計算機で脳の障害を再現できるか

浅川伸一
asakawa@twcu.ac.jp

2002 年 4 月 6 日

1　はじめに

編集部から頂いたテーマは「計算機で脳の障害は再現できるか」であったが、「言語」という雑誌の性格から「計算機で言語の障害は再現できるか」という問いのことであると解釈して、ニューラルネットワークによる言語の障害を扱った最近の研究動向を概説してみたい。

以下に紹介する研究では、特定の言語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であろう。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だろう。

本稿では、まず最近の研究動向を理解するのに必要だと思われる 3 つのキーワード、すなわち (1) ニューラルネットワークの特徴、(2) チョムスキー (Chomsky) の生成文法理論とニューラルネットワークモデルとの違い、および (3) 神経心理学における二重乖離の原則の再考、を概説する。その後ニューラルネットワークを用いた言語障害の研究を 2 つ紹介することにしたい。取り上げる話題は、失読症 (alexia) あるいは難読症 (dyslexia) と呼ばれる読みの障害の研究、文章理解 (sentence

comprehension) と失文法 (agrammatia) と呼ばれる障害の研究、の 2 つである。

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行なうものである。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。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。この意味において、ニューラルネットワークは「強い AI」[25] を主張している。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次の 3 点に要約できる [10, 19]

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 節でも述べたとおり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。4 節に示す二重乖離の原則への再考はもっとも大きな影響と言えるだろう。

3　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。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 [7, 9]。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 [26, 27]。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 [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 [27] のかも知れない。

6.1 節で説明するエルマン (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 [26, 27]。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。

エルマンの示したシミュレーションではいわゆる「刺激の貧困」「否定証拠の欠如」でも言語獲得が可能である点が重要である。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。このことは生後わずか 2 日の新生児でも母国語を好んで聞く傾向があるという実験結果 [16] からも支持されるであろう。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 [18] も関係していると思われる。

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 [26]。

4　二重乖離の原理再考

二重乖離の原則によって脳内の認知的機能局在が論じられる。例えば 5 節において説明するように、低頻度例外語の読みだけが選択的に障害される表層失読症と呼ばれる症状を呈する患者が存在する。これに対して、非単語が読めない音韻失読症と呼ばれる患者も存在する。表層失読症の患者は非単語を読むことができ、音韻失読症の患者は低頻度例外語を読むことができる。このようにある特定の認知機能 A だけに障害を示し、別の認知機能 B は正常に保たれている患者が存在し、逆に認知機能 B だけが選択的に障害され、認知機能 A は正常の範囲である患者も存在する場合、認知機能 A と 認知機能 B とは二重に乖離しているという。二重乖離の現象が観察された場合、それぞれの認知機能に対応する脳内機構が存在し、それら脳内の認知機構は他方の影響を受けずに独立に機能すると主張するのが二重乖離の原則である。二重乖離の原則は神経心理学における基本原理とされてきた。二重乖離の原則の背後には、脳内の認知機能は完全に独立して営ま

れており、この機能的に独立した単位は脳内の他の部位とは単純な情報を伝えあうだけであるという認知機能の局在性の強い仮定に基づいている [10]。

このような二重乖離を示す例は、上記の例外語と非単語の例ばかりでなく、日本語における仮名と漢字の乖離、意味記憶における生物と非生物の知識の乖離、具体語と抽象語の知識の乖離、自発話において名詞のみが損傷される失名辞失文法と助詞や冠詞が抜け落ちる機能的失文法との乖離、文章理解と文法性判断の乖離など、さまざまなものがある。こうした認知機能の独立性と脳の画像診断結果との対応を考察することが認知神経心理学における標準的な研究方法であった。

脳は複雑な相互作用をするシステムであるので、二重乖離の原則によって、ある認知機能の特別な要素と損傷の効果とを結びつけることは大変困難な作業だと考えられる。にもかかわらず、伝統的な神経心理学では患者の検査成績を文字通りの意味で解釈する傾向があった。例えば、視覚性の誤り (5.2 節参照) は、視覚処理段階における障害であると解釈されてきた。換言すれば、神経心理学的な解釈の特徴は、脳における認知機能を、相互作用をする複雑なシステムとしてではなく、個別的な働きをする機能的構造体の個別の働きとして解釈する傾向にあったのである。二重乖離の原則はきわめてナイーブな見方と言わざるを得ない。

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複雑な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な検証、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である。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、かつて信じられていた程明らかな関係にはないのである。

