階層型ニューラルネットによる語彙的曖昧性の解消

高橋 直人†

本稿では，階層型のニューラルネットを用いて日本語形態素解析時に生じる語彙的曖昧性を解消する実験およびその結果について述べる．ここで使用されているニューラルネットは，先行する自立語の系列を文脈情報として受け取り，その文脈において適切と考えられる単語を出力する．実験に用いたニューラルネットは，単純なフィードフォワード型と，フィードバックループを含むElman型との2種類である．入力層と出力層においては単語の局所表現，すなわち1ユニットが1単語を表現するような表現を採用した．実験の第1段階では，先行する自立語の系列を与えたとき，それに後続する単語と後続しない単語を正しく区別するようニューラルネットに学習を行わせた．データとして用いたのは，実際の出版物に現れた108文である．この結果，いずれの型のニューラルネットにおいても，10個の隠れユニットがあれば上の108文の学習には十分であることがわかった．実験の第2段階では，学習の終了したニューラルネットに対し，別の出版物からの33文を入力することで，獲得された単語選択能力がどの程度一般化されうるかを調べた．結果として，どちらの型のニューラルネットも語彙的曖昧性の生じた552か所のうち90%以上の局面で正しい単語を選択することが確認された．

Lexical Disambiguation by Layered Neural Networks

NAOTO TAKAHASHI †

In this paper, we describe an experiment of layered neural networks that we have done to disambiguate lexical ambiguity which arises in Japanese morphological analysis. The neural networks used here (one is simple feedforward type and the other is so called Elman type) accept the sequence of preceding conceptual words as context, and output the words which seem suitable for that context. In the input- and output-layer of the networks, we adopted local representation (i.e. each unit represents one word). In the first stage of the experiment, neural networks were trained to distinguish the words which follow the preceding words from the words which do not. The 108 sentences, which were taken from a real publication, were used to train the networks. This experiment showed that 10 hidden units were enough to learn all of the training sentences. In the second stage of the experiment, the generalisation ability of the trained neural networks was examined by giving them 33 sentences taken from another publication. Both type of neural networks gave correct words in more than 90% of the 552 ambiguous situations.

1. はじめに

通常の日本語文を表記する場合は，単語と単語の間に空白を置かないのが普通である．したがって計算機を用いた日本語処理においては，文を単語単位に分解する過程，すなわち形態素解析が重要になってくる．形態素解析を実行する際には，語彙的曖昧性を常に考慮しなくてはいけない．複数の候補の中から正しい単語を決定するためには，選択しようとしている候補と他の単語との意味的関係，および現在の文脈を正しく理解する必要がある．また効率の点からは，すべての候補を同等に扱うよりも，正解となる見込みの高い単語から順に調べていく方が望ましい．この際，見込みのある単語を発見するための方法が問題になる．

本稿では，文脈が与えられたとき，その文脈と関連の強い単語を選択するニューラルネットについて述べる．語彙的曖昧性が生じた場合，文脈と関連の強い単語は，そうでない単語よりも正解となる可能性が高いと考えられる．したがって，このニューラルネットを日本語形態素解析システムに組み込むことで，形態素解析の効率向上が期待できる．

次章以降の構成は次のとおりである．まず第2章で，文脈と関連の強い単語の選択にニューラルネットを用いた理由と，その適用範囲について述べる．第3章では，実際の出版物を対象にして，文脈と単語の関連をニューラルネットに学習させる実験について説明し，実験結果を示す．続く第4章では，ニューラルネットの一般化能力を調べる．これは，第3章の実験を通じ

† 電子技術総合研究所
Electrotechnical Laboratory

て文脈と単語の関連を学習させたニューラルネットに，学習に用いたのとは別の他の出版物の文章を適用することで行う．第5章では拡張性に関して議論する．最後に第6章で全体をまとめ今後の展望を述べる．

2. 文脈とニューラルネット

2.1 文脈処理に要求される能力

ここではまず，文脈と関連の強い単語を選択する際に要求される能力を列挙し，次にニューラルネットがその条件を満たすものであることを示す．

文脈と関連の強い単語を選択する際に要求される能力としては，以下の3種類が考えられる．

[パターン的情報処理能力]

与えられた文脈と関連の強い単語を選択するためには，まずその文脈を何らかの形で表現する必要がある．一口に文脈と言っても，そこには無限のバリエーションがあり，それらすべてをあらかじめ列挙することは不可能である．

また，文脈同士を比較した場合，互いに似ているものもあれば，そうでないものもある．もし文脈Aと文脈A'が互いに似ているならば，文脈Aと関連の強い単語の集合aと，文脈A'と関連の強い単語の集合a'には共通の要素が多く存在すると考えられる．したがって，互いに似た文脈に対しては似た処理を行うべきである．

