脳の科学 第 11 回

担当：浅川伸一
brain_science@cis.twcu.ac.jp

師走 19 日, 2008 年

2

ニューラルネットワークから見える言語の風景 (「言語」に 2002 年に書いた原稿より)

第1章　ニューラルネットワークによる脳の障害の説明

ニューラルネットワークから見た言語の風景ということで，これから 6 回にわたって言語の障害とニューラルネットワークモデルの関連を紹介する．

第 1 回目の今回は，次号以降の話題を理解するための基礎的な用語の解説をし，脳の障害とニューラルネットワークの研究に関する「局在性仮説」批判などの話題を取り上げる．ニューラルネットワークの心理学的応用に興味のある方は文献 [28] などを参照していただきたい．

1.1　脳の構成論的研究

近年，脳はさまざまな方法で研究されている．fMRI に代表される機能的脳画像研究，ネコやサルの脳にマイクロ電極を差し込んで細砲の動作を測定する電気生理学的手法，動物を用いた脳の破壊実験，脳波，薬理学的方法，神経心理学と呼ばれる障害を持った脳の観測，心理学実験などである．これらの方法に加えて脳の構成論的研究，すなわち，ニューラルネットワークと呼ばれる脳のモデルを作って，このモデルが実際の脳と同じ機能を果たしていると考えられるのならば，モデルの持っている機構が脳にも存在する可能性があるとする研究分野がある．

脳の構成論的研究とは，このようなモデルを作ってコンピュータによるシミュレーションを通して脳の機能を類推してゆく研究である．脳の構成論的研究で重要なことは，生理学的に分かっていることはできるだけモデルに取り入れ，分かっていないことについては大胆に仮定してモデル化を行なうということである．

本稿では人間の脳に特徴的である言語機能を脳はどのようにして実現しているのかをコンピュータシミュレーションと実際の言語障害の事例とを比較しながら考察してみたい．

1.2　ニューラルネットワークの特徴

ニューラルネットワーク—あるいはコネクショニストモデル，並列分散処理 (PDP) モデルとも呼ばれる— とは脳の神経細胞 (ニューロン) の動作を抽象化した表現であるユニットの集合を用いて人間の情報処理のモデル化をめざす研究分野である．脳内では莫大な数のニューロンが互いに密接に結合されており，ニューロンのネットワークを構成している．これがニューラルネットワークである．

ニューロンの動作を抽象化して考えれば比較的簡単な図式で理解できる．ニューロンとニューロンとの間はシナプス結合と呼ばれる結合によって結ばれている．実際のシナプスにはシナプス間隙と呼ばれるわずかな隙間があって，このシナプス間隙の間を神経伝達物質が放出 (入力を受け取る側から見れば吸収) されて情報伝達がなされる．シナプス結合には，この情報伝達の種類と効率に従ってシナプス結合強度が各シナプス毎に異なる．正の結合強度なら興奮性の，負の結合強度ならば抑制性の結合となる (結合が興奮性か抑制性かの違いは神経伝達物質

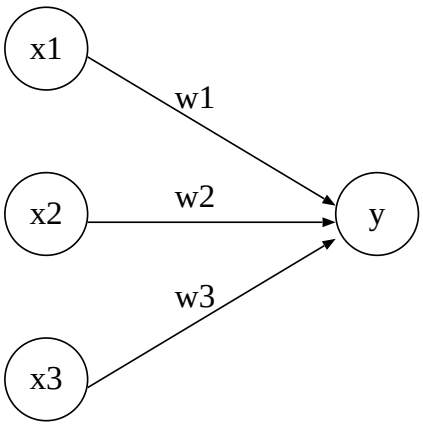

図 1.1: ニューロン y は x_1, x_2, x_3 からの入力を受けて，入力ニューロンからの活性値を結合係数 w によって重み付けられた値の関数として自身の出力値が定まる

の違いによる)．情報を受け取る側のポストシナプスニューロンは，この伝達強度で重み付けられた他のニューロンからの情報を時間的，空間的に加算して，この値が一定の強度に達するとニューロンの電位が急激に変化する．このときニューロンが活性化したとか興奮したとかいう．ニューロンが活性化する値のことをしきい値という．しきい値が高くなれば活性化しにくく，反対に低ければ活性化しやすくなる．このニューロンの活性値がシナプスを介して他のニューロンに伝達されることによって脳内の活動，ひいては言語などの高次認知機能が発現する．

もっとも簡単なマッカロックとピッツの形式ニューロンのモデルでは，活性化した状態を 1 で表し，そうでなければ 0 の値をとる 2 値関数，従って真か偽かを表す命題関数と同一視できるモデルである．従ってこの場合のしきい値は 0 である．そのニューロンが活性化するか否かは，出力を送るニューロンの活性値にそのシナプス結合強度を掛けた値の合計値が 0 より大きいか否かで決まるというのがマッカロックとピッツのニューロンモデルである．ニューラルネットワークとはこのような抽象化されたニューロンの表現 (ユニットと呼ばれることもある) を用いて人間の知的活動をモデル化する研究分野である．

人間の行なうさまざまな行為はすべてニューロンの活動とニューロン間の結合の強度として表現可能であると考えるのがニューラルネットワークである．ニューラルネットワークにおける特徴を挙げるとすれば，分散表現と統計的構造の漸進的学習，および相互作用の 3 点に要約できる [13, 20]．

分散表現:　ニューラルネットワークにおいては，知識はそれぞれのユニット集団の活性化パターンとして表現される．例えば，ある単語の意味は別の単語の意味とは異なる活性化パターンとして表現されており，類似した概念は互いに類似した活性化パターンとして表現される．

統計的構造の漸進的学習:　長期的な知識はユニット間の関係，すなわちユニット間の結合強度としてネットワーク内に埋め込まれている．ユニット間の結合強度は学習によって徐々に変化する．すなわち，学習により外界から提供される情報 (単語間の類似度や相互関係など) の統計的性質が徐々に獲得される．

相互作用:　ユニットは密接に連結されており，相互に影響しあう．すなわちユニット間の結合強度に応じて，互いに活性化パターンを強め合ったり，弱め合ったり，振動したりと

いうような複雑な相互作用をする．

1.3 ニューラルネットワークの損傷の解釈について

ニューラルネットワークを破壊することで言語を含むさまざまな認知機能の障害をコンピュータ上に再現できる．人間の認知機能とニューラルネットワークプログラムとを同一視し，かつ，ニューラルネットワークを部分的に破壊することと脳損傷を同一視することとによって，近年の認知神経心理学は大きく変化してきた．

ニューラルネットワークを用いた脳損傷のシミュレーション研究では特定の認知機能を遂行するためのニューラルネットワークモデルをコンピュータプログラムとして実現し，構築されたニューラルネットワークの一部を破壊することによって対応する部位が損傷を受けたときに生ずる症状をプログラムの出力として表現することをめざしている．こうした研究はいわば人工脳損傷とも言える研究である．人間の言語活動を理解する上でも，あるいは実際の脳損傷患者の症状を理解するためにも，コンピュータを用いた人工脳損傷研究は重要だと考えられる．倫理上の制約から実際の人間の脳を破壊して実験を行なうことは不可能だからである．ニューラルネットワークによる人工脳損傷研究は，実際の脳損傷患者を扱う神経心理学に対して強力な道具を提供していると言える．

ニューラルネットワークは相互作用をする複雑なシステムとしての行動と，そのシステムが損傷したときの効果との関係を推論する手段を提供している．そのような推論が明示的で機能論的に検証できる，すなわちシミュレーションによる検証が可能になったことが重要なのである．

このときモデル化した現象が，脳内で対応する機能が存在するのかという疑問がある．脳の構成論的研究者の中には，情報論的必然性という概念を用いてこの分野の研究に理論的根拠を主張する人達もいる．情報論的必然性とは入出力関係が脳とモデルとの間で一致したとき，そのメカニズムも一致している可能性が高いというものである．入出力関係が複雑で巧妙であるほど，それを実現する情報処理の機構はそういくつもあるはずが無いという主張である．これにはさまざまな異論反論があるだろう．筆者もこの情報論的必然性に完全に同意しているわけではない．むしろ脳の構成論的研究の意義は，今までの神経心理学的証拠から導き出された脳観に疑問を投げかけ，新しい脳観を求めていることではないかと思う．今回はそうした従来のモデルへの批判のひとつ「局在性仮説」への批判をとりあげてみたい．

1.4 神経心理学における局在性仮説への批判

神経心理学の分野では機能局在性仮説 (locality assumption) が重要な役割を果たしてきた．機能局性在仮説とは，心の機能が独立した下位モジュールによって構成されているとする考え方である [14]．そしてこのモジュール間では，比較的簡単な情報を伝達するにすぎない．各モジュールは情報論的にカプセル化 (encapsulated) されていて，あるモジュールの損傷が別のモジュールの機能に影響を与えることはほとんどなく，あるモジュールの損傷は比較的単純なひとつの認知機能の低下として表出する，というのが局在性仮説である．

だが，この局在性仮説は単純すぎる [13]．特定の認知機能の検査結果とその認知システムの障害の部位を特定することの関連は，それ程明らかな関係にはない．

アービッブ [2] は次のように書いている．ニューラルネットワークの破壊実験，切除実験は重要な情報を与えてくれるが，その解釈には注意を要する．「ラジオから抵抗を取り外したらピーッという音が鳴ったからといって，その抵抗がピーの抑制中枢であるとはいえない．ほと

んどの人は噂話が好きである，グループはある人がそのグループを離れると，その人の噂話をする．このときその人をそのグループの噂話抑制中枢とみなすようなものである.」

具体例を挙げてこのことを考察してみよう.