次節以降で紹介する研究は二重乖離の原則をニューラルネットワークの枠組でとらえ直し、新たな解釈を提供するものである。

5　失読症のシミュレーション

5.1　英単語の特徴

英語における読字過程のモデルとその破壊実験による失読症 (alexia) あるいは難読症 (dyslexia) と呼ばれる脳損傷のシ

ミュレーション研究を紹介する。

英語ではおなじ a でも make と cat では発音が異なる。a がどのように発音されるかはその単語の前後の文字によって定まる。その点、日本語の場合の平仮名やカタカナは表音文字であるので英語のような問題が無い。日本語の仮名はモーラを単位とする表記法であり、文字と音との対応関係に、少数の例外を除いて 1 対 1 対応が存在し、前後の文脈によって影響されることがほとんどない。一方、漢字については先の英語例と同じように単一文字だけでは読みが決まらない場合が多く、1 つの文字が前後の文脈によりいくとおりもの発音が存在する。英単語は、仮名と漢字の中間に位置すると考えられる。すなわち文字と音との間の対応関係が規則的である規則語 (regular words) と規則に従わない不規則語 (irregular words または exception words) とが存在する。多くの単語と読みを共有している単語をコンシステントな単語 (例えば hint, mint, print, lint) と言い、例外語をインコンシステントな単語と呼ぶ (例えば pint) ことがある。

単語の集まりを語彙 (レキシコン) と呼ぶが、多くの理論では脳内にこのレキシンコンが存在すると仮定している。レキシコンの分散表現を仮定するのがニューラルネットワークの発想である [28]。ニューラルネットワークを用いる言語研究における基本的なアイデアは単語は 3 つの要素 (あるいは脳内の部位と言っても良い) に分散表現されていると仮定することである。3 つの要素は書記素、音韻、意味と呼ばれる。書記素とは視覚提示された単語の脳内表現を指し、音韻とは音声単語の脳内表現、意味とは視覚や聴覚などの感覚入力に依存しない抽象的な表現である。

5.2 3 つの失読症

大きく分けて読みの障害には 3 種類ある。音韻失読症 (phonological dyslexia)、表層失読症 (surface dyslexia)、深層失読症 (deep dyslexia) 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し、非単語を書き取ることもできるし、復唱することもでるが、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。5.3 節で説明する二重経路モデルに照らし

て考えれば音韻失読は書記素から音韻へ変換 (直接経路という) の障害と考えることができる。

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。このことに加えて、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。また、視覚性の誤りを示すことがある。ときには上記二つの混ぜ合わさった読みの誤り、たとえば sympathy を orchestra と言ったりする (おそらく sympathy と綴りの似ている symphony を介して)。この種の患者は、抽象語と具体語の成績にも差があることがある。具体語である「椅子」よりも抽象語である「真実」の読みの成績が悪い。二重経路モデルによる深層失読の説明では、書記素から音韻への直接経路の障害の他に意味経路の障害が加わったとされる。

以下では古典的な読みのモデルである二重経路モデルを説明し、二重経路モデルへの批判から産まれたニューラルネットワークモデルであるトライアングルモデルを紹介する。トライアングルモデルを用いた労働の分割問題 (5.4 参照) によって、どのように表層失読と音韻失読と捉えられるかを説明する。

5.3 二重経路モデルからトライアングルモデルへ

コルトハートら [4, 5] によって開発された記号処理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単語ベースのルックアップテーブルをもつ間接経路 (語彙経路) とから構成されている。

さまざまな議論はあるものの、心理学における論争の一つは、単一経路か二重経路かという議論ではなく、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明示的で記号処理的なルールを用いることである。すなわち二重経路モデルでは 2 つの経路間の離散的なスイッチが仮定されている。

これに対して、トライアングルモデルではそのようなルールあるいはスイッチの存在を仮定しない。代わりに、トライ

アングルモデルでは 2 節で説明したような同時的、相互作用的処理が仮定される。トライアングルモデルにおける直接経路では、5.1 節で紹介したコンシステントな規則語と高頻度の不規則語が学習される。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。すなわちトライアングルモデルにおける直接経路では規則語と高頻度例外語が学習され、従って直接経路は単語の頻度効果に、すなわち単語の統計情報 (生起確率) に敏感である。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 (5.4 節) と呼ばれる作用が関与する。

5.4 トライアングルモデルによる読みの障害の説明

トライアングルモデルは、プラウト、マクレランド、サイデンバーグ、パターソン (Plaut, McClelland, Seidenberg, and Patterson) により開発された [28, 21]。この論文とモデルのことを PMSP96 ということがある。PMSP96 のモデルはもっとも成功したニューラルネットワークの一つであると言われている。