このように文脈には，1)無限のバリエーションが存在し，2)各々の間に類似度を考えることができる，というアナログ的特徴がある．アナログ的な情報には，シンボル的な表現よりもパターン的な表現が適している．したがって，文脈を処理するシステムは，パターン的な情報の操作に適したものが望ましい．

[帰納的学習能力]

既に述べたように，文脈のバリエーションは無限に存在する．したがって，各々の文脈に対し，それと関連の強い単語を，人間の手によって先見的に与えることはできない．もし，考慮の対象とする領域を厳密に制限するのであれば，その中に現れうる文脈を制限することは可能であろう．しかしその場合でも，労力あるいは結果の首尾一貫性という点を考えると，人間が内省に基づいて文脈と単語を関連付ける方法は好ましくない．文脈と単語の関連付けは，解析システム側が，1) 実際の例文から，2) 統計的な操作に基づいて，行うべきである．このためには，例文からの帰納的学習能力を備えたシステムが必要である．

[一般化能力]

システムが帰納的学習能力を備えていたとしても，それだけではまだ不十分である．無限に存在する文脈の全パターンを学習することは到底不可能であるから，学習していない文脈が入力されることも当然ありうる．未知の文脈が入力された際にシステムがとるべき行動は，既に学習したパターンの中からそれに類似したものを捜し出して外挿することであると考えられる．このためには，獲得した知識を一般化する能力が必要とされる．

以上，パターン的情報処理能力・帰納的学習能力・一般化能力，の3種類の能力の必要性を述べたが，これらはいずれも，ニューラルネットが得意とする分野である．ニューラルネットを用いた場合，文脈をユニットの活性値パターンとして表現することができる(パターン的情報処理能力)．また，学習すべき文脈と，その文脈に関連の強い単語とを入出力ペアとしてニューラルネットに提示することで，両者の関係を帰納的に学習させることが可能である(帰納的学習能力)．さらにニューラルネットには獲得した知識を一般化する能力が備わっているので，未知の文脈に対しても適切な行動をとるものと期待できる（一般化能力)．

2.2 階層型ニューラルネットの採用とその適用範囲

これまでにもニューラルネットを用いた語彙的曖昧性解消の研究（形態素解析を含む）はいくつか行われてきたが，それらの中では相互結合型のニューラルネットを利用するものが比較的多い[1)~4)]．一方，今回の実験で使用したのは，階層型のニューラルネットである．階層型を採用した理由は以下のとおりである．

- 理論的な研究が広く行われており，その挙動に対する定性的・定量的な解析が進んでいる．
- バックプロパゲーション[5)]という強力な学習法が存在し，しかもそれを高速化するための手法が色々と開発されている．
- 活性値の伝搬が一方向に決定されており，データ入力後，比較的短時間で出力を得ることができる．そのため実用性が高い．

また，今回作成したニューラルネットは，日本語形態素解析システムの一部に組み込まれるような使われ方を想定している．そのため，単語の表記が解析すべき入力文字列とマッチするかどうかのチェック，および単語と単語との間の品詞的接続性のチェックは形態素解析システムにまかせ，ニューラルネット側では処理しないことにした*．文字表記的・品詞接続的なチェ

* 文字表記的・品詞接続的なチェックまで含めてニューラルネット上で実行しようという試みもある[6)]．

ックというものは適格か不適格かが2値ではっきりと決定できる種類のものであるから，特にニューラルネットを使わなくても，従来のシンボル的なアプローチ（すなわちLispあるいはPrologといった言語によるチェック）で十分対応できるというのがその理由である．

3. 文脈的整合性の学習

3.1 学習用例文

自然言語処理システムの有用性を評価する際には，意図的に作った例文ではなく，実際の言語データを用いることが重要である．文脈にふさわしい単語を選択するという今回の目的を考慮すると，対象としてはある主題について客観的に叙述した文章，たとえば解説文等が適していると考えられる．また，通常の日本語という観点からすると，数式や化学式が頻出する文は好ましくない．そこで今回の実験には，歴史上の事件に関する説明文（対象はフランス革命とその周辺）を使用することにした．ニューラルネットを訓練するための学習用例文としては，

村川，江上，山本，林：詳説世界史（再訂版），山川出版社（1982）

の中から「第12章 市民社会の成長 §2. フランス革命とナポレオン」(pp. 221-231) に出てくる文章を使用した．この中に含まれる総文字数は5708，異なり単語数は871，文の数は108である．