1.4.1　システムの等価性

下図のニューロン n_1 とニューロン n_2 はいずれもしきい値が 0 であるとし，入力がそのまま出力となるとしよう．ニューロン n_3 はニューロン n_1 や n_2 とは全く無関係にこの回路に信号を送っているとする．ニューロン n_4 はしきい値が 1.5 であるので，二つ以上の入力があったときに活性化する．したがって x_3 に入力がなければ，この回路全体の出力は x_1 と x_2 とが同時に入力されたときだけ活性化する．他方，もし x_3 からの入力があれば，この回路の出力は，x_1 または x_2 からの入力があれば活性化する．すなわち n_3 は，x_1 と x_2 とが AND 回路 (論理積) として動作するか OR 回路 (論理和) として動作するかを決定する制御ニューロンと見なすことができる．このシステムを実験者が外から見るとき，x_1 と x_2 はシステムへ

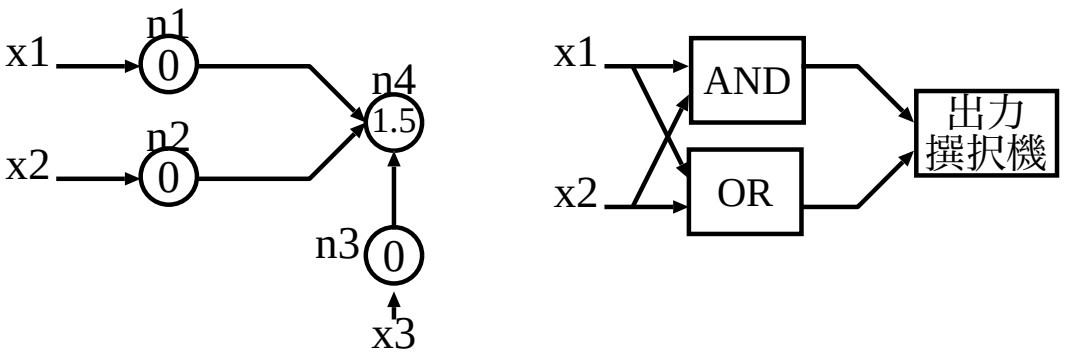

図 1.2: 同じ機能でも内部構造が異なる 2 つのシステム．入出力は両者とも同じなので，損傷実験をするか内部変数をモニターしないかぎり分離できない

の入力ニューロンであることが分かるが，どこか別のサブシステムから来ている AND と OR を計算する別々のサブシステムがあり，そのどちらかをシステムの出力として選択する出力選択機があると思って，右のようなブロック回路を描くことも可能である．すなわち外部から観察している限りこの二つのシステムは分離不可能である．入出力は両者とも同じなので，損傷実験をするか内部変数をモニターしないかぎり分離できない．神経心理学は，こうした破壊を取り扱う研究分野であり，神経心理学的事実を考慮した構成論的研究であるニューラルネットワークは脳のメカニズムに迫る有力な研究手法であるということができるだろう.

1.4.2　システムの構造と解釈

一つのシステムの行動が完全にわかったとしても，そのことから構造が一義的に決まるわけではない．システム S を分析して，それが行動的に二つのシステム S_1 と S_2 の結合とみなすことができ，またそのように分解するとシステム S がよく説明できるにしても，それをもってシステム S の構造を S_1 と S_2 の二つのシステムに機械的に分解することはできない．もし心をいくつかの心的過程に分解できたとしても，その心的過程を脳の別々の部位に帰する

ことはできないし，その逆も成り立たない．観察された行動のもとにある内部交互作用の詳細を説明する前に，内部構造と状態についての解剖学と生理学に留意しなければならない．

ある機能がないからと言って，その脳部位が無活動であるとは言えない．何かの異常によってニューロン n_3 が発火し続けるとニューロン n_4 はいつも OR 回路になり，AND 回路にはならない．したがってある機能が働かないということは，その神経回路が働かないということではなくて，ある仕方で働かないということなのである．すべてのニューロンが活動し，大筋においてすべての神経路が適切に結合していても，小さな結合やしきい値に異常があると異常行動となる．ラジオの「ピー音抑制」の例を考えればわかりやすいだろう．脳構造の意味づけは慎重に行なう必要がある．

1.5 まとめ

上記のようにシステムの評価，とりわけ破壊の効果をそのシステムの機能や性能と結びつけるには慎重な調査，観察が必要である．だが，驚くべきことにこのような認識が産まれたのは比較的最近になってからのことである．神経心理学は，言語野の存在を示したブローカ (Paul Broca, 1824-1880) 以来 100 年以上の歴史を持ち，厖大な数の観測データがあるにもかかわらず，ここに示したようなシステムの解釈に関する考察が行なわれるようになったのはここ十数年でしかない．これから多くのことがなされなければならないし、実り多き研究が盛んになることを望んでいる．

本稿は 6 回にわたってニューラルネットワークを用いた言語に関する話題を取り上げて解説するものである．ここでの試みは 100 年の伝統を持つ神経心理学的証拠—たとえば失語症のような—から帰納される言語観について脳の構成論的研究であるニューラルネットワークという視点で眺めてみることにする．ニューラルネットワーク研究者の視点から見た言語情報処理とは何か．次回は言語学者の間でも有名になったエルマン (Jeffrey Elman) のネットワークを紹介する．

第2章　エルマンネットの衝撃

エルマン (Elman)[8] の考案した単純再帰型ネットワーク (通称エルマンネット) によって文章の処理が可能である。このことは言語学者にとってインパクトの強い研究であった。今回はエルマンの考案したニューラルネットワークモデルを紹介し、最後に生成文法理論との対比について考察してみたい。

エルマンネットではいわゆる「刺激の貧困」「否定証拠の欠如」でも文章理解が可能である点が重要である。エルマンのネットワークでは明示的な教師信号による誤り訂正機構を仮定しないからである。また、言語の記号処理モデルで用いられるような書き換え規則や木構造の構文木を必要とせず文法構造に関する知識が創発する点も強調される。エルマンの示した系列学習の枠組みでの言語獲得とは、子どもが (大人による) 言語環境に曝されることから引き起こされる学習の結果であると主張される。

2.1　エルマンネットにできること

エルマンネットでは、入力層は入力信号を処理する入力層ユニットと、直前の時刻までの中間層の状態を入力とする文脈層ユニットとで構成される (図 2.1)。文脈ユニットは以前の中間層をコピーするためだけ (すなわち中間層から文脈層ユニットへの結合強度は 1.0) である。結合強度の学習は順方向の結合についてだけ行なわれるので、通常の誤差逆伝播法がそのまま適用できる。

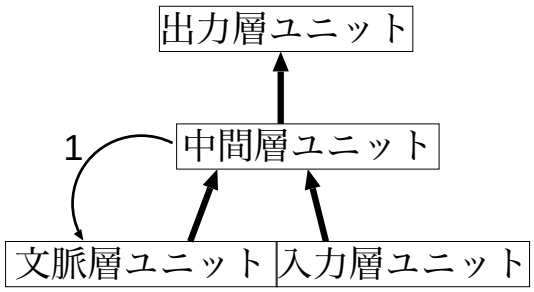

図 2.1: エルマンネット

ある時刻 t で処理される内容は、その時点での入力信号と、それ以前の時刻 $t-1$ までで処理された回路の状態を表す信号とが同時に処理される。すなわち、文脈層は $t-1$ 時刻までの過去の状態を記憶していることを意味する。この結果、ある時刻 t でのネットワークの状態は現在の入力と過去の入力履歴の集合によって決まることになる。

2.2　単語予測課題

ここでは有名な単語予測課題のシミュレーションを紹介しよう。

エルマン [9] は、自身の考案したエルマンネットを用いて文法学習などの複雑な構造を表現できることを示した。文章を構成する単語を逐次入力層に与え、ネットワークは次の単語を予測するように訓練される。この訓練手続きを系列予測課題 (または単語予測課題) という。エルマンは、系列予測課題によって次の単語を予想することを繰り返し学習させた結果、文法構造がネットワークの結合係数として学習されることを示した。エルマンネットによって、埋め込み文の処理、時制の一致、性や数の一致、長距離依存などを正しく予測できることが示されている [8, 9, 10]。

表 2.1: エルマンの用いた文法規則

S	→	NP VP “.”
NP	→	PropN | N | N RC
VP	→	V (NP)
RC	→	who NP VP | who VP (NP)
N	→	boy | girl | cat | dog | boys | girls | cats |dogs
PropN	→	John | Mary
V	→	chase | feed | see | hear | walk | live | chases| feeds | seeds | hears | walks | lives

これらの規則にはさらに 2 つの制約がある。(1) N と V の数が一致していなければならない。(2) 目的語を取る動詞に制限がある。例えば hit, feed は直接目的語が必ず必要であり、see と hear は目的語をとってもとらなくても良い。walk と live では目的語は不要である。

表 2.1 にエルマンが用いた文章生成則を示した。文章は 23 個の項目から構成されている。8 個の名詞と 12 個の動詞、関係代名詞 who、及び文の終端を表すピリオドである。入力層においては一ビットが一単語に対応するように単語の数だけユニットが用意された出力層のユニットも一ユニットが一単語を表すように入力層と同じ数だけのユニットが用意された。中間層は 70 個のユニットが用意された。エルマンネットの特徴である文脈層ユニットは中間層のユニット数と同数の 70 個である。

エルマンは表 2.1 に従って生成された文章を一単語ずつ次々にネットワークに示し、次に来る単語を予測させる訓練を行なった。すなわち入力層にある単語を提示し、出力層における教師信号として次に来る単語を与えたのである。

訓練の結果、ネットワークは次に来る単語の種類を予測できるようになった。例えば boy が提示されるとネットワークは次に来る単語として、関係代名詞 who もしくは単数を主語とする動詞 feeds, seeds, hears, walks, lines を表わすユニットがほぼ等確率で活性化され、複数形を主語とする s の付かない動詞や他の名詞を示すユニットは全く活性化されなかった。反対に、複数名詞である boys が提示されると who, chase, feed, see, hear, walk, live が等確率で活性化された。ネットワークに boys who Mary chases まで提示されると文頭の主語 boys が複数であるために複数形を主語とする動詞が等しく活性化された。このようにエルマンのネットワークは中央埋め込み文のある、いわゆる長距離依存を正しく予測できたのである。

ここで大切なことは、エルマンのネットワークでは文法知識はネットワークの結合係数の大きさとして表象されていることである。明示的な書き換え規則のようなルールは全く与えられていない。さらに、関係代名詞による文章の再帰的構造は中間層の活性値で表現される状態空間の中に表現されていることである。そして、この文法知識は否定的な証拠を提示されることによって獲得されるのではなく、単純に次の単語を予測するだけしか行なっていない点も強調される。

2.3 小さく始めることは本当に重要なのか

エルマンの主張には、さらに２点ほど重要な点がある。それらは、「小さく始めることの重要性」と「言語獲得期における記憶容量の制限」と呼ばれる。エルマンの主張によれば、言語獲得期の幼児における記憶容量の制限は言語獲得に対して否定的な要因としてではなく、むしろ記憶容量が制限されている結果として、複雑な文章を処理しないで済むことで言語獲得が可能になるという。文法学習では記憶容量を制限することがむしろ有利に働くと主張している。現生人類が他の種と異なる特徴は、長い成育期間と顕著な学習能力である。進化の過程において、成体に達するまでの発育期間が長いことは自然淘汰から見て不利なはずである。にもかからず人類が滅びずにここまで文明社会を発展させたのは、まさにこの学習能力によるものであり、幼児から大人へと成長する過程で記憶容量が徐々に増加することが、我々ホモサピエンスにとって決定的に重要だったというのである。ニューラルネットワークによるシミュレーションから進化の問題を論じてしまう破天荒なところが、良くも悪くもエルマンのすごいところでもあるのだが。