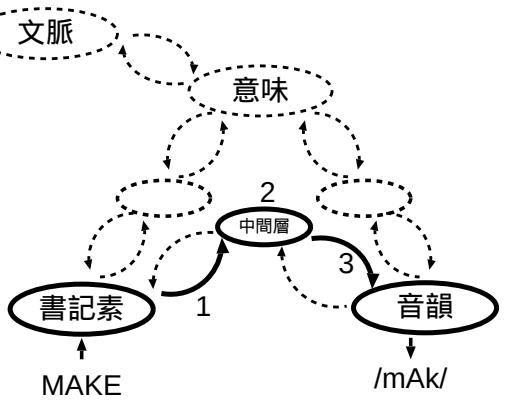

図 1: トライアングルモデル

トライアングルモデルでは英語の単音節単語約 3000 語を読むことができる。初期のトライアングルモデル [28] では、非単語を読むことができなかったため、人間の読みのモデルとしては適当でないという批判を受けていた。PMSP 96 では入力表現を工夫し、オンセット、母音、コーダという 3 つの部分から構成される母音中心表現と呼ばれる表現を用いている。

この入力表現の工夫により非単語の読みについても人間と同程度の能力があることが示された。この例のようにニューラルネットワークモデルではネットワークに与える入力表現にどのような工夫を凝らすかが重要である場合が多い。最近では PMSP96 で用いた母音中心表現を自動的に学習させる試みもなされている [17]。

トライアングルモデルでは、以下の心理実験結果を説明することができる。

1. 高頻度語の音読潜時は低頻度語の音読反応時間よりも速いこと
2. コンシステントな単語はインコンシステントな単語よりも音読反応時間が速いこと
3. 頻度効果との交互作用があること。コンシステント効果は低頻度語の方が効果が大きく、高頻度語ではコンシステントの効果が小さいこと

などである。これに加えて PMSP96 では、図 1 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。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患者の検査結果とも一致していることが示された。

トライアングルモデルはニューラルネットワークの特徴である相互作用 (2 節参照) をする。一旦書記素層へ入力された文字は、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。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、また、個人毎に異なるとされる。直接経路では規則語と高頻度語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 2 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。労働の分割による説明では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。

上記のようにトライアングルモデルによって、表層失読と音韻失読という二重乖離 (4 節) 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 PMSP96 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。

5.5 アトラクタネットによる深層失読のシミュレーション

失読症の最後に深層失読を扱ったモデルを紹介する。プラウト (Plaut) とシャリス (Shallice)[22] はアトラクタネットと呼ばれるニューラルネットワークを使って意味性の錯読や視覚性の誤りと意味性の誤りとの混ぜ合わさった誤り (5.2 節を参照) の説明を試みた。2 節で述べたとおりニューラルネットワークではユニットが相互に結合されている。この相互作用から生じる活性化パターン上で定義された空間の中で、ユニットの状態がある特定の状態に遷移することによって意味記憶の検索を表現する。すなわち単語が提示されてからその単語に対応する意味表現を検索するときに、ユニットの活性値が意味空間上で遷移する。任意の初期値からある領域へと引き込まれることをアトラクタに吸引されたと言い、各単語毎に吸引領域が異なると考えるのである。

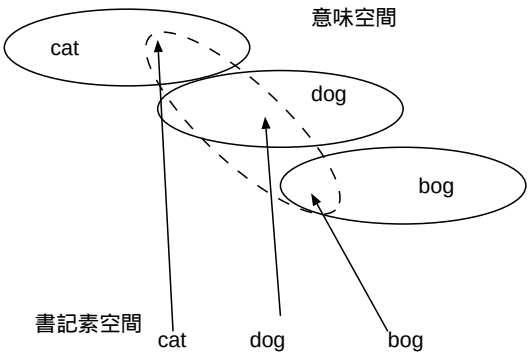

図 2: アトラクタネットによる意味記憶空間

図 2 は、損傷前の吸引領域を実線で、損傷を受けて吸引領域が変化した様子を点線で模式的に表現したものである。cat を dog という意味性の誤りに加え、bog を dog と言い間違える視覚性の誤りも起こり得ることを示している。なお、アト

ラクタネットの考え方は PMSP96 にも取り入れられている。

上記のようにプラウトらは意味記憶における障害によって深層失読の症状が再現できることを示した。彼らのシミュレーションでは、さらに意味記憶の構造を操作することで、具体語と抽象語との二重乖離の説明にも成功している [20, 19]。具体語における意味表現は抽象語の意味表現に比べてより多くの特徴を持っているとして意味記憶を構成し、損傷の程度が軽度ならば具体語よりも抽象語の方がより損傷の程度が大きく、損傷が重篤な場合には反対に具体語の方が損傷されやすいという結果を示した。ここでも二重乖離 (4 節) の原則が 1 つのモデルで説明できる可能性が示唆されている。

