

単純再帰型ネットワーク

浅川伸一 <asakawa@twcu.ac.jp>

エルマン (Elman, 1990) 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。その理由は、エルマンの用いたニューラルネットワークでは、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だからである。

「刺激の貧困」とは、プラトンの考えた問題であるとされている。言語の発達過程にある幼児が耳にする言葉は、多くの言い間違いや不完全な文を含んでおり、限りある言語データしか与えられない。それにもかかわらず、どうしてほとんど無限に近い文を発話したり解釈したりできるようになるのか、という問題である。

「否定証拠の欠如」とは、ゴールド (Gold, 1967) によって定式化された。ある文法規則に従う正しい文と、文法規則に従わない誤った文 (否定証拠) を一つずつ与えていき、元の文法を推定することを文法推論という。正しい文法推論とは、与えられた文の列から元の文法を推定するときに、ある有限の数の例から先では、その推定が同じ正しい結果に収束して変わらないということである。文法規則に従う正しい文だけを提示したときは文法推定が不可能である、すなわち獲得できないことをゴールドは証明した。

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。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。

エルマンの研究 (Elman, 1991) のめざしたことは (1) 言語的知識の表象とはどのようなものか、(2) 時制の一致や文章の関係代名詞による文章の再帰的構造などの複雑な文法がどのように表象されているのか、(3) ほぼ無限とも思われる単語の組み合わせが限られた資源のネットワークの中でどのように調整されているのか、という疑問に答えることであった。

1 エルマンネット

エルマンネットでは、入力層は入力信号を処理する入力ユニットと、直前の時刻までの中間層の状態を入力とする文脈ユニットとで構成される(図1)。文脈ユニットは以前の間層をコピーするためだけ(すなわち中間層から文脈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。

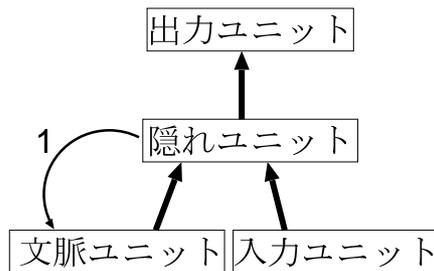


図 1: エルマンネット

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。例えば、図2において過去の影響を考えれば時刻 t における中間層の状態は

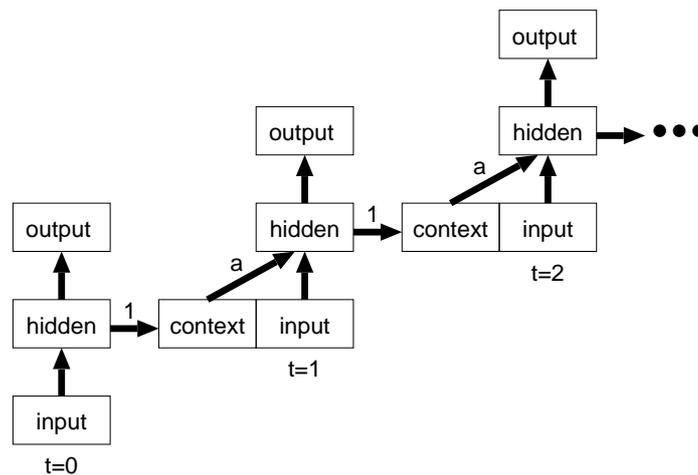


図 2: エルマンネットの時間発展

$$\begin{aligned}
h(t) &= I(t) + ah(t-1) \\
&= I(t) + a(I(t-1) + h(t-2)) \\
&= I(t) + aI(t-1) + a^2I(t-2) + a^2h(t-3) \\
&= \sum_{\tau=0}^T a^\tau I(t-\tau)
\end{aligned} \tag{1}$$

と表すことができる。ここで $I(t)$ は時刻 t における入力を表している。文脈層からの影響 (図中では a と表記) が 1 より小さければ過去の入力からの影響が指数関数に従って小さくなることを表している (1 より大きければ過去の状態からの影響は大きくなる)。このことは、エルマンネットが一つ前の状態を保存しておくという単純な構造にもかかわらず、過去の全履歴に依存した出力を生じることを示している。

