに、McClelland (1986) の PABLO は、これを複数の単語に拡張した。しかし、神経生理学的な妥当性が低い・出力が不明確である、といった問題が生じることになった。このような問題の原因は、入力表現において位置に関する不変性が実現されていないことにあると考えられる。Mozer (1991) の BLIRNET は、網膜像における位置のずれを徐々に許容していくような構造を取り入れることによって、入力表現に関する問題を解決した。さらに、BLIRNET に注意のメカニズム等を加えた MORSEL は、様々な心理実験の結果に対して明確な説明を与えることができた。現在のところ、MORSEL は単語認知過程に関するもっとも優れたモデルであるといえるだろう。

第3章では、文字の位置に関する情報の表現に焦点を当てて、モデルの妥当性を検討するための実験を行った。BLIRNET においては、

文字クラスターユニットと呼ばれる内部表現が用いられている。このユニットは、隣接する文字同士の位置関係にもとづいて文字の位置を表現する。実験1では、このような文字位置の表現が実際の単語認知過程で用いられているのかどうかを検討した。単語の探索課題を用いて実験を行った結果、単語内における文字同士の相対的位置関係が文字の混合エラーの頻度に対して影響を与えることが分かった。この結果は、文字クラスターユニットのような内部表現の存在を支持している。

一方、日本語の文章は、空白によって単語を区切らないという点で、英語などと異なっている。実験2では、このような表記の特性が単語認知過程に対してどのような影響を与えているのかを検討した。得られたデータは、漢字とひらがなの区別が文字同士の相対的位置関係に関する情報を抽出する以前に行われていることを示していた。日本語の認知過程においては、漢字とひらがなの境界が、文を単語に分節化するための手がかりとして用いられているのかもしれない。

第4章では、総合的な考察と今後の展望に関する議論を行った。第1章で提案した主成分分析による情報圧縮という枠組みが単語の内部表現についても適用できるのかどうかに関しては、更なる検討が必要である。一方、より広い枠組みにおいては、言語によるコミュニケーション自体を一連の情報圧縮過程としてとらえることができる。また、ニューラルネットワークを用いて単語の内部表現について検討することは、脳が言語をどのように表現しているのかという問題を解明するための基礎として位置づけることができる。

注

^① 文字や単語を同定する過程は、「認知」ではなく「知覚」と呼ばれることもある。本論では、これらの過程がパターンの学習・記憶を必要とすることを重視する立場から、「認知」という用語を用いることにする。

^② 神経生理学や知覚心理学においては、「対象」ではなく「物体」という用語が用いられることが多い（英語では、両者ともに“object”である）。本論では、主として文字や単語のような抽象的属性をもつものを扱うので、「対象」という用語に統一した。

^③ Young and Yamane (1992) の研究は、個別の顔を表現するいわゆる「おばあさん細胞 (grandmother cell)」の存在を否定しているわけではない。実際、Perret et al. (1989) はサル STS に個別の顔に対して選択的に反応する細胞が存在することを見出している。これらの細胞は、大局的な特徴を表現する細胞からの入力を受けて、個別の顔を符号化しているのかもしれない。

^④ バイアスとは、刺激がまったく見えなかった場合に、被験者が各文字を反応として選択する確率の推測値である。Luce の選択モデルでは、混同行列の非対称性はこのバイアスに起因すると仮定されている。

^⑤ ニューラルネットワークを用いたモデルは、「コネクショニスト・モデル」や「PDP (Parallel Distributed Processing) モデル」と呼ばれることもある。本論では、「ニューラルネットワーク」という用語に統一した。

^⑥ 以下の分析においても、これと同様な手続きで次元数を決定した。その結果、それぞれの分析におけるストレスは異なった値をとることになった。しかし、すべての分析においてこの値はかなり小さく（.12 以下）、ストレスの大きさの違いが結果に対して意味のある影響を与えたとは考えにくい。