これら「小さく始めることの重要性」と「言語獲得期における記憶容量の制限」と呼ばれる２点については否定的な証拠も提出されていることに言及しておきたい。この２点を仮定せずとも言語入力にある種の意味構造を仮定することでエルマンネットの言語獲得能力が劇的に向上することが示されているからである [22]。ロードとプラウトによれば言語獲得には小さく始めることが重要なのではなく、ソフトな意味論的制約 — 例えば犬は猫を追いかけるが、猫が犬を追いかけることはほとんどない— を付加することで最初から複雑な構文を与えても学習が可能であることが示されている。エルマンの訓練したネットワークでは、「犬が猫を追いかける」と「猫が犬を追いかける」とが等確率で訓練文に含まれていた。さらにエルマンが作った訓練文には、「少年が追いかけた少年が追いかけた少年が歩いた」などというような構文的には正しくても実際にはほとんど用いられることのない文章が含まれていた。ロードとプラウトはこのような点を改善した文章 —彼らの用語ではソフトな制約という —を用いて訓練することにより、小さく始める必要は必ずしも必要ではないことを示した。ロードとプラウトの研究によれば第二言語獲得が難しいのは第一言語である母語の獲得の必然的結果であるとされる。二つの言語を最初から同時に学習する条件のエルマンネットは、単一言語を学習する条件のネットワークと比べて僅かに学習が成立するのが遅れるが、単一言語条件とほぼ同時期に二つの言語を獲得することが可能であった。一方、単一言語を習得したエルマンネットに対して第二言語を習得させた場合学習が進行し難いことが示された。このことはバイリンガルの成立に関する常識的な見解とも合致していると思われる。

2.4 生成文法理論と統計的構造学習モデル

チョムスキーの生成文法理論においては、言語獲得には普遍的で言語固有の生得性が要求される。生成文法を前堤とした言語習得理論は、連続仮説に基づき大人の文法と同じ強力な装置 (例えば統語範疇、句構造規則) が幼児の文法にも存在すると仮定するため、幼児の発話に表われる意味的、形態的な制限を説明するために、様々なアドホックな原則に訴えざるを得なかった。しかし、1990 年代に入ってニューラルネットワークの分野で開発されたモデルにおいては、このような生得性を仮定せずとも言語知識が学習によって創発し、記号処理的な書き換え規則を仮定せずとも統語規則を学習しうると主張されている [9, 11]。さらに、最近ではこの考え方を先鋭化させ、言語獲得とは言語の持つ多様な統計的確率的性質を学習することであるというアイデアに発展してきている [24, 25]。このような立場を取る研究は、統計的 (または確率的) 構造学習モデルあるいは多重制約充足仮説と呼ばれる。

言語の持つ統計的な性質を獲得することが重要であるという多重制約充足仮説のアイデアは、単純なマルコフ連鎖だけを用いた確率的言語モデルでは文法の問題は説明できないとしてチョムスキーの生成文法理論においては長い間無視されてきた。例えば、チョムスキーが考案した文章 “Colorless green ideas sleep furiously” は、英語を母語とする聞き手には文法的に正しいと判断できるが意味をなさないことが了解されるが、統計的言語モデルでは文法判断ができないとされてきた。統計的構造学習モデルの枠組では、この文章でさえ、Property, Property, Things, Action, Manner という自然な英語の文法構造を反映しているということになる [1]。最近のニューラルネットワーク研究の動向を見ると、子どもは普遍文法の知識を持って生まれて来るという生成文法の仮説だけが言語獲得の諸事実を説明する仮説ではない [25] のかも知れない。

今回紹介したエルマン (Elman) の研究や統計的 (確率的) 構造学習モデルに代表されるニューラルネットワーク理論は、どのように言語知識が学習されて行くのかという問題や、言語能力と言語運用とを区別して考える必要がない、という生成文法理論では説明が難しかった問題に答えることができる。このことは理論上大きなアドバンテージを持つと言えるだろう [24, 25]。加えて「子どもは規則に違反する例文を提示されないのになぜ正しく文法を学習するのか」というベーカーのパラドックス (Baker’s paradox) をも矛盾なく説明できる。

上記のような統計的構造学習モデルの視点で言語獲得を考えれば、言語獲得における子どもの課題は、生成文法理論の主張する普遍文法におけるパラメータ設定問題ではなく、むしろ言語の使用そのもの、および背後にある言語の統計的 (確率的) 構造を学習することであると言えるだろう [24]。

言語が人間という種に特異的で領域固有であるという言語学者の主張も近年疑われ始めている。他の種は確かに我々人間のような言語を持っていないが、同時に我々人間のようにバイオリンを引いたりゴルフを楽しんだりしない。言語が種に特異的で領域固有であるのならバイオリンやゴルフも種に特異的で領域固有の知識だということになってしまうからである。

第3章　失読症モデルの変遷

言語を司る脳の一部が障害を受けると言語行為に影響が表われる。そのなかで、今回は失読症を取り上げ、ニューラルネットワークによる説明モデルとその論争を紹介する。最初に 三種類の失読症を概説し、失読症を説明するモデルである二重経路モデルとトライアングルモデルとの間の論争を紹介したい。そして、エキスパート混合ネットワーク [15, 16] というモデルの観点から見れば、両者は統一的に解釈できることを示そう。

3.1　三つの失読症

大きく分けて読みの障害には三種類ある。音韻性失読症、表層失読症、深層失読症である。音韻性失読症の患者は実在する単語は読むことができるが、非単語を読むことができない。例えば must は読むことができても、実際には存在しないが発音可能な単語 nust を発音することができない。表層失読の患者は、規則語や非単語を読むことができるが、低頻度の例外語を読むことができない (例えば yacht)。表層失読の患者は視覚性の誤り (dog を dot と言ったりする) もするが意味の誤りはない。深層失読の患者は音韻失読の患者と同じく非単語を読むことができない。加えて、深層失読の患者は意味性の錯読を示す。例えば dog を cat と言ったりする。

3.2　二重経路モデルによる読みの障害の説明

コルトハートら [3, 4] によって開発された記号処理的な読みのモデルである二重経路モデルでは、印刷文字を音韻へ変換するための明示的な規則に基づく直接経路と、規則にあてはまらない単語を読むためのルックアップテーブルをもつ間接経路とから構成されている。印刷された文字を読むときには、直接文字を音韻へと変換する直接経路と、例外語の語彙テーブルまたは意味を介して発話に至る間接的な経路との２つの経路を仮定するのが二重経路モデルである。

心理学における読みのモデルの論争とは、直接経路と間接経路の処理の違いに関してである。二重経路モデルの特徴は記号処理的なルックアップテーブルを用いることである。すなわち二重経路モデルでは２つの経路間のスイッチが仮定されている。

3.3　トライアングルモデルによる読みの障害の説明

これに対して、トライアングルモデル [21] では同時的、相互作用的処理が仮定される。書記素、音韻、意味の情報は各ユニット群内/群間で分散表現表現されており、類似した単語は、類似した活性パターンとして表現される。

トライアングルモデルにおける直接経路では、多くの単語と発音規則が一致する規則語と高頻度の不規則語が処理される。一方、低頻度の不規則語は意味系に依存すると仮定される。

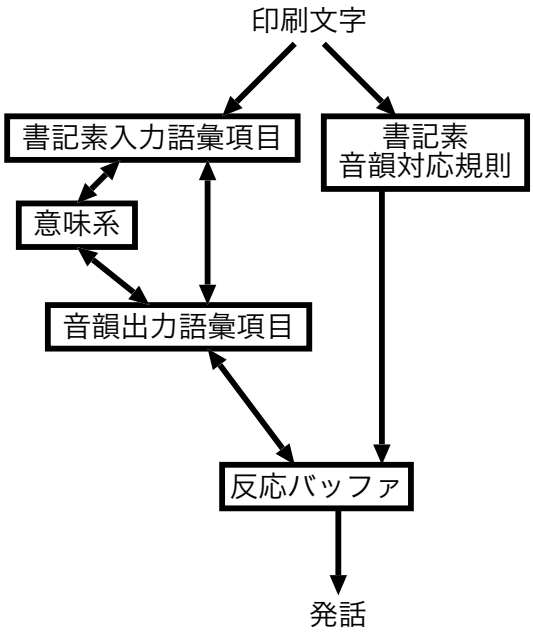

図 3.1: 二重経路モデル

規則語および高頻度例外語と低頻度例外語との処理の違いには労働の分割と呼ばれる作用が関与する。

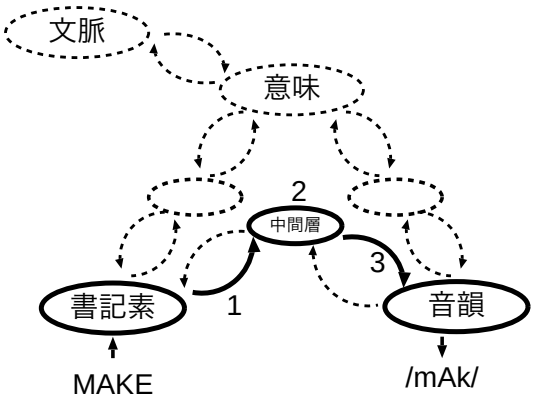

図 3.2: トライアングルモデル

トライアングルモデルでは、図 3.2 の実線で描かれた部分を実装し、図中の数字が描かれている部分を破壊することで音韻失読の症状が再現された。特に低頻度の例外語を規則化して発音する誤りが見られた。すなわち、音韻失読は直接経路への損傷の結果生じると見なすことができ、モデルの出力結果は失読症患者の検査結果とも一致していた。

トライアングルモデルでは入力された文字は意味層を介する間接経路と音韻層に直接出力を送る直接経路との両方の影響を受けるとされる。ある単語がどちらの経路をたどって読まれるかは、単語毎に異なると仮定される。直接経路では規則語と高頻度例外語とが処理され、低頻度例外語は間接経路である意味層のサポートを必要とする。このように単語毎に二つの経路の影響が異なって表現されることを労働の分割問題という。

表層失読は、この労働の分割問題によって説明される。労働の分割によれば、直接経路に損傷がある場合、間接経路を経由した読みは労働の分割の程度によって不完全な読みが生じるからである。音韻性失読と表層失読という二重に乖離した失読症状を同一機構で説明することに成功したことがトライアングルモデルの特徴である。一方、二重経路モデルによる表層失読の説明では、なぜ表層失読患者に視覚性の誤りが生じるのかが不明確である。二重経路モデルにおいては、表層失読は直接経路の障害に加えて間接経路も障害されているという説明になり、複数の認知機能が同時に損傷を受けたと仮定せざるを得ないからである。