ニューラルネットに提示するための学習パターンは，上記の文章から以下の手順で作成した．

学習パターン生成手順

[step 1] 学習用例文のすべてを手作業で単語単位（厳密には形態素単位）に分解する．

[step 2] 分解されたすべての学習用例文に対し，step 3からstep 4までを適用する．

[step 3] 選択された文の中に含まれる総単語数をNとするとき，$2 \leq n \leq N$のnに対してstep 4を繰り返す．

[step 4] 与えられたnに対して，1) 先行自立語列，2) 後続単語集合，3) 非後続単語集合，からなる3つ組を作る．**先行自立語列**とは，文の先頭から$n-1$番目までの単語のうち，自立語類のみを出現順に並べたものである．**後続単語集合**とは，文の先頭から数えてn番目の単語を唯一の要素とする集合である．**非後続単語集合**とは，形態素解析システムの辞書に登録されている単語の中で，文字表記的・品詞接続的に$n-1$番目の単語に後続する可能性のあるものすべてから，後続単語集合の要素を除いたものからなる集合である（図1）．

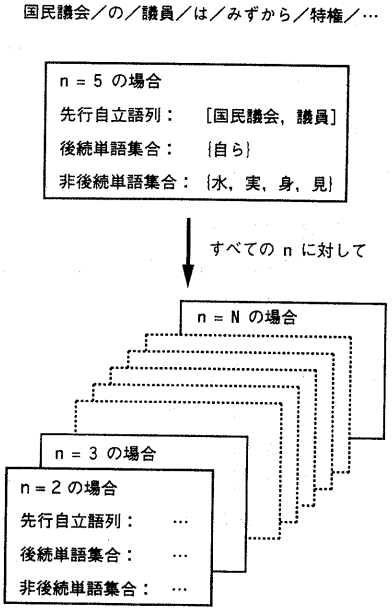


図1 先行自立語列・後続単語集合・非後続単語集合の作成
Fig. 1 Generation of Preceding Word Sequence, Following Word Set and Non-following Word Set from a sentence.

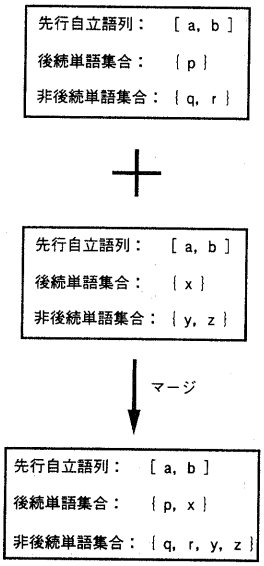


図2 同一の先行自立語列を持つ3つ組をマージする
Fig. 2 Merging triples that have the same Preceding Word Sequence in common.

1)．先行自立語列はその中の単語の出現順序が意味を持つが，後続単語集合および非後続単語集合は集合であるから要素の順序は意味を持たない．なお，品詞的接続性の検査は$n-1$番目の単語とn番目の単語の間のみで行っている．

[step 5] 得られたすべての3つ組のうち，同じ先行

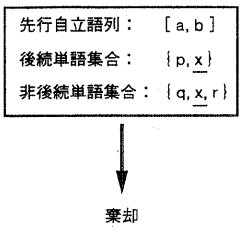


図3 矛盾した3つ組を捨てる
Fig. 3 Discarding contradictory triples.

入力文1：「1812年ナポレオンはロシア遠征軍をおこし、…」
先行自立語列1：［1，8，1，2，ナポレオン，ロシア，遠征，軍］
後続単語集合1：｛興｝
非後続単語集合1：｛…，起こ，…｝

入力文2：「13年、ロシア・プロシア・オーストリアの同盟軍は解放戦争をおこした。」
先行自立語列2：［1，3，ロシア，プロシア，オーストリア，同盟，軍，解放，戦争］
後続単語集合2：｛起こ｝
非後続単語集合2：｛…，興，…｝

入力文3：「…、革命暦・統制経済・理性崇拝の宗教をつくるなど、各方面に革新的な施策をおこなった。」
先行自立語列3：［…，革命，暦，統制，経済，理性，崇拝，宗教，作，方面，革新，施策］
後続単語集合3：｛行な｝
非後続単語集合3：｛…，起こ，興，…｝

図4 先行自立語列に基づいた後続単語選択の例
Fig. 4 Examples of following word selection based on Preceding Word Sequence.

自立語列を持つものをマージする（図2）.

[step 6] マージの終了した3つ組のうち，後続単語集合の要素と非後続単語集合の要素に重複があるもの*を捨てる（図3）.