このような単純回帰型ネットワークの強力な性質を利用して、エルマンネットでは言語情報処理などへの応用が試みられてきた。

2 エルマンネットの処理能力

もっとも広い文法の階層は、チューリングマシン (0 型文法) と呼ばれている。その中に自然言語、さらにその中に「正規言語」がある。自然言語はチューリングマシン程広い必要はない。しかし、正規言語では狭すぎて人間の使う言語を説明できない。両者の中間に文脈依存文法 (1 型文法)、文脈自由文法 (2 型文法) という文法が存在する。自然言語が文脈依存文法であるか、文脈自由文法であるかは議論が分かれるところだが、自然言語はチューリングマシンと正規言語の中間に位置することは確かである。

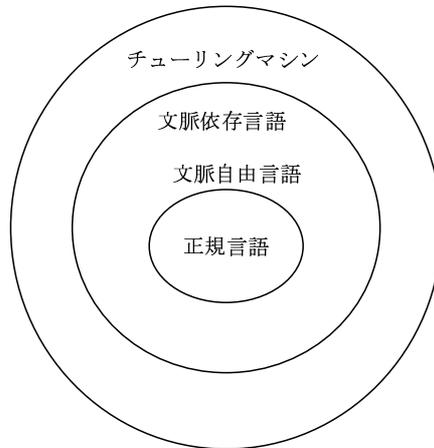
ちなみに文脈依存文法とは、左辺から右辺への書き換え規則で、右辺の長さが必ず左辺の長さ以上となっている場合、ある規則の適用が対象となる部分文字列の左側の文字列に依存するために「文脈依存」と呼ばれる。

文脈自由文法とは、左辺値から右辺への書き換え規則で、左辺が必ず非終端記号 1 個で表現されるものを指す。ある規則を適用する際に、対象となる部分文字列の左側の文字列に依存しいことから「文脈自由」と呼ばれる。

ここでは文献 (Rodrigues, Wiles & Elman, 1999) に従ってエルマンネットの形式的処理能力を説明する。

オートマトンの理論によれば正規言語は有限状態機械によって処理することができる。しかし、中央埋め込みのある言語は少なくとも文脈自由言語であり、その処理のためには少なくともプッシュダウンオートマトンであることが要請される。プッシュダウンオートマトンは埋め込みを保持 (追跡) するようにスタックあるいはカウンターのような付加的な記憶装置を持つ有限状態機械だからである。

時間的情報の複雑な構造的関係 — 文脈自由言語など — を学習することは、明示的な記憶装置を持たないシステムに対する深刻な要求であるが、再帰型



ニューラルネットワークは記憶装置に匹敵する能力を発揮することが示されている。事実、簡単な人工文法を用いて再帰型ニューラルネットワークを使った実験的研究では、再帰型ニューラルネットワークが正規言語の文字列を認識し、学習できることが示されている。

さらに、再帰型ニューラルネットワークは有限状態機械の計算 — プッシュダウンオートマトンにおけるスタック管理や push-pop 遷移 (Sun *et al.*, 1990)、チューリングマシンにおける read/write/shift 遷移を実装しうることが示されてきた。事実、再帰型ニューラルネットワークは、有限状態機械における状態と状態間の遷移を模倣するために、隠れ層ユニットで構成される空間の領域と領域間の遷移を使用できることが解析されている (Cleeremans, Servan-Schreiber & McClelland, 1989)。