^⑦ MDS によって得られた布置は項目間の距離に関する情報しかもたないため、軸の向きや符号自体には意味がない。そのため、布置同士の比較を行う

際には回転と反転によって2つの布置をできるだけ一致させておく必要がある。そこで、相関係数を算出する際に、各軸上の順位相関係数の平均値が最大になるように2つの布置を回転させた。さらに、相関係数がより高くなる場合には、各軸に関して布置を反転させた。このような操作は相関係数を高くする傾向があるが、表 1.2 に示されるように、相関係数の値は条件間で十分に異なっているようである。

本文中の図においては、ネットワークから得られた布置だけを、対応する心理実験から得られた布置とできるだけ一致するように回転・反転させて表示し、心理実験から得られた布置は回転・反転させていない。これは、1つの実験に対応するシミュレーションを複数行ったために、実験から得られた布置の図を1つに統一する必要があるためである。

⁽⁸⁾後述するように、中間層の個々のユニットが単体で1つの固有ベクトルに相当する働きをしているわけではない。固有ベクトルは入力層から中間層への結合パターン全体の中に分散していると考えられる。

⁽⁹⁾ Farah (1990) が「連想失認」という伝統的な用語を変更しなかったことは、かえって誤解を招きやすい状況を生むことになった。彼女の枠組みにおいては、「連想失認」は「連想」に関する障害ではなく、統覚失認と同様、対象の形の分析過程の障害である。統覚失認は形の分析過程の初期における障害であり、連想失認は同じ過程のより高次の段階における障害とみなすことができる。

⁽¹⁰⁾ 実験心理学においては、プローブ (probe) という用語は「仮説を検討するための刺激」という意味で用いられる。

⁽¹¹⁾これは、英単語を構成する2～3文字単位の文字列が、複数の単語に共有されることが多いためである。

⁽¹²⁾ これらの特徴の組み合わせがランダムであることは、むしろ積極的な意味を持っている。できるだけ少ない数の結合によって情報を伝達しようとするならば、結合の分布の仕方に偏りがある場合より、一様な乱数にもとづいて分布していた方が効率がよいからである。

⁽¹³⁾ M 単位は形態にもとづいた単語の単位であるが、日常的な感覚では独立した単語とみなさないような項目も含まれている（例：「紫外」）。そのような M 単位を除外するために、国語辞典に収録されているかどうかによってふるい分けを行った。

⁽¹⁴⁾ 文字を構成する画素の数に大きな差があると、文字全体の密度を手がかりとしてターゲットの有無を判断することが可能になる。これを避けるために、画素数に関する統制を行った。

⁽¹⁵⁾ ただし、ウィルスのように世代交代のサイクルが非常に短い生物の場合は、進化の速度は人間の長期的な思考と比較しうる程度に速い。

参考文献

- Aderman, D. & Smith, E. E. 1971 Models of word recognition. *Cognitive Psychology*, 11, 133-176
- Allport, D. A. 1977 On knowing the meaning of words are unable to report: The effect of visual masking. In Dornic, S. (Ed.), *Attention and Performance Vol. 6, The Psychology of reading*. Hillsdale, NJ: Lawrence Erlbaum Associates, Pp. 505-533.
- 甘利俊一 1989 神経回路網とコネクショニズム. 東京大学出版会
- Barlow, H. B. 1995 The neuron doctrine in perception. In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences*, Cambridge, MA: The MIT Press.
- Blough, D. S. 1985 Discrimination of letters and random dot patterns by pigeons and humans. *Journal of Experimental Psychology: Animal Behavior Processes*, 11, 261-280.
- Bub, D. N., Black, S. & Howell, J. 1990 Word recognition and orthographic context effects in a letter-by-letter reader. *Brain and Language*, 36, 357-376.
- Cottrell, G. W., Munro, P., & Zipser, D. 1988 Image compression by back propagation: an example of extensional programming. In N. E. Sharkey (Ed.), *Advance in Cognitive Science*, 3. Norwood, NJ: Ablex.
- Duncan, J. 1987. Attention and reading: Wholes and parts in shape recognition - A tutorial review. In Coltheart, M. (Ed.), *Attention and Performance vol. 12, The Psychology of reading*, Hillsdale, NJ: Lawrence Erlbaum Associates, Pp. 39-61.
- Edelman, S. 1995 Representation of similarity in three-dimensional object discrimination. *Neural Computation*, 7, 408-423.
- Elman, J. L. 1992 Implicit learning in neural networks: The importance of starting small. In Umiltà, C. & Moscovitch, M. (Eds.), *Attention and Performance vol. 15, Conscious and nonconscious information processing*, Hillsdale, NJ: Lawrence Erlbaum Associates, Pp. 861-888.
- Estes, W. K., Allmayer, D. H., & Reder, S. 1976 Serial position functions for letter identification at brief and extended exposure durations. *Perception & Psychophysics*, 19, 1-15.
- Fang, S. & Wu, P. 1989 Illusory conjunctions in perception of Chinese characters. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 434-447.
- Farah, M. J. 1990 *Visual Agnosia*. Cambridge, MA: MIT Press/Bradford books.