3.4 エキスパート混合ネットワークによる統合

ただし、トライアングルモデルでは労働の分割問題を実装しているわけではない。トライアングルモデルでは書記素層から音韻層への直接経路だけがシミュレートされただけであり、モデルで説明できない部分を労働の分割と呼んでいるだけである。

ここで、あらためてニューラルネットワークモデルに単語の読みを学習させることの意味を考えてみよう。ニューラルネットワークに単語の読みを学習させるということは、書記素から音韻への変換規則を学習させることである。低頻度例外語にエラーが大きいのは、大部分の単語の読みに共通する書記素–音韻対応規則を学習し、その結果を適用しているからであって、ニューラルネットワークの見地からすれば正しい一般化と解釈することができる。すなわち、未学習のデータに対して、学習によって獲得した書記素–音韻対応規則を適用しているという意味である。むしろ、高頻度例外語は学習のしすぎ、すなわち過学習である。直接経路と間接経路との労働の分割問題は 2 つのネットワーク間の競合作用とみなすことができる。

もし、書記素–音韻対応規則を学習し、例外語と規則語を自動的に分類して学習できるアルゴリズムが存在すれば、労働の分割問題を解決できるモデルになる。この語彙の自動分類機構を実現するのがエキスパート混合ネットワーク Mixture of experts (以下 ME と略記) とよばれるニューラルネットワークモデルである。

ME とは入力データ空間をいくつかの小領域に分割し、その分割された各領域に対してひとつのニューラルネットワークを割り当てることによって、複雑な問題を解くための手法である [15, 16]。ME における学習とは、入力空間を分割し、分割された各小領域に属する入力に対する最適な答えを見つけ出すことである。このような手法を分割統治と言ったりする。分割統治 は科学における一般原理であるといって良い。この分割統治を自動的に行ない、学習させようと言うのが ME の発想である。

ME はエキスパートネットワークとゲーティングネットワークから成り立っている。ゲーティングネットワークは問題空間を分割するために用いられ、各エキスパートは分割された領域内での局所的な解を出力する。ME は階層的な問題空間の分割とエキスパートの割り当てが可能である。2 階層の場合の ME を図 3.3 に示した。

トライアングルモデルは直接経路と意味層を介した間接経路という 2 つのエキスパートネットワークを持つ ME とみなすことができる。そして、トライアングルモデルにおける労働の分割問題は ME におけるゲーティングネットワークによる領域の分割そのものである。各々の低頻度例外語ごとに限局された極限では、その単語のみに応答するルックアップテー ブルと同一視できる。すなわち、ME により、二重経路モデルにおけるルックアップテーブルも、トライアングルモデルにおける労働の分割問題も統一的に記述できる。この意味において、二重経路モデルとトライアングルモデルの間に本質的な違いはない。両者の間に存在する違いとは、入力空間の領域分割の大きさという量的な問題に帰結され、本質的には二重経路モデルとトライアングルモデルは同じ ME という、より一般的な範疇のニューラルネットワークモ

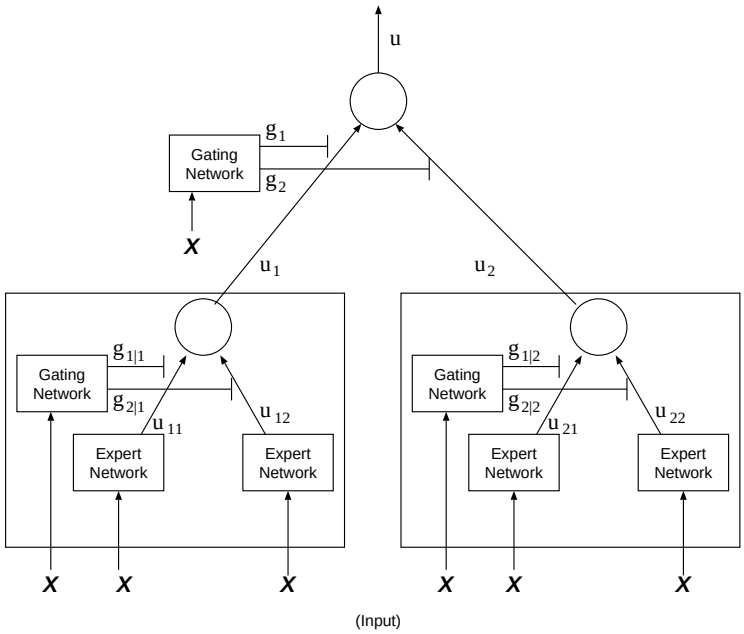

図 3.3: 2 段階のエキスパート混合ネットワーク。各エキスパートは単純なフィードフォワード型のネットワークであり、すべてのエキスパートは同じ入力を受け取り同じ数の出力ユニットを持っている。ゲーティングネットワークもフィードフォワード型のネットワークであり、エキスパートネットワークと同じ入力を受け取る。図中の g はゲーティングネットワークの出力 (確率) を表し、すべてのゲーティングネットワークの和は 1 となる。u はエキスパートの出力である。エキスパートネットワークの出力はゲーティングネットワークによって重み付けられた合成変量となる。

デルとして同一視できる。紙面の都合上具体的なデータを示す余裕がなくなってしまったが、実際 ME は驚くほどよく動作し、現象を説明できる。

3.5　新しい視点の重要性

二重経路モデルとトライアングルモデルにおける直接経路と間接経路との処理の違いに関する論争は、より一般的な読みのモデルである ME の一形態にすぎないということができる。二重経路モデルのおけるルックアップテーブルも、トライアングルモデルにおける労働の分割問題も、ME による領域分割として記述可能である。すなわち ME により、論争に結着をつけることができるのではないだろうか。この例は、科学における論争の結着方法を思い起こさせる。すなわち、論争の争点となっている問題は、新しい理論 —新しい世界観と呼んでもよい— によって統合された再解釈がなされるということが科学史においてしばしば起こっている。例えば、ガリレオがアリストテレスの物理学を覆し、アインシュタインの理論がニュートンのそれを特殊な場合として含んでいたように。同じようなことが二重経路とトライアングルモデルとの間の論争においても起こっていたと考えられる。このことは、伝統的な科学の知識体系に新しい世界観を導入し、既成事実を考え直すことがいかに大切かを教えてくれていると思えるのである。

第4章　意味記憶の構造試論

4.1　アルツハイマー症の物体呼称課題における成績低下

アルツハイマー症の初期症状の一つとして、物体呼称課題における障害が挙げられる。患者は椅子のことを尋ねられてテーブルと答えたり、梨のことを果物と言ったりする。このことは意味記憶の構造を考える上で興味深い。同一カテゴリーに属する別のメンバーである椅子とテーブルとを間違えることと、具体的な事物を表す梨をより上位概念である果物と答えることが起こっている。アルツハイマー症の意味記憶構造にはどのような障害を考えればよいのだろうか。

アルツハイマー症の患者の物体呼称課題における障害には、上記の意味記憶の障害説以外にも、二つの別の説明が存在する。一つめは、視覚刺激の質を低下させたときに、例えば写真を用いた物体呼称課題と線画を用いた課題で、線画を用いた方が呼称成績が悪いのである。すなわち、視覚性の障害だと解釈できる。二つめは語彙性の障害の可能性である。高頻度語の想起の方が低頻度語の想起よりも成績がよいという頻度効果が存在する。

アルツハイマー症患者の物体呼称課題における成績低下は、意味性、視覚性、語彙性のいずれの障害によって引き起こされるのであろうか。個々の患者ごとに障害の場所が異なると言う説明の仕方ももちろん可能ではある。しかし、ティペット (Tippett) とファラー (Farah) [26] によるニューラルネットワークを用いた研究によれば、意味記憶の障害のみによってアルツハイマー症の物体呼称課題における成績低下を説明できる可能性がある。ニューラルネットワークの特徴は知識の分散表現と相互作用であり、意味的な知識表象と視覚的および語彙的な知識表象は密接に関連しあっているのである。その結果、一つの構成要素、例えば意味記憶に障害があると、その障害の影響は、視覚性の知識にも語彙の知識にも影響を与える可能性がある。

4.1.1　モデル

彼女らのモデルを図 4.1 に示す。3つのユニット群、意味、名前、視覚がある。実際には各層の間に中間層が存在するのだが、ここでは中間層の存在は本質的ではないので省略した。例えば視覚ユニット層に入力が与えられると、各ユニットの活性化 (あるいは負の結合であれば抑制) は意味ユニット層に伝播する。意味ユニット層の活性化が、さらに視覚ユニット層と名前ユニット層の活性化に影響を与える。図中の矢印で示されているとおり各ユニット間の結合は層内、層間で双方向である。従って、あるユニットの活性は別のユニットの活性を引き起こし、全体として複雑な活性化パターンを示す。彼女らは視覚ユニット層 (あるいは名前ユニット層) に入力を与えたときに、意味ユニット層と名前ユニット層 (あるいは視覚ユニット層) に、特定の活性化パターンが示されるようにニューラルネットワークを訓練した。

4.1.2　視覚性障害仮説についてのシミュレーション

訓練後、意味層のユニットを除去することで脳損傷がシミュレートされた。脳損傷を受けていないネットワークと脳損傷を受けたネットワークとに対して、視覚刺激が完全である場合

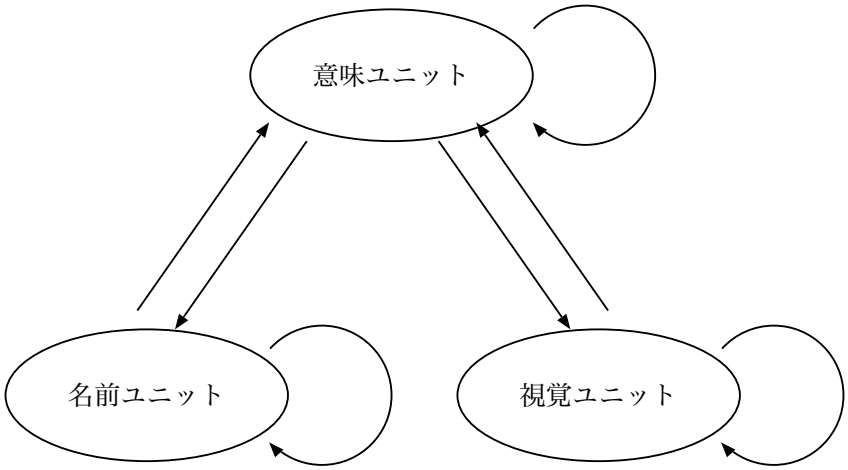

図 4.1: ティペットとファラーの用いたモデルの概略図

と、不完全な場合とで名前ユニット層に現われる活性化パターンがどのように変化するのかが観察された。

結果は交互作用が観察された。正常なネットワークに、不完全な視覚刺激を与えても、名前ユニットに現われる活性化パターンは学習したパターンに近いものが観察されたが、脳損傷を受けたネットワークでは物体呼称成績が著しく障害されたのである。