5.6 トライアングルモデルの拡張

最近のトライアングルモデルでは、発達性難読症への拡張 [11] もなされている。更に、失読症の読みの特徴に非常に時間がかかることが挙げられる。この逐語読み (英語では Letter By Letter Reading) という単語の音読に非常に時間がかかる現象を説明するために、6.1 節で説明する単純再帰型のニューラルネットワークを用いて音読時間を説明するように拡張されている [20, 19]。

6 文法判断と失文法のシミュレーション

6.1 エルマンネット

エルマン (Elman)[6] 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。

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。しかしこれら 2 点について

は否定的な証拠も提出されていて、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 [24]。

エルマンネットを簡単に紹介しておく。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 3)。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。

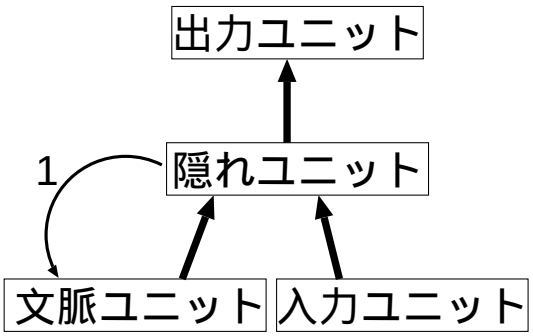

図 3: エルマンネット

エルマン [7] は、自身の考案したエルマンネットを用いて入力における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という。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表現されることになる。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 [6, 7, 8]。

6.2 エルマンネットの処理能力

ここで文献 [23] 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。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チューリングマシンにおけるリード/ライト/シフト 遷移など — を実装しうることが示されてきた。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 [3]。たとえばエルマン [7] は再帰型ニューラルネットワークが 3 段階までの中央埋め込み文の学習可能であることを報告している。もし付加的な意味論的制約があれば再帰型ニューラルネットワークはより多段の埋め込み文を学習できることが示されている ([27] の議論も参照してほしい)。

6.3 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 4 に示す。図中 clean up と書いてあるユニット群はアトラクタネット (5.5 節) でも用いられたもので、エルマンネットの文脈層 (図 3) 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

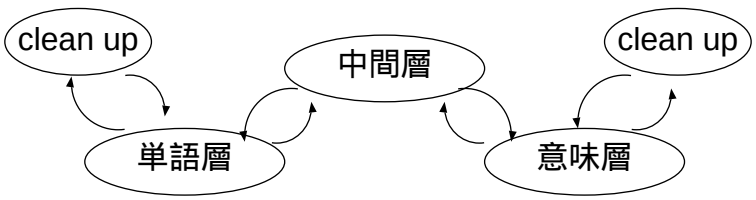

図 4: 文章理解と文法判断のためのネットワーク図

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 (6.1 節を参照) 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 (5.5 節参照) 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 (2 節) 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。

7　おわりに

本稿では、3 つのキーワードと 2 つの研究領域から脳の障害をコンピュータで実現する方法 — 最近のニューラルネットワークによる神経心理学的研究 — を概観した。

ニューラルネットワーク批判には 2 つの点を挙げることができる。一つはモデルが特定の現象を解釈するための post hoc なモデルにすぎないのではないかという点である。Post hoc なモデルであっても、今まで信じて疑われなかった理論 — いわばセントラルドグマとでもいうべき— 普遍文法や二重乖離の原理に対する alternative を提出し、問題の再考を迫っている点は評価されるべきであると考える。ある分野の科学的知識が深まるとは、絶えず alternative からの挑戦を受け、それに答える形で理論が洗練されて行くことなのだろう。

もう一つの批判はニューラルネットワークのプログラムの多くはおもちゃのように小さく、実際の脳と比べると著しく見劣りするという批判である。確かに、このスケーラビリティの問題は深刻であると言って良い。かつてミンスキー (Minsky)

とパパート (Papert)[15] が批判したとおり、蟻の体型をそのまま拡大して象の大きさにしたのでは、おそらく巨大な蟻は自分の体重を支えきれずに動けないであろう。地球の重力に反して体を支えるためには象のような太さの足が必要なのである。すなわち現在のニューラルネットワークプログラムを単に拡張しただけでは解決できない問題が存在するに違いない。しかし、このことはニューラルネットワーク研究を全否定することにはならないだろう。スケーラビリティの問題を解決するためにはどのような手法が有効であるのかを議論しゆくのはこれからの課題である。