2.1 エルマンネットの数理

エルマンネットを数学的に解析するためには、出力ベクトル \mathbf{y} に対して、文脈層からの結合係数行列を \mathbf{W} とし時刻 t における入力を $I(t)$ と表すことにすれば

$$\mathbf{y}(t) = f(\mathbf{W}\mathbf{y}(t-1) + I(t)), \quad (2)$$

と表すことができる。各ユニットへの入力は互いに独立であると考えれば、 \mathbf{W} の対角成分に $I(t)$ の対応する値を加えることで \mathbf{W} と $I(t)$ をまとめて $\mathbf{W}(t)$ と表すことができる。これにより、(2) 式は

$$\mathbf{y}(t) = f(\mathbf{W}(t)\mathbf{y}(t-1)), \quad (3)$$

と表現できる。

この式は離散状態の状態方程式だが、これを連続近似して関数 f が原点

近傍では連続近似できると考えれば

$$\frac{dy}{dt} = W(t)y \quad (4)$$

とすれば、エルマンネットの挙動は行列 $W(t)$ の固有値によって定まると言える。例えば、固有値が 1 より大きければ、対応する固有ベクトルの方向へ大きくなる。固有方程式が複素解を持つならばリミットサイクルなどのような解が得られることになる。

3 単語予測課題

エルマン (Elman, 1991) 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可能であることが示されている (Elman, 1990, 1991, 1993)。

表 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 が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 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。この文法規則から生成される文 S は、名詞句 NP と動詞句 VP と最後にピリオドから成り立っている。名詞句 NP は固有名詞 PropN か名詞 N が名詞に関係節 RC が付加したものの何れかである。動詞句 VP は動詞 V と名詞句 NP から構成されるが名詞句が付加されるか否か

は動詞の種類によって定まる。関係節 RC は関係代名詞 who で始まり、名詞句 NP と動詞句 VP か、もしくは動詞句だけのどちらかが続く。

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(実際には別の目的もあって 23 ビットではなく 26 ビット、すなわち 26 個のユニットが用意されていた)。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。実際にエルマンが用いたシミュレーションでは入力層と中間層の間にもう一層、中間層と出力層の間にもう一層の計 5 層のネットワークであったのだが中間層の数は、問題の本質でない。エルマンは表 1 に従って生成された文章を一単語ずつ徐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、ネットワークは次の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っていない点も強調される。

4 文字予測課題

エルマンら (Elman, Bates, Johnson, Karmiloff-Smith, Parisi & Plunkett, 1996; Elman, 1991) は

Manyyearsagoaboyandgirlivedbytheseatheyplayedhappily.

のような文章の区切りを見つけることをネットワークに要求した。ここでの入力は文中の 1 音素で、出力は次の 1 音素を予測するように訓練された。

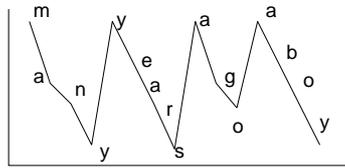


図 3: エルマンによる音素予測課題のエラー曲線

図からエラーは語頭で高く、語末まで減少している様子が分かる。エラー曲線を確信度と解釈すると、単語内の次にくる音素をかなり確信を持って予測していることが分かる。一方、入力が単語の終りに達すると次にどの語がくるか予想できないので、結果としてノコギリ状のエラー曲線になる。エラーの特徴については、実際観察される幼児の言葉の誤りと類似していて a boy を aboy と切り出してしまうオーバーシューティング (overshooting) や、they を the y とするアンダーシューティング (undershooting) のエラーが観察される。

5 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。にもかかわ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 (Rohde & Plaut, 1999)。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

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国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。

また、子どもが人称代名詞の知識を正しく獲得するためには自分以外の大人たちの間で交わされている会話を聞くことが重要だという実験事実とニューラルネットワークによるシミュレーション研究(百合子, 1997)も関係していると思われる。

6 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)(Allen & Seidenberg, 1999) 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4に示す。図中 clean up と書