4.1.3　語彙の頻度効果に対するシミュレーション

続いて、語彙の頻度効果を調べるために、高頻度語と低頻度語との違いをニューラルネットワークの訓練回数の違いとして表現した。視覚性障害仮説についてのシミュレーションと同じように、訓練後、意味層のユニットを除去することで脳損傷が表現された。脳損傷を受けたネットワークの名前ユニット層に低頻度語を提示したときの成績は著しく低下するが、高頻度語を提示したときの成績はそれほど低下しなかった。すなわち、語彙性の障害と考えられてきたような症状を意味層の障害によって再現できたのである。

以上見てきたように、ニューラルネットワークを用いると、意味記憶の障害だけを仮定すれば、視覚性の誤りも語彙性の誤りも説明できる。すなわち従来から考えられてきた課題成績によるアルツハイマー症の障害の分類に全く新しい視点が与えられるのである。それでは、意味記憶そのものの構造はどのようになっているのであろうか。

4.2　意味記憶の構成 —生物、非生物の二重乖離—

認知心理学でしばしば話題になる記憶表象論争に関して、意味記憶は、個々の対象についてカテゴリーごとに構成されているのかそれとも、それともモダリティーごとに構成されているのか、という論争がある。ファラー (Farah) とマクレランド (McClelland)[12] が行なったニューラルネットワークによる研究によれば、モダリティーに依存した意味記憶表象を考えれば、カテゴリーに基づく意味記憶表象は説明できる。

4.2.1 神経心理学的症状

実際の脳損傷患者の中には、動物や植物などの生物の知識について障害がある一方で、非生物の知識については健常のまま保たれている患者が存在する [27]。古典的な二重乖離の原則から、生物と非生物の知識の脳内での意味記憶には、生物と非生物とを独立に表象している意味記憶が存在すると仮定される。しかし、ウォリントン (Warrington) とシャリス (Shallice) は、生物の知識と非生物の知識との間で選択的な障害が起こるのは、異なる感覚運動経路からの情報の重みの差異を反映しているためではないか、と述べている。すなわち、生物は主に感覚的な性質によって互いを区別することが多いが、非生物は主に機能によって分類される。ある動物、例えばヒョウは、他の肉食動物と比べて主に視覚的な特徴によって差別化される。これとは対照的に、机の知識については、他の家具との違いを記述するときには主に機能、すなわち何のために使うのか、によって差別化される。それゆえ、障害のある知識と健全に保たれている知識との違いは、生物–非生物の違いなのではなく、対象を記述している特徴が感覚–機能の違いであるかも知れない。

ファラーとマクレランド [12] のモデルは上記の感覚–機能仮説が意味記憶障害を説明できることを例示するために作成された。

4.2.2 モデル

彼女らのモデルを図 4.2 に示す。3 つのユニット群、記憶を表現する意味記憶系と、入出力を表現する二つの周辺系、視覚ユニット群と言語ユニット群とがある。言語ユニット群と視覚

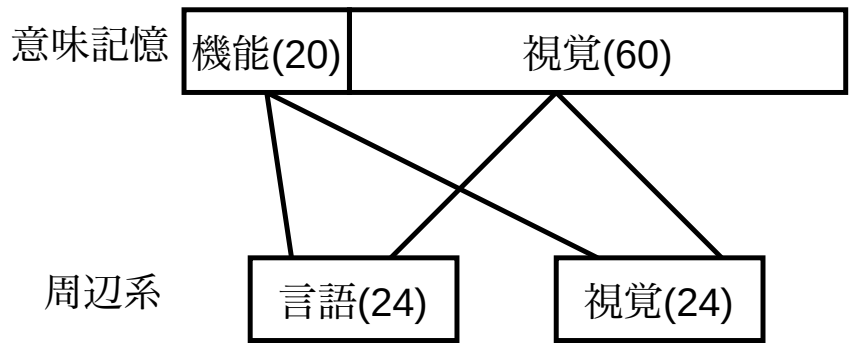

図 4.2: ファラーとマクレランド (1991) の意味記憶モデルの概念図。カッコ内の数字は数値実験で用いられたユニット数を表す。意味記憶内で機能的記憶と視覚的記憶のユニット数が異なるのは、彼らの論文中の実験 1 (心理実験) の結果を反映している。

ユニット群との間を除いて、全てのユニットに群間および群内結合が存在した。

このモデルに生物と非生物を表す刺激が提示された。生物と非生物とを表す項目は、視覚情報と機能情報との比率が変えられた。生物項目では平均して 16.1 の視覚意味記憶ユニット、2.1 の機能意味記憶ユニット。非生物では 9.4 の視覚意味記憶ユニット、6.7 の機能意味記憶ユニットを用いて表現された。視覚パターンが提示されたときには対応する意味記憶パターンと言語パターンが産出されるように、また、言語パターンが提示されたときには対応する意味記憶パターンと視覚パターンが産出されるよに訓練された。各訓練試行では、生物、もしくは非生物に対応する視覚入力や言語入力が言語ユニット群あるいは視覚ユニット群に対して提示され、ネットワークは解が安定するまで活性値の更新が行なわれた。

4.2.3　破壊実験

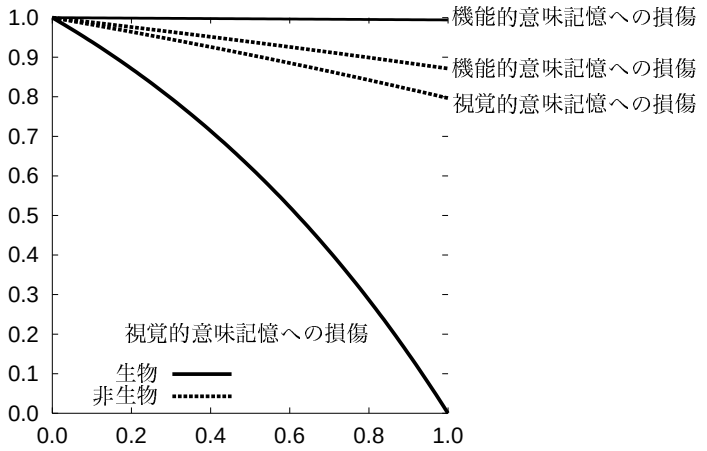

図 4.3: 生物–非生物別の意味記憶内の損傷の程度と課題成績との関係 (ファラーとマクレランド (1991) の表 3 と図 2 より改変)。

彼女らは、モダリティーに依存した意味記憶障害、すなわち、耳で聞いたときには理解できるが、目で見たときには特定のカテゴリーについての知識に障害を生じる患者のシミュレーションを行なった。さらに、ネットワークに雑音を加える方法によって脳損傷を表現し、シミュレーションを行なった。彼女らのシミュレーション結果は、モダリティー依存の意味記憶構造を用いれば、カテゴリー依存の障害を説明できることを示している。すなわち生物、非生物という異なるカテゴリーに属する事物は脳内で異なって表象されているのではなく、意味記憶の構造はは入力モダリティーに依存して形成されているという結果が得られたのである。

4.3　ニューラルネットワークによる障害の再解釈

脳損傷を扱った認知障害の研究では二種類の実験を使って脳損傷患者の障害の場所を特定することが試みられてきた。一つめは患者の示す誤りの種類を分析することである。例えば、物体の呼称課題において、視覚的に似ているものを答える誤りについては視覚性の誤りと分類し患者の視覚機能に問題があると推論する。同様に意味性の誤りについては意味記憶に障害があると判断される。二つめは課題の難易度を操作して障害の場所を特定することである。例えば呼称課題において、低頻度語の呼称成績が選択的に障害されていれば、語彙に障害があるとする。視覚刺激の質を低下させたときに、例えば写真を用いた物体呼称課題と線画を用いた物体呼称課題で、線画を用いた方が呼称成績が悪ければ、視覚性の障害とみなす、などである。

ところが、今回説明したようにニューラルネットワークモデルによる研究では、認知機能の局在を示す脳損傷患者のデータと、その認知機能を推論する伝統的な認知神経心理学的手法に疑問を投げかけている。ニューラルネットワーク研究によって、従来からの神経心理学的障害分類論に本質的な変更が迫られているように思えるのだ。

第5章　文法知識の創発と失文法

チョムスキーの生成文法理論によれば言語獲得とは、任意の言語の文法を収得することである。この文法の知識に障害が起きると失文法と呼ばれる神経心理学的症状が現われる。今回はこの失文法をめぐるニューラルネットワーク研究を紹介しよう。チョムスキー派の人たちの関心も高いのではないだろうか。

5.1　失文法 agrammatia の種類

失文法患者には、発話の中でも品詞によって障害されやすさに不均衡があることが知られている。すなわち図 5.4 と図 5.5 のようなことが実際に観察されている。また、語順障害と語尾の欠落や誤用との二つに区別されるとする研究者もいる。重要なことは、発話の流暢性と文法的な発話の産出は二重に乖離していることである。すなわち

1. 表出が乏しい、内容語は産出できるが機能語が産出できないタイプ (電文体のような発話になる) 形態的失文法
2. 内容語が乏しいが機能語は流暢で豊富である統辞的失文法だと

という二種類の患者が存在する。

実際、構文的に乏しく努力性の失文法的発話を示す患者では前方言語野 (前頭葉) や深部の構造である島皮質、さらに側頭葉前部が損傷を受けることが知られており、流暢な発話における構文の障害では側頭葉、頭頂葉下部、および弓状回などに障害がある場合が多いとされている。

5.2　失文法と文法判断

ブローカ失語とよばれる患者の中には、文章の理解は困難であるが、与えられた文章が文法的に正しいか否かを判断する文法判断課題の成績は保たれている患者が存在する。この種の患者の発話の特徴は電文体と言われるもので、文章中の冠詞や前置詞などの機能語が脱落する傾向にある。アレン (Allen) とサイデンバーグ (Seidenberg)[1] は文章理解と文法判断との乖離を説明するニューラルネットワークモデルを作成した。彼らの用いたニューラルネットワークモデルの概略を図 5.1 に示す。図中 clean up と書いてあるユニット群はエルマンネットの文脈層を拡張した仕様になっている。エルマンネットがシステムの状態更新に離散時間を用いたニューラルネットワークであったのに対し、アレンとサイデンバーグのニューラルネットワークは連続時間を用いている。さらにエルマンネットでは中間層から文脈層への結合強度が 1 に固定されていたのに対し、彼らのモデルでは clean up 層への結合係数も、経時的誤差逆伝播法 (Back Propagation Through Time) を用いて学習を可能にした。