解決の一部はすでに示されていると言って良い。ランダウアー (Landauer) とデュマス (Dumais)[12] は、百科辞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。なお、同じような発想から非負行列因子化 (NMF) と呼ばれる手法も注目を集めている [13]。NMF を説明するスペースがなくなってしまったが、簡単な説明が http://www.twcu.ac.jp/~asakawa/nmf/ あるので参照して頂きたい。

本稿で概観してきたような視点で考えれば、「計算機で脳の障害は再現できるか」という最初の問いには、イエスと答えたい。コンピュータ上に実装され実際に動作するプログラムが存在することは、他の理論にはない大きな特徴であろう。克服すべき点が多いにもかかわらず、ニューラルネットワークによる人工脳損傷の研究によってこれからますます多くの成果が期待できると考えている。

なお、紙面の都合上割愛せざるを得なかった話題も多い。特に意味記憶の構造について説明する余地が全く無くなってしまったことをお詫びしたい。関連した情報が文献 [2] にあるので参照して頂きたい。本稿についての感想、コメント、批

判などは電子メールで asakawa@twcu.ac.jp 宛に送って頂ければ幸いである。

参考文献

[1] J. Allen and M. 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] 浅川 伸一. 脳損傷とニューラルネットワークモデル — 神経心理学への適用例 —. In 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, chapter 5, pages 51–66. 北大路書房, 2001.

[3] A. Cleeremans, D. Servan-Schreiber, and J. L. McClelland. Finite state automata and simple recurrent networks. *Neural Computation*, 1(3):372–381, 1989.

[4] M. 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.

[5] M. 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.

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

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

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

[9] J. L. Elman, E. A. Bates, M. H. Johnson, A. Karmiloff-Simth, D. Parisi, and K. 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).

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

[11] M. W. Harm and M. S. Seidenberg. Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review*, 106(3):491–528, 1999.

[12] 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.

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

[14] J. L. McClelland, D. E. Rumelhart, and T. P. R. Group, editors. *Parallel Distributed Porcessing: Explorations in the Microstructures of Cognition.* MIT Press, Cambridge, MA, 1986.

[15] M. Minsky and S. Papert. *Perceptrons, Expanded Edition.* MIT Press, Cambridge, MA, 2 edition, 1988.

[16] C. Moon, R. Panneton-Cooper, and W. Fisher. Two–day–old infants prefer their native language. *Infant Behavior and Development*, 16:495–500, 1993.

[17] R. C. O'Reilly and Y. Munakata. *Computational Explorations in Cognitive Neuroscience: Understanding in mind by simulating the brain.* MIT Press, 2000.

[18] 大嶋 百合子. 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, 1997.

[19] D. C. Plaut. A connectionist approach to word reading and acquired dyslexia: Extension to sequential processing. In M. H. Chirstiansen and N. Charter, editors,

Connectionist Psycholinguistics, chapter 8, pages 244–278. Ablex Publishing, Westport, CT, 2001.

[20] D. C. Plaut, J. L. MaClelland, and M. S. Seidenberg. Reading exception words and pseudowords: Are two routes really necessary? In J. P. Levy, D. Bairaktaris, J. A. Bullinaria, and P. Cairns, editors, *Proceedings of the Second Neural Computation and Psychology Workshop*. University College London Press, London, 1995.

[21] D. C. Plaut, J. L. McClelland, M. S. Seidenberg, and K. Patterson. Understanding normal and impaired word reading: Computational principles in quasi–regular domains. *Psychological Review*, 103(1):56–115, 1996.

[22] D. C. Plaut and T. Shallice. Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology*, 10(5):377–500, 1993.

[23] P. Rodrigues, J. Wiles, and J. L. Elman. A recurrent network that learns to count. *Connection Science*, 11(1):5–40, 1999.

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

[25] J. R. Searle. Minds, brains, and programs. In D. R. Hofstadter and D. C. Dennett, editors, *The Mind's I — Fantasies and Relections on Self and Soul*, chapter 22. Basic Books, 1980. /サール「心・脳・プログラム」. ホフスタッター, デネット編著 . 坂本百大監訳「マインズ・アイ (下) —コンピュータ時代の心と私—」. 第 22 章 p. 178–210. TBS ブリタニカ . 1992.

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

[27] M. S. Seidenberg and M. C. MacDonald. Constraint satisfaction in language acquisition and processing. In

M. H. Christiansen and N. Charter, editors, *Connectionist Psycholinguistics*, chapter 9, pages 281–318. Ablex publication, Westport, CT, 2001.

[28] M. S. Seidenberg and J. L. McClelland. A distributed, developmetal model of word recognition and naming. *Psychological Review*, 96(4):523–568, 1989.

(東京女子大学/心理学)