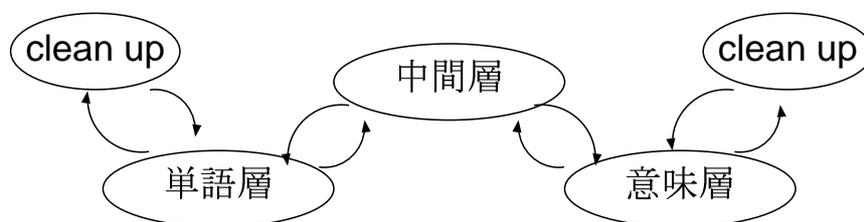


図4: 文章理解と文法判断のためのネットワーク図

いてあるユニット群は、エルマンネットの文脈層(図1)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。

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葉の持つ統計的構造)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。

7 文法知識の創発と失文法

7.1 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。このシルビウス列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 (Dronkers, 1996)。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。その証拠に自分が話した言葉をマイク

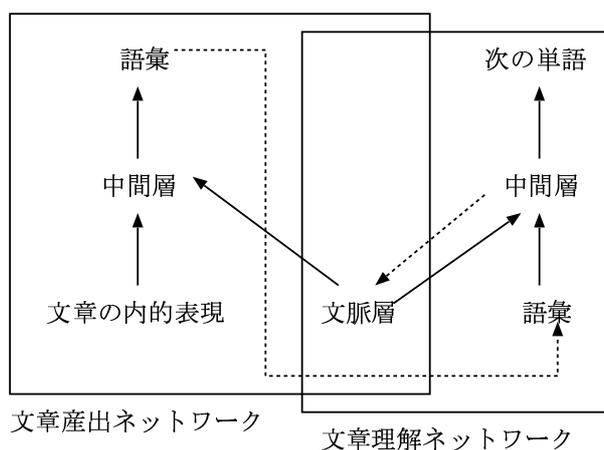


図 5: 2重経路単純再帰型ニューラルネットワークモデル

ロフォンで録音し、一定の遅延を置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 効果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、切り離すことは難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プ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 (Dell, Schwartz, Martin, Saffran & Gagnon, 1997; Dell, Chang & Griffin, 2001)。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 に二重単純再帰型ニューラルネットワークを示した。

7.2 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 6 である。入力表現

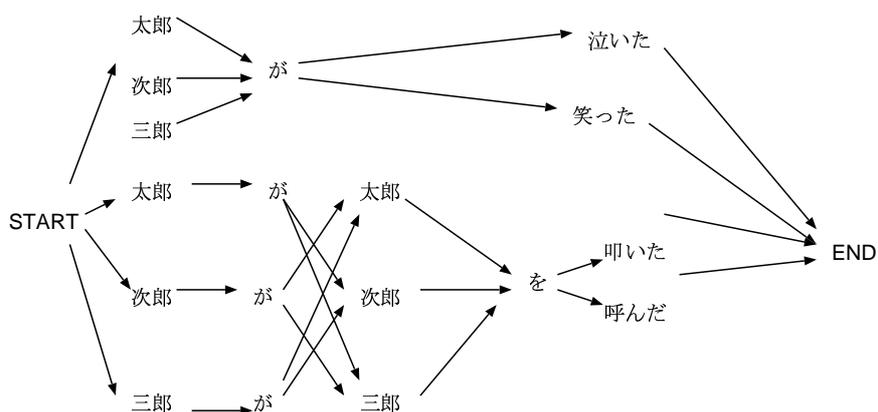


図 6: 二重単純再帰型ニューラルネットワークの訓練に用いた文章

は、主格太郎、主格次郎、主格三郎、目的格太郎、目的格次郎、目的格三郎、笑った、泣いた、呼んだ、叫んだ, 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

「太郎が泣いた」は

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 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。

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つ

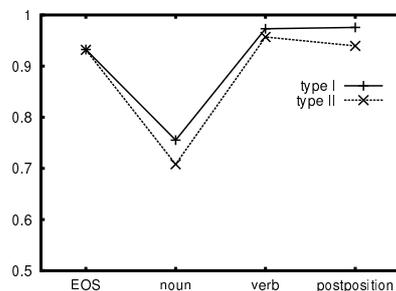


図 7: 失名辞タイプの損傷例

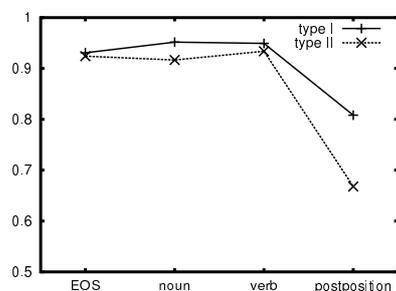


図 8: 形態的失文法タイプの損傷例

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の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。