学習はエルマンの系列予測課題とほぼ同様の手続きを 2 種類行なった。単語層に単語を逐次提示し、中間層を介して対応する意味表現を学習させる文章理解課題と、反対に文章に対応す

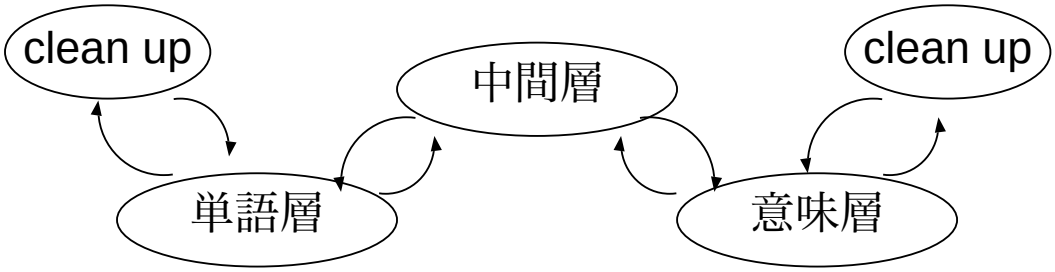

図 5.1: 文章理解と文法判断のためのネットワーク図

る意味の系列を意味層に逐次提示し対応する単語を出力するように学習させる文章産出課題とである。中間層ユニットは単語層と意味層とに結合され双方向の結合を持ち、中間層ユニットの介在によって文章理解課題と文章産出課題の系列再生の橋渡しがなされる。

学習の成立したネットワークに対して、与えられた文章が文法的に正しいか否かを判断させる文法判断課題は次のように定義された。単語層に逐次単語を入力し、意味層を介して逆方法に計算されて戻ってきた出力文が入力文と異なるか否かで判断された。すなわち、入力文と、意味層を介してフィードバックされた文との差に基づいて文法判断がなされると仮定された。学習の結果、ネットワークは文法的に正しい文章については正確に予測することができ、文法的に誤った文章については予測ができなかった。すなわちこのネットワークは文章理解と共に文法判断も正しく行なう能力を持っていたと言える。

彼らは学習の成立したネットワークを破壊し、動詞や名詞などの内容語に比べて、冠詞や前置詞などの機能語 (高頻度単語だが意味を持たない) の産出に失敗やすいことを見出した。このことは失文法患者の電文体の発話に対応するものと考えられる。この現象は、意味層における表現において内容語によって形成されるアトラクタの方が機能語のアトラクタよりも損傷に対して頑健であったと説明されている。

彼らはさらに、ネットワークの損傷によって文章理解の障害 (与えられた文章の再活性化に失敗する) を再現できることを示した。損傷後のネットワークは文章の理解には失敗するものの、文法的に正しい文章と文法的に正しくない文章とを区別する文法判断課題では、与えられた文章の文法性を正しく判断する能力を持っていた。しかも、文型毎に比較すると、損傷後のネットワークによる文法判断の出力と、失文法患者が文法判断課題において示す誤りのパターンとは一致することが分かった。すなわち彼らのネットワークでは、文章理解と文法判断の乖離をシミュレートできたことを意味する。

換言すれば、アレンとサイデンバーグのモデルは系列予測課題によって文法知識 (あるいは単語間の遷移確率という言語の持つ統計的構造) を獲得したと見なすことができる。このモデルは与えられた文章が文法的に正しいか否かを判断する能力を持っていた。モデルの示した文法判断能力はネットワークが学習を通して徐々に形成されたものであり、この意味においてニューラルネットワークの特徴をすべて持っている。アレンとサイデンバーグのモデルは、言語学者がその理論的根拠だとしている文法判断課題をニューラルネットワークの枠組で説明したモデルであると言えよう。

5.3　言語産出と聴理解の二重単純再帰型ネットワーク

左半球のシルビウス裂によって二つの言語野、ブローカ野とウィルニッケ野は離されている。このシルビウス裂を開いてみると島皮質 insula という部分が現われる。島皮質はブロー

カ野とウィルニッケ野の中間に位置すると考えることができ、最近では言語の発話に関しても島皮質が関与している可能性も指摘されている [7]。ここでは島皮質の計算論的役割としてブローカ野とウィルニッケ野を結びつける役割の可能性があることを指摘しよう。島皮質で起こっていることが文章産出と言葉の聴理解に密接に関っている可能性があり、大胆に仮説を構成すれば二重単純再帰型ニューラルネットワークである。二重単純再帰型ニューラルネットワークは言語産出と言語理解が密接に関っていることを表すおそらく最も単純なモデルである。我々が何かを話すときに起こっていることは、おそらく話したい内容がブローカ野に形成され、ブローカ野の指示に従って補足運動野や運動野を介して発話にいたる。一旦発話した内容は側頭平面にある第一次聴覚野を介してウィルニッケ野に入力される。つまり我々が話しているときには文章産出と文章理解の両者を同時に行なっているのだ。その証拠に自分が

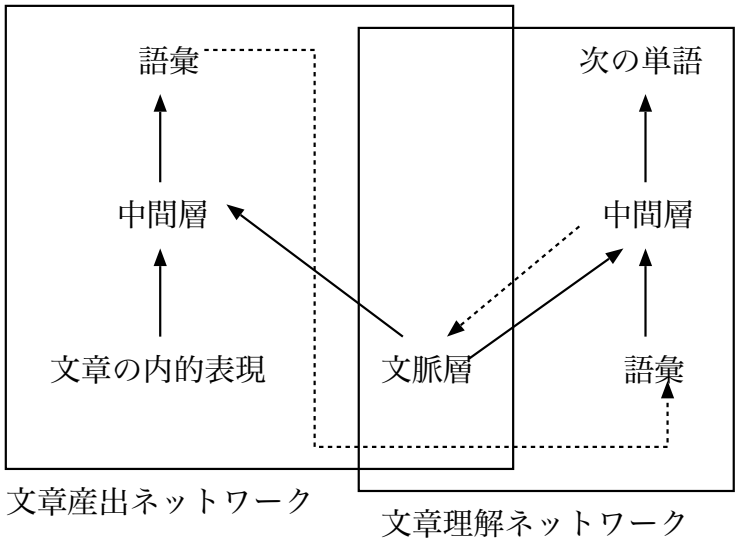

図 5.2: 2 重経路単純再帰型ニューラルネットワークモデル

話した言葉をマイクロフォンで録音し、一定の遅延をおいてヘッドフォンでその言葉を聴かせると言語産出が困難になる。このような心理実験課題とその効果のことを DAF (Delayed Auditory Feedback) という。DAF の存在が示していることは文章産出と文章理解とは密接にからみ合っており、切り離すことはむずかしいということである。

エルマンネットを使うと文章理解が可能であることは以前既に述べた。同じようにして入力刺激を一定の値に固定しておいて文脈層の変化によって文章産出を指せることも可能である。このような方法をプロダクション SRN と言ったりする。この二つのエルマンネットの文脈層を共有させるというモデルが二重単純再帰型ニューラルネットワークである。このモデルは大まかにブローカ野とウィルニッケ野という脳内の言語を司る領野とが文脈層 (島皮質？) を介して結びついているということを表す、もっとも単純なモデルであると見なすことができよう [5, 6]。二重単純再帰型ニューラルネットワークを使うことによって最も基本的な文章産出と文章理解の相互作用をモデル化することができるのである。図 5.2 に二重単純再帰型ニューラルネットワークを示した。

5.4　シミュレーション

文章産出ネットワークで生成された単語が次の時刻の聴理解を担当するエルマンネットである文章理解ネットワークへの入力となる。この二重単純再帰型ニューラルネットワークを用いて簡単な日本語の文章の産出と理解を訓練させてみた。訓練に用いた文型は全 18 文で以下の図 5.3 である。入力表現は、主格太郎、主格次郎、主格三郎、目的格太郎、目的格次郎、目的

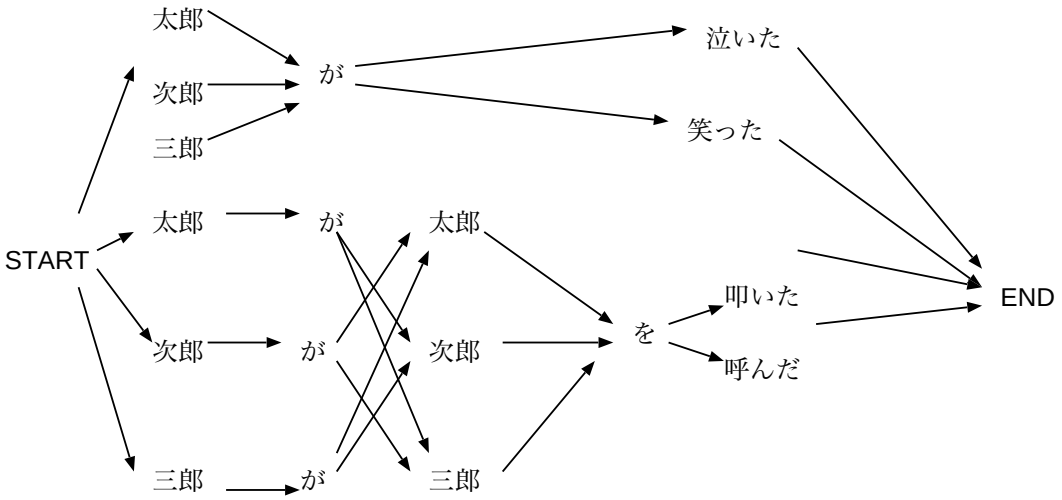

図 5.3: 二重単純再帰型ニューラルネットワークの訓練に用いた文章。上の文が type I, 下の文が type II

格三郎、笑った、泣いた、呼んだ、叩いた，の 10 ビットを 0,1 で表現した。例えば
「太郎が泣いた」→ 1,0,0, 0,0,0, 0,1,0,0
となる。

出力表現は EOS(文章の終わり), 太郎、次郎、三郎、笑った、泣いた、呼んだ、叩いた、を、が、の各ビットを 0,1 で表現した。例えば、「太郎が泣いた」という文は、時刻 t=1 で「太郎」を表すビットが 1 となり、かつ、他のビットはすべて 0 であるように 0,1,0,0, 0,0,0,0,0, 0,0 と表現された。次の時刻 t=2 では格助詞「が」を表すビットが 1 であり、かつ、他のビットは 0 となるように、0, 0,0,0, 0,0,0,0,0, 0,1 と表現された。以下、同様に時刻 t=3 では「泣いた」を表すビットが 1 であり、かつ他のビットはすべて 0 と表現され、最後の時刻 t=4 では EOS(文章の終わり) を表現する最初のビットが 1 で他のビットがすべて 0 で表現された。すなわち「太郎が泣いた。」という文は、時刻 t=1 から t=4 までの時間刻みを用いて

```
0, 1,0,0, 0,0,0,0,0, 0,0 # (t=1)
0, 0,0,0, 0,0,0,0,0, 0,1 # (t=2)
0, 0,0,0, 0,1,0,0,0, 0,0 # (t=3)
1, 0,0,0, 0,0,0,0,0, 0,0 # (t=4)
```