- Fujita, I., Tanaka, K., Ito, M., & Cheng, K. 1992 Columns for visual features of objects in monkey inferotemporal cortex. *Nature*, 360, 343-346.
- Friedman, R. B. & Alexander, M. P. 1984 Pictures, images, and pure alexia: A case study. *Cognitive Neuropsychology*, 1, 9-23
- Gilmore, G. C., Hersh, H., Caramazza, A., & Griffin, J. 1979 Multidimensional letter similarity derived from recognition errors. *Perception & Psychophysics*, 25, 425-431.
- Hancock, P. J. B., Baddeley, R. J., Smith, & L. S., 1992 The principal components of natural images. *Network*, 3, 61-70.
- ホフスタッター, ダグラス 1985 ゲーデル, エッシャー, バッハ —あるいは不思議の環. 白揚社
- Holland, J. H. 1992 Genetic algorithms. *Scientific American*, 267, 44-50.
- Hubel, D. H. 1988 *Eye, Brain, and Vision*. Scientific American Library, A division of HPHLP, New York.
- Hubel, D. H. & Wiesel, T. N. 1977 Functional architecture of macaque monkey visual cortex. *Proceedings of the Royal Society of London, B*, 198, 1-59.
- 乾敏郎 1993 Q & Aでわかる脳と視覚 人間からロボットまで. サイエンス社
- 伊藤克浩 1989 単語認知における文字の移動. 日本心理学会第 53 回発表論文集, 626.
- 国立国語研究所 1983 高校教科書の語彙調査. 秀英出版
- 小松左京 1987 虚無回廊. 徳間書店
- Luce, R. D. 1963 Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*, 1. New York: Wiley.
- 牧岡省吾 1989 漢字知覚における migration 現象. 京都大学文学部卒業論文 (未公開)
- 牧岡省吾 1994 単語の視覚的認知過程における文字位置情報の表現. 心理学研究, vol. 65, 224-232.
- 牧岡省吾, 乾敏郎, 山下博志 1993 文字パターンの心理空間と脳内表現. 電子情報通信学会技術報告, 93, 33-40.
- 松沢哲郎 1991 チンパンジーから見た世界. 東京大学出版会
- McClelland, J. L. 1986 The programmable blackboard model of reading. In McClelland, J. L., Rumelhart, D. E., & the PDP research group (Eds), *Parallel Distributed Processing*, vol 2. Cambridge, MA: MIT Press/Bradford books, Pp. 122-169.