7.3 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 7 と図 8 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

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。

7.4 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ため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 priori 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。

8 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。生成文法を前提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ようと主張されている (Elman, 1991; Elman, Bates, Johnson, Karmiloff-Smith, Parisi & Plunkett, 1996)。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 (Seidenberg, 1997; Seidenberg & MacDonald, 2001)。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。事実、乳児は明示的なガイダンスや報酬なしに保育者の発話の統計的な側面を自然にかつ自動的に符号化しているという証拠がある (Saffran & Richard N. Aslin, 1996)。そしてこのような学習は子宮内で始まっていると思われる。なぜなら新生児は母親の話す言語を好んで聞く傾向があることが実験的に確かめられている (Moon, Panneton-Cooper

& Fisher, 1993)。

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、英語を母国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 (Allen & Seidenberg, 1999)。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産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 (Seidenberg & MacDonald, 2001) のかも知れない。

エルマンの研究や統計的 (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 (Seidenberg, 1997; Seidenberg & MacDonald, 2001)。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 (Seidenberg, 1997)。

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。

文献

- Allen, J. & Seidenberg, M. S. (1999). The emergence of grammaticality in connectionist networks. In B. MacWhinney (Ed.), *The Emergence of Language* (pp. 115–151). Mahwah, NJ: Lawrence Erlbaum.
- Cleeremans, A., Servan-Schreiber, D. & McClelland, J. L. (1989). Finite

- state automata and simple recurrent networks. *Neural Computation*, 1(3), 372–381.
- Dell, G., Schwartz, M., Martin, N., Saffran, E. & Gagnon, D. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, 104.
- Dell, G. S., Chang, F. & Griffin, Z. M. (2001). Connectionist models of language production: Lexical access and grammatical encoding. In M. H. Chirstiansen & N. Charter (Eds.), *Connectionist Psycholinguistics* chapter 7, (pp. 212–243). Westport, CT: Ablex Publishing.
- Dronkers, N. F. (1996). A new brain region for coordinating speech articulation. *Nature*, 384, 159–161.
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14, 179–211.
- Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning* (pp. 195–225).
- Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition* (pp. 71–99).
- Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Simth, A., Parisi, D. & Plunkett, K. (1996). *Rethinking Innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press. (邦訳「認知発達と生得性」, 乾, 今井, 山下訳, 共立出版).
- Gold, E. M. (1967). Language identification in the limit. *Information and control*, 10, 447–474.
- Moon, C., Panneton-Cooper, R. & Fisher, W. (1993). Two-day-old infants prefer their native language. *Infant Behavior and Development*, 16, 495–500.
- 大嶋百合子 (1997). 言葉の意味の学習に関するニューラルネットワークモデル—人称代名詞の場合—. *心理学評論*, 40(3), 361–376.
- Rodrigues, P., Wiles, J. & Elman, J. L. (1999). A recurrent network that learns to count. *Connection Science*, 11(1), 5–40.
- Rohde, D. L. T. & Plaut, D. C. (1999). Language acquisition in the abasence of explicit negative evidence: How important is starting small? *Cognition* (pp. 69–109).

- Saffran, J. R. & Richard N. Aslin, E. L. N. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.
- Seidenberg, M. S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science* (pp. 1599–1603).
- Seidenberg, M. S. & MacDonald, M. C. (2001). Constraint satisfaction in language acquisition and processing. In M. H. Christiansen & N. Charter (Eds.), *Connectionist Psycholinguistics* chapter 9, (pp. 281–318). Westport, CT: Ablex publication.