などとなる。日本語の文章としては単純すぎるという反論は十分予想されるのだが、ここではいかに複雑な文章を産出、理解させるのかを目的にしているわけではなく、文章産出と文章理解の基本的な相互作用の在り方をシンプルに考えてみようという試みである。

二重単純再帰型ニューラルネットワークモデルでは、図 5.2 中の中央にある文脈層は、左側の文章産出ネットワークのための状態空間を遷移する。一方、図中の右側のネットワークは文章理解のためのエルマンネットであるから、次の単語を予測するために中央の文脈層が用いら

れる。すなわち図中の文脈層は文章産出の状態と単語予測のための状態とを同時に処理しなければならなかったのである。

学習の成立した二重単純再帰型ニューラルネットワークの文脈層では、いったいどのようなことが起こっていたのであろうか。実は二重単純再帰型ニューラルネットワークの理論的解析についての研究は少なく、多くのことが分からずに残されている。従って文章産出と文章理解のエルマンネットで共有される文脈層で起こっていることは不明な点が多いのである。しかし、ニューラルネットワークが言語研究において有力な手段を提供しているのは直接シミュレーションをして調べてみることができることである。

学習の成立した二重単純再帰型ニューラルネットワークに対して共有されている文脈層を破壊することによって人工脳損傷をおこさせてネットワークの振舞を観察してみることにした。人工脳損傷は、自由に、いつでも、どの場所でも破壊することができるので、理論認知神経心理学の有効な技法の一つになりうると考えている。

シミュレーションの詳細を記述すると、中間層ユニット数を 10 にして、全 18 文のうちランダムに 15 文を学習データとして訓練した。残りの 3 文を用いて般化誤差を測定した。般化誤差が小さくなったパラメータを用いて文脈層のユニットを破壊した。

人工脳損傷によるシミュレーションからから非常に興味深い結果が得られた。結果を二つのグラフに示す。

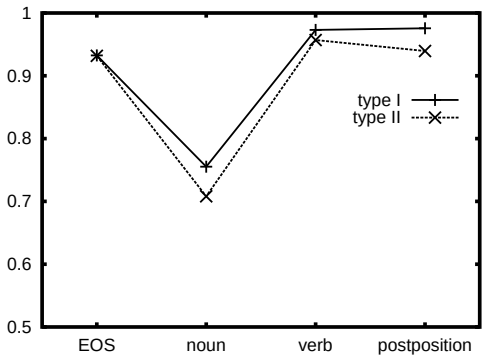

図 5.4: 失名辞タイプの損傷例。縦軸は正解率を表す。type I,II の違いについては図 5.3 のキャプションを参照

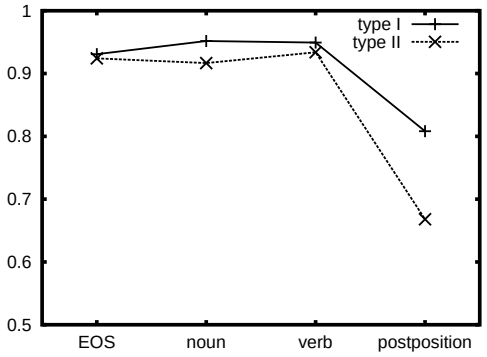

図 5.5: 形態的失文法タイプの損傷例。縦軸は正解率を表す。type I,II の違いについては図 5.3 のキャプションを参照

この二つのグラフは、人工脳損傷を文脈層に起こしたときに、各品詞の正解率を表している。すなわちグラフで高い位置にある品詞は人工脳損傷の影響を受けなかったことを表し、逆にグラフの下にきている品詞は成績が悪かったことを示している。両グラフの違いは、破壊した文脈層ユニットの違いである。すなわち別の文脈層ユニットを破壊すると別の品詞の成績が悪くなることがこの二つのグラフから読み取れるのだ。一方では格助詞が、他方では名詞が選択的に障害されている。

5.5　文法知識の創発と失文法

従来の認知神経心理学の枠組みでは、形態的失文法と統辞的失文法とは二重に乖離しているので異なる脳内モジュールが障害を受けたと考えざるを得なかった。

ところが、紹介した二重単純再帰型ニューラルネットワークの人工脳損傷のシミュレーションでは同一モデルで二重に乖離した二つの失文法を、文脈層ユニットに形成されたと考えられる文章産出と文章理解のために同時に利用される文法知識の障害として説明できるのである。しかも、この文法知識は筆者がアプリオリに与えたものではなくニューラルネットワークの訓練の結果として文法知識が創発されたのである。結果を表す二つの図を見ると、従来の神経心理学的症例分類学をみなおす必要があるのではないかと思えてくるのである。

第6章　自己組織化の意味と意味の自己組織化

6.1　はじまりは...

「自己組織化」とは非常に壮大なテーマである。この問題に直接答えるのには私には荷が重すぎるし、この連載の主旨からもはずれてしまう。だがあえて自己組織化の定義を試みれば「自己組織化とは、経験と環境の関数として基本構造が変化し、合目的的システムができあがること」と定義することができるだろう。

例えば、人間は自己組織化システムである。だれもが一個の有精卵から次第に複雑な構造を発生させて行ったのだから。もっとも、すべての生物は自己組織化システムであるし、太陽系も自己組織化システムだと言うことができるかも知れない。さらに初めの初めから始めるとすれば、30 億年前原始地球の原始スープの中から長い年月をかけて自己複製を始めた生物の発生にさかのぼることができる。原始生命の出現に超越的な創造者の存在を仮定するべきなのだろうか？それとも、現在の生物の持つ自己複製機能の創発を認めるべきなのだろうか？現代生化学の研究成果は、超越的な創造者の存在を仮定しない生命発生のシナリオを描き始めているように思われる。単純な自己増殖機能を持ったタンパクから、やがて細砲が作られ単細胞生物へ、さらに多細胞生物へ、さらに陸上へと進出し、火を発見し、文字を発明し、知的活動を行なうようになった実例が今の私たちである。生物が自身の知的活動をシミュレートするようになるまでには、多様なレベルでの自己組織化が行なわれて来たのだと想像できる。

現代的な意味でのニューラルネットワークにおいては、上記のような意味での「自己組織化」は実現されていない。現在のニューラルネットワークにできることは、極論すれば、外界の統計的構造を獲得することができるという点である。もうすこし具体的にいえば、外部入力の統計的構造を内部のシナプス伝導効率の変化として表現することができる、ということである。ここから、知的な活動を創発できることまでの間には厖大な距離がある。だが、外界の情報すなわちデータの相互関係を効率良く表現することは情報科学の分野でも中心的な問題であり、おそらくこのような能力が脳の働きの特徴の１つであるということができる。今回は自己組織化という壮大なテーマの入口、外界の情報から意味のある構造を作りだす、という点に的を絞って説明しよう。

6.2　トポグラフィックマッピング

外界の構造が脳内の地図として表現されていることは一般に知られた事実である。例えば、網膜と第一次視覚野の間には連続的な１対１対応が存在する。鼓膜の周波数選択特性と第一次聴覚野との間にも対応関係が見られる。同様に体表面の感覚と体制感覚野の間にも対応関係が見られる。すなわち感覚器官と第一次感覚野との間の神経結合は、類似した刺激に対して皮質上の同じような位置に対応する受容野を持つことが知られている。このような 2 つの神経場間の連続的な結合関係のことをトポグラフィックマッピング topographic mapping と言

う。このような構造は、大まかには遺伝子によって決定されているが、細かい構造については神経回路の自己組織化によって達成されると考えられている。

6.3　自己組織化の意味するもの

大脳皮質全体の 10 % を占める第一次感覚野で起こっていることの類推から、特定のカテゴリーにおける知識表現が脳の各部位の位置関係として表現されているという可能性があるだろうと考える。すなわち、さまざまなレベルの情報表現の自己組織化に対して、たった 1 つの同じ機能的原理が働いているのではないか、という仮説が提起できる。第一次感覚野で表現されている情報表現と同じ機能的原理が、知的なレベル (各種の連合野、あるいは前頭葉) でも同じであると考えてはいけない理由はないはずである。仮に、この同一の機能的原理が高次の知的活動のためにも働いているのなら、低次の感覚受容野から、階層的に高次の連合野にいたるまで自己組織化によって我々の知的活動のある部分が説明可能なのかも知れない。自己組織化によって高度に抽象的な概念が階層的に重ね合わさっていた場合にどのようなことが起こるのだろうか。第 1 次感覚野が物理的な特徴量を表現し、第 2 次感覚野が具体的な概念を表現しているとしたら、連合野は抽象的な概念を表象しているのかも知れない。連合野の連合野である前頭葉では概念の概念の概念が形成されているというのは誇張のしすぎなのだろうか。概念の概念の概念は知的な能力とみなしても良いと思う。すなわち自己組織化が多段階に重なることによって抽象度の高い知的能力が創発すると考えても良いのではないだろうか。

6.4　意味の抽出

ランダウアー (Landauer) とデュマス (Dumais)[17] は、百科事典のすべての文章における単語の見出し語項目との間の共起関係に特異値分解を適用し、数百個の次元からなるベクトル表現を構成した。このベクトル表現によって単語間の類似度を定義し、TOEFL の類義語問題に結果を適用することで人間の受験者に近い正答率が得られることを示している。彼らによればベクトルの次元数を 300 としたとき一致率が最大になるという。このことから人間の意味処理として 300 次元程度の意味空間を用いることでコンピュータに人間に近い振る舞いをさせることができるという結果が得られている。このことは従来曖昧な定義であった意味に対して計量的なアプローチが可能であることを示していて興味深い。