- McClelland, J. L. & Rumelhart, D. E. 1981 An Interactive activation model of context effect in letter perception: Part 1. On account of basic findings. *Psychological Review*, 88, 375-407.
- McClelland, J. L. & Mozer, M.C. 1986 Perceptual Interactions in two-word displays : familiarity and similarity effects. *Journal of Experimental Psychology : Human Perception and Performance*, 12, 18-35.
- Miyashita, Y., Date, A., & Okuno, H. 1993 Configurational encoding of complex visual forms by single neurons of monkey temporal cortex. *Nuropsychologia* , 31, 1119-1131.
- 森崎礼子, 乾敏郎 1995 類似性判断における顔の表現. 電子情報通信学会技術報告, PRU95-87, HIP95-14 (1995-07).
- Mozer, M. C. 1983 Letter migration in word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 9, 531-546.
- Mozer, M. C. 1989 Types and tokens in visual letter perception. *Journal of Experimental Psychology : Human Perception and Performance*, 15, 287-303.
- Mozer, M. C. 1991 *The Perception of Multiple Objects*. Cambridge, MA: MIT Press/Bradford books.
- Nelson, T.O. 1987 The Goodman-Kruskal gamma coefficient as an alternative to signal-detection theory's measures of absolute-judgment accuracy. E. E., Roskam & R., Suck (Eds.) *Progress in Mathematical Psychology*, vol. 1., Amsterdam: Elsevier Science Publishers B.V. (North-Holland), 299-306.
- 旺文社 1973 国語辞典. 旺文社
- Oja, E. 1982 A simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology* , 15, 267-273.
- Perret, D. I., Harries, M. H., Bevan, R., Thomas, S., Benson, P. J., Mistlin, A. J., Chitty, A. J., Hietanen, J. K., & Ortega, J. E. 1989 Frameworks of analysis for the neural representation of animate objects and actions. *Journal of Experimental Biology* , 146, 87-113.
- Poggio, T. & Edelman, S. 1990 A network that learns to recognize three-dimensional objects. *Nature*, 343, 263-266.
- Reicher, G. M. 1969 Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, 81, 274-280.
- Rumelhart, D. E. 1971 A multicomponent theory of confusions among briefly exposed alphabetic characters. *Tech. Rep. No. 22 Center of Human Information Processing* . La Jolla, CA.

- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. 1986 Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.) *Parallel distributed processing*, 1. Cambridge, MA: MIT press.
- Rumelhart, D. E. & McClelland, J. L. 1982 An Interactive activation model of context effect in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the models. *Psychological Review*, 89, 60-84.
- Rolls, E. T. 1994 Brain mechanism involved in perception and memory, and their relation to consciousness. *Proceedings of International Symposium "Computation, Cognition, and Consciousness"*, Kyoto, Japan: International Institute for Advanced Studies.
- SAS institute Inc., 1992 SAS/STAT software: changes and enhancements, release 6.07, in *SAS technical report P-229*. Cary, NC.
- Shallice, T. & McGill, J. 1978 The origins of mixed errors. In J. Requin (Ed.), *Attention and Performance Vol. 7*, Hillsdale, NJ: Lawrence Erlbaum Associates, Pp. 193-208.
- St. John, M. F. & McClelland, J. L. 1990 Learning and applying contextual constraints in sentence comprehension, *Artificial Intelligence*, 46, 217-257.
- 杉山高一 1983 多変量データ解析入門. 朝倉書店
- 田中啓治 1994 対象認知の脳内メカニズム. 甘利俊一, 酒田英夫 (編) 脳とニューラルネット, 朝倉書店, Pp. 49-74.
- Treisman, A. & Gelade, G. 1980 A feature integration theory of attention. *Cognitive Psychology*, 12, 97-136.
- Treisman, A. & Schmit, H. 1982 Illusory conjunctions in the perception of objects. *Cognitive Psychology*, 14, 107-141.
- Ungerleider, L. G., & Mishkin, M. 1982 Two cortical pathways. In D. J. Ingle, M. A. Goodale, R. J. W. Mansfield (Eds.), *Analysis of Visual Behavior*, Cambridge, MA: MIT Press.
- Van der Velde, F. 1992 Effect of orthographic preactivation on letter migration. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 2, 449-459.
- Van der Velde, F., van der Heijden, A. H. C., & Schreuder, R. 1989 Context-dependent migrations in visual word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 133-141.
- Warrington, E. K. & Shallice, T. 1980 Word-form dyslexia. *Brain*, 10, 99-112.
- Widrow, B. & Hoff, M. E. 1960 Adaptive switching circuits. *Institute of Radio Engineers, Western Electronic Show and Convention, Convention Record, Part 4*, 96-104.
- 柳井晴夫, 高根芳雄 1985 新版多変量解析法. 朝倉書店

Young, M. P. & Yamane, S. 1992 Sparse population coding of faces in the inferotemporal cortex. *Science*, 256, 1327-1330.