6.5　自己組織化アルゴリズム

ランダウアーとデュマスの使った特異値分解とは、数学的手段であり、固有値問題と関連が深い。固有値問題は、多次元の情報を情報の損失を最小にしながら低次元の情報に変換する情報圧縮のために使われたりもする。従って与えられたデータの固有値問題の解を自己組織的に学習して解くニューラルネットワークがあれば、ランダウアーとデュマスたちの示した結果をニューラルネットワークでも表現できることになる。固有値問題を解く自己組織化ニューラルネットワークには、ヘッブ (Hebb) の学習則、およびヘッブの学習則を拡張したオヤ (Oja) の学習則 [19]、オヤの学習則を拡張したザンガー (Sanger) の学習則 [23] などが知られている。固有値問題とはどのようなものかを説明せずに、この文章を読んでも意味不明であるとは思うが、固有値問題とは情報の圧縮であり、抽象化である、と思っていただいてよい。ヘッブの学習則を使うとシナプスの結合係数が最大固有値に対応する固有値ベクトルの方向と一致し、オヤの学習則を使うとその固有ベクトルが 1 に規格化され、ザンガーの学習則を使うと望む

数だけ固有ベクトルが大きい順にとりだせるということである。数式を用いずにこれらのことを説明するのは大変なのだが、多数のニューロンと結合を持つ一つのニューロンを考えたとき、このニューロンへのシナプス結合係数の変化は、このニューロンの発火率とこのニューロンへ信号を送っているニューロンの発火率の積で表されるというのがヘッブの学習則であり、ヘッブの学習則に正則化項を取り入れたものがオヤの学習則であり、オヤの学習則を多層化したものがザンガーの学習則なのである。

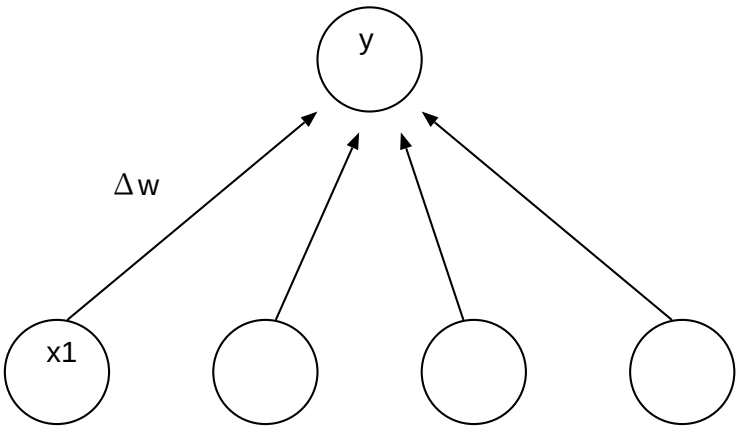

図 6.1: 自己組織化の例。2 層のネットワークを考え、下位層の各ユニットと上位層にあるユニットとの結合係数 w が固有ベクトルに対応するような自己組織化アルゴリズムが提案されている。$\Delta w_i = \eta y x_i$ がヘッブの学習則であり、$\Delta w_i = \eta y(1-y)x_i$ がオヤの学習則であり、$\Delta w_i = y(x_j - \sum x_k w_k)x_i$ とするのがザンガーの学習則である。

数式を使わないで説明したため、いささか面妖な日本語になったがお許し願いたい。要するに初めに戻って、外界の統計情報を効率良く学習するニューラルネットワークモデルが実在するのだと言いたいのだ。そしてランダウアーとデュマスの結果を信じれば 300 次元程の意味次元を考えれば人間の知識を表現することができ、また、それは自己組織化アルゴリズムを用いて実現可能だと言うことである。

6.6 NMF による意味の分解

同じような発想から非負行列因子化 (NMF) と呼ばれる手法も最近注目を集めている [18]。NMF は入力データを構成する基底を抽出する自己組織化アルゴリズムである。実際に NMF を顔画像処に用いていろいろな人物の正面顔を入力した場合では、顔を構成するパーツ、目や鼻や口のような画像が基底として抽出された。NMF は基底と展開係数の成分が共に非負であるという性質を持っている。NMF は基底と展開係数を更新することによって外界情報の持つ性質を抽出する自己組織化アルゴリズムの一手法である。NMF を事典の各項目に対して応用した例では、例えば「アメリカ合州国憲法」は大統領、議会、権力などの各因子を展開係数を用いて加算した形で表される。このように各概念が、下位概念 (因子) とその重みである展開係数の積とで表現される点が NMF の特徴である。

NMF の応用を示した簡単な例が `http://www.twcu.ac.jp/~asakawa/nmf/`にあるので参照して頂きたい。ここでは小学生が学習する学習漢字 1006 字に対して NMF を実行し、ごんべんやしんにょうなどのパーツが基底として抽出されたことが示されている。

NMF を使えば、ランダウアーとデュマスの使った語彙も、各々の単語の下位のパーツとなる語彙とその展開係数で表すことが可能であろう。「意味の自己組織化」の試みは確実に進歩してきているといってよいだろう。これは一昔前にくらべてコンピュータの処理速度と記憶容量とが十分になってきていることと関係している。いよいよ面白い時代になってきたと言えるのではないだろうか。

6.7　ニューラルネットワークから見える言語の風景

以上 6 回にわたって、ニューラルネットワークの言語への応用に関する話題を取り上げて、平易に解説したつもりである。言語にまつわる話題をニューラルネットワーク研究の立場から見たときに、伝統的な言語理論や神経心理学に対する批判、いわば、「新しい脳観」が見えてくると思うのだ。

第二次ニューロブームといわれた 1980 年代からすでに十五年以上が経過し、ニューラルネットワークという言葉も一般的な言葉になってきた。しかし、その実態は必ずしも正確に知られていないように思う。この連載で、少しでもニューラルネットワーク研究を理解していただけたとしたら望外の喜びである。御意見、ご批判、抗議などを電子メールで asakawa@cis.twcu.ac.jp に送っていただければ幸いである。

関連図書

[1] Joseph Allen and Mark S. Seidenberg. The emergence of grammaticality in connectionist networs. In B. MacWhinney, editor, *The Emergence of Language*, pages 115–151. Lawrence Erlbaum, Mahwah, NJ, 1999.

[2] Michael A. Arbib. *The metaphorical Brain 2: Neuran Networks and Beyond*. John Wiley and sons, 1989. ニューラルネットワークと脳理論第２版, 金子隆芳訳, 1992, サイエンス社.

[3] Max Coltheart, B. Curtis, P. Atkins, and M. Haller. Models of reading aloud: Dual-route and parallel-distriputed-processing approaches. *Psychological Review*, 100(4):589–608, 1993.

[4] Max Coltheart and K. Rastle. Serial processing in reading aloud: Evidence for dual-route models of reading. *Journal of Experimental Psychology: Human Perception and Performance*, 20:1197–1211, 1994.

[5] G. Dell, M. Schwartz, N. Martin, E. Saffran, and D. Gagnon. Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, 104, 1997.

[6] Gray S. Dell, Franklin Chang, and Zenzi M. Griffin. Connectionist models of language production: Lexical access and grammatical encoding. In Morten H. Chirstiansen and Nick Charter, editors, *Connectionist Psycholinguistics*, chapter 7, pages 212–243. Ablex Publishing, Westport, CT, 2001.

[7] Nina F. Dronkers. A new brain region for coordinating speech articulation. *Nature*, 384:159–161, 1996.

[8] Jeffrey L. Elman. Finding structure in time. *Cognitive Science*, 14:179–211, 1990.

[9] Jeffrey L. Elman. Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, pages 195–225, 1991.

[10] Jeffrey L. Elman. Learning and development in neural networks: The importance of starting small. *Cognition*, pages 71–99, 1993.

[11] Jeffrey L. Elman, Elizabeth A. Bates, Mark H. Johnson, Annette Karmiloff-Simth, Domenico Parisi, and Kim Plunkett. *Rethinking Innateness: A connnectionist perspective on development*. MIT Press, Cambridge, MA, 1996. (邦訳「認知発達と生得性」, 乾, 今井, 山下訳, 共立出版).

[12] Martha J. Farah and James L. McClelland. A computational model of semantic memory impairment: Modality specificity and emergent category specificity. *Journal of Experimental Psychology: General*, 120(4):339–357, 1991.

[13] Matha J. Farah. Neuropsychological inference with an interactive brain: A critique of the locality assumption. *Behavioral and Brain Sciences*, 17:43–104, 1994.

[14] J. A. Fodor. *The modularity of mind*. MIT press, 1983.

[15] Robert A. Jacobs, Michael I. Jordan, Steven J. Nowlan, and Geoffery E. Hinton. Adaptive mixtures of local experts. *Neural Computation*, 3:79–87, 1991.

[16] Michael I. Jordan and Robert A. Jacobs. Hierarchical mixtures of experts and the em algorithm. *Neural Computation*, 6:181–214, 1994.

[17] T. K. Landauer and S. T. Dumais. A solution to plato's problem: The latent semantic analysis theory of acquistion, induction, and representation of knowledge. *Psychological Review*, 104:211–240, 1997.

[18] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791, 1999.

[19] E. Oja. A simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, 15:267–273, 1988.

[20] David C. Plaut. A connectionist approach to word reading and acquired dyslexia: Extension to sequential processing. In Morten H. Chirstiansen and Nick Charter, editors, *Connectionist Psycholinguistics*, chapter 8, pages 244–278. Ablex Publishing, Westport, CT, 2001.

[21] David C. Plaut, James L. McClelland, Mark S. Seidenberg, and Karalyn Patterson. Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, 103:56–115, 1996.

[22] Douglas L. T. Rohde and David C. Plaut. Language acquistion in the abasence of explicit negative evidence: How important is starting small? *Cognition*, pages 69–109, 1999.

[23] T.D. Sanger. Optimal unsupervised learning in a single–layer linear feed–forward neural network. *Neural Networks*, 2:459–473, 1989.

[24] Mark S. Seidenberg. Language acquisition and use: Learning and applying probabilistic constraints. *Science*, pages 1599–1603, 1997.

[25] Mark S. Seidenberg and Maryellen C. MacDonald. Constraint satisfaction in language acquisition and processing. In Morten H. Christiansen and Nick Charter, editors, *Connectionist Psycholinguistics*, chapter 9, pages 281–318. Ablex publication, Westport, CT, 2001.

[26] Lynette J. Tippett and Martha J. Farah. Parallel distributed processing models in alzheimer's disease. In Randolph W. Parks, Daniel S. Levine, and Debra L. Long, editors, *Fundametals of Neural Network Modeling: Neuropsychology and Cognitive Neuroscience*, chapter 17. MIT press, 1998.

[27] Elizabeth K. Warrington and Tim Shallice. Category specific semantic impairment. *Brain*, 107:829–854, 1984.

[28] 守一雄, 都築誉史, and 楠見孝, editors. **コネクショニストモデルと心理学**. 北大路書房, 2001.