1. バックプロパゲーション法による学習に使用したパラメータ

出力層のユニット j と中間層のユニット i の間の結合の変化量は、次式によって決定される。

$$\Delta w_{ji}(t) = -\varepsilon \frac{\partial E}{\partial w_{ji}}(t) + \alpha \Delta w_{ji}(t-1)$$

E は現在の出力パターンと教師パターンとの間の誤差を、 ε は学習の速さを、 α は結合強度を更新する際に前回の結合の変化量を加味する割合を示す。第 1 章のすべてのシミュレーションにおいて、 ε の値は .1、 α の値は .4 であった。

2. 複雑型細胞による入力パターン生成のアルゴリズム

単純型細胞に類似した受容野を持つ線分検出器でランダムドット・パターンの中の一定の面積を走査することによって、複雑型細胞の出力を決定した (図 A1)。線分検出器の受容野を図 A2 に示す。受容野の大きさは 7×7 ドットで、興奮性の領域の重み付けは 1.0、抑制性の領域の重み付けは -0.3 であった。線分検出器の出力は、もとのランダムドット・パターンとこれらの重みとの積和である。これらの線分検出器の中心を 3×3 ドットの範囲内で移動させ、それぞれの位置において出力を計算した。少なくとも 1 つの位置で出力が閾値を超えた場合には、対応する複雑型細胞の出力を 1 とした。すべての位置において出力が閾値を超えなかった場合には、対応する複

雑型細胞の出力を0とした。閾値の値は1.0であった。以上のような手続きによって、各ランダムドット・パターンに対する複雑型細胞の出力を算出し、その結果をネットワークへの入力パターンとして使用した。

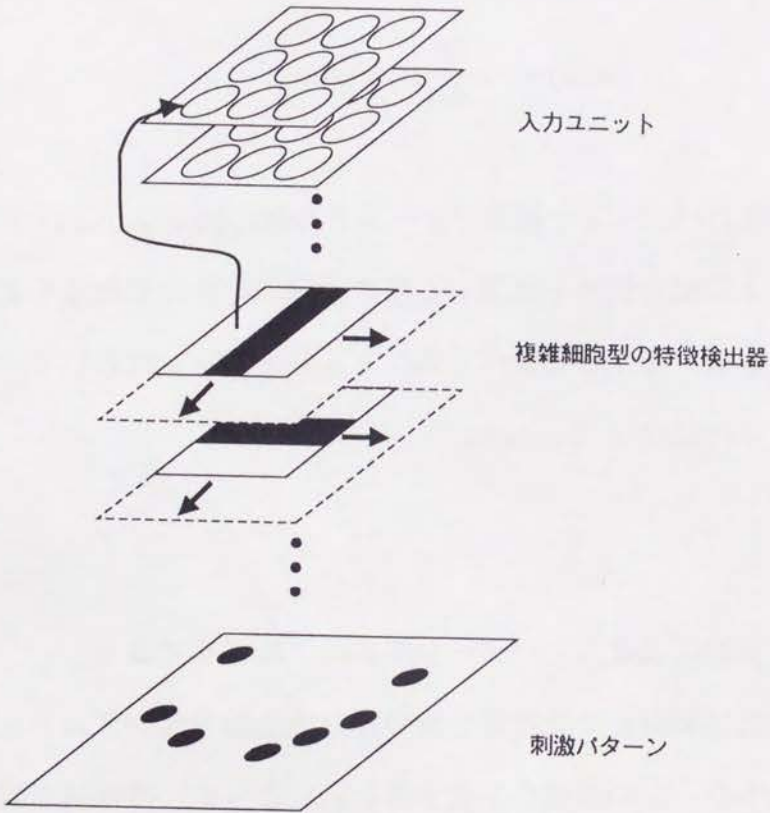


図 A1 複雑型細胞の出力を決めるための手続き。

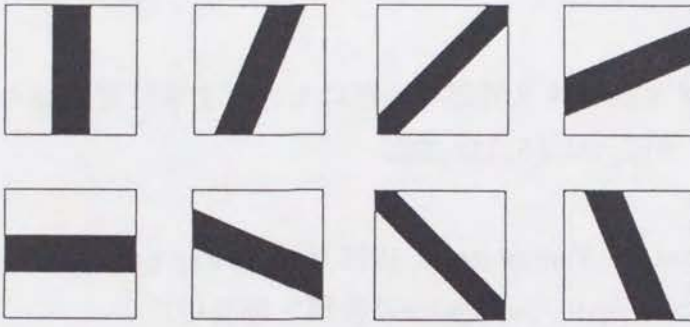


図 A2 線分検出器の受容野. 黒の領域は正の重み付け (1.0) を表し, 白の領域は負の重み付け (-0.3) を表す. 重み付けの値は, 入力パターンが一様である場合に出力が 0 になるように定められた.

3. Widrow & Hoff (1960) の最小二乗規則 (LMS rule)

Mozer (1991) の使用した学習規則は, Widrow & Hoff (1960) の最小二乗規則にもとづくものである. 第 5 層の特徴検出器 u から第 6 層 (出力層) の文字クラスターユニット v へ向かう結合 w_{uv} の変化の仕方は, 以下のような式で記述される.

$$\Delta w_{uv} = \eta b_v^{6*} (1 - b_v^{6*}) (d_v - b_v^6) b_u^5$$

$$\text{ただし, } b_v^{6*} = \begin{cases} .1 & \text{if } b_v^6 < .1 \\ b_v^6 & \text{if } .1 \leq b_v^6 \leq .9 \\ .9 & \text{if } b_v^6 > .9 \end{cases}$$

b_v^6 は第 6 層の文字クラスターユニット v の活性化度を, d_v は望ましい活性化パターン (教師パターン) を, b_u^5 は第 5 層の特徴検出器 u の活性化度を表す. η は学習の速度を決定するパラメータ (学習係数) である.

本研究に関連する研究業績

<論文>

1. 牧岡省吾 1994 単語の視覚的認知過程における文字位置情報の表現. 心理学研究, vol. 65, 224-232.
2. Makioka, S., Inui, T., Yamashita, H. 1995 Internal Representation of 2-dimensional Shape. (*perception*に投稿・審査中)

<著書>

1. 牧岡省吾 1995 単語認知過程のコネクショニスト・モデル. 荻阪直行 (編) 読みの情報処理, 朝倉書店 (収録予定)

<学会発表>

1. 牧岡省吾 1990 単語認知におけるmigration error -文字位置情報の処理に関して-. 日本心理学会第54回発表論文集, 577
2. 牧岡省吾 1991 単語認知過程における文字位置情報の符号化様式. 日本心理学会第55回発表論文集, 189
3. 牧岡省吾 1992 日本語の視覚的認知における文字位置情報の表現. 日本心理学会第56回発表論文集, 872
4. 牧岡省吾, 乾敏郎, 山下博志 1993 文字パターンの心理空間と脳内表現. 電子情報通信学会技術報告, 93, 33-40.
5. 牧岡省吾, 乾敏郎, 山下博志 1993 ニューラルネットワークを用いた文字の脳内表現の検討. 日本心理学会第57回発表論文集, 512
6. 牧岡省吾, 乾敏郎, 山下博志 1994 2次元パターンの脳内表現. 日本心理学会第58回発表論文集, 660

謝辞

本論文の作成にあたりご指導をいただきました京都大学文学部の乾敏郎先生と山下博志先生に深く感謝いたします。また、研究をご指導下さいました京都大学文学部の故平野俊二先生，清水御代明先生，苧阪直行先生に感謝いたします。

さらに，私が京都大学在学中に貴重な助言をいただいた，心理学研究室の研修員・研究員・大学院生・学部生（当時）のみなさんに感謝いたします。そして，予備実験も含めた数多くの実験に被験者として協力していただいた学部生（当時）のみなさんに感謝いたします。