こうして得られた3つ組の学習パターンは，全部で1718組になった．すぐ後で述べるように，以下の実験では，この3つ組の中の先行自立語列を文脈情報として用いた．これは文脈を「同一文内で先行する自立語の系列」と定義したことを意味する．

3つ組の具体例を図4に示す．四角い枠で囲まれた各々は，入力文と書かれた文字列のうち，下線を引いた部分の直前まで解析が進んだときの先行自立語列・後続単語集合・非後続単語集合を表している．この3文はいずれも学習用例文の原テキストに現れた例であるが，動詞部分が平仮名で表記されているため（このテキストでは和語動詞はしばしば平仮名表記されている），語彙的曖昧性が生じている．

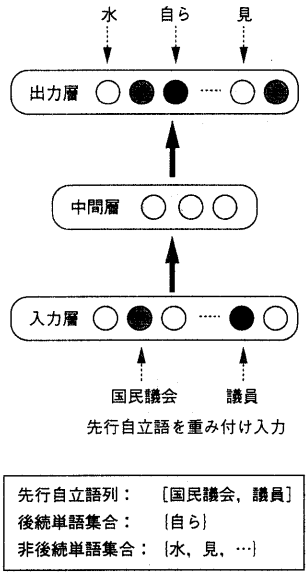


図5 フィードフォワード型ニューラルネットの構成
Fig. 5 Structure of a feedforward neural network.

この3例は，いずれも直前の単語が格助詞「を」であるので品詞的接続条件が同一であり，また文字表記的にも最初の2文字「おこ…」が同一であるが，先行自立語列が異なり，それに伴って後続単語集合および非後続単語集合が異なる，という例になっている．

入力文1の場合，動詞「興す」の語幹「興」は後続単語集合に属し，動詞「起こす」の語幹「起こ」は非後続単語集合に属している．一方，入力文2の場合は，これとは反対に「起こ」が後続単語集合に，「興」が非後続単語集合に属している．また入力文3の場合は，「興」も「起こ」も非後続単語集合に属している．

3.2 フィードフォワード型ニューラルネットによる学習

最初の学習実験は，フィードフォワード型の3層ニューラルネット（図5）上で行った．ニューラルネットが学習すべきタスクは，ある先行自立語列が入力されたとき，それに対応する後続単語集合と非後続単語集合を出力することである．以下，このニューラルネットの各部分について解説する．

入力層

入力層の各ユニットは，それぞれが辞書（単語数は6619）中の単語と1対1に対応している．すなわち単語は局所表現されている．

入力層には先行自立語列中の単語の重み付き和を与える．具体的には，先行自立語列中の最後の単語に対

* 今回の実験では全部で10組が該当した．

応するユニットには4.0，その1つ前の単語に対応するユニットには3.5，さらにもう1つ前の単語に対応するユニットには3.0，のように0.5きざみで減少する値を，それぞれの単語に対応するユニットに入力する．先行自立語列中で最後の8語よりも前の単語に対応するユニット，および対応する単語が先行自立語列に含まれないユニットへの入力はすべて0とする．直観的には，現在解析しようとしている部分よりも前にあって，位置的に近い自立語に対応するユニットほど大きな値が入力されると考えればよい．なお，入力層のユニットの入出力関数は恒等関数である．すなわち入力された値がそのまま出力される．

出力層

出力層における単語表現も，入力層と同様の局所表現である．今，入力層に与えられた先行自立語列と同じ3つ組に属する後続単語集合と非後続単語集合を，それぞれS^+，S^-で表わすことにする．このとき出力層における各ユニットの理想出力値は以下のようになる：

- S^+の各要素に対応するユニットは1.
- S^-の各要素に対応するユニットは0.
- S^+にもS^-にも属さない単語に対応するユニットは何を出力してもかまわない*.

出力層におけるユニットの入出力関数としては，標準的なsigmoid関数を採用した．

中間層

中間層（隠れ層）のユニットは，その数を10，30，50と3通りに変えて実験を行った．入出力関数は，出力層と同様にsigmoid関数である．

リンク

入力層と中間層，および中間層と出力層との間はリンクによって全結合されている．入力層と出力層を直接つなぐリンクはない．各リンクの初期値は，$[-0.01, +0.01]$の間の乱数である．この値は試行ごとに設定し直される．

このニューラルネットに，3.1節で作成した3つ組の学習パターンを学習させた．学習パターンを1つ提示するたびに，バックプロパゲーションに基づいてリンクの重みを更新した．ただし，出力層のユニットのうち，対応する単語が後続単語集合にも非後続単語集合にも属さないもの（出力値が何であってもよいとされたユニット）に関しては出力誤差を0とみなし，そのユニットに向かうリンクに関しては重みを更新しないものとした．それ以外のユニットが実際に出力する値と，理想的な出力値との間の許容誤差（tolerance）は，0.5, 0.3, 0.1の3通りに変化させて実験を行った．また学習の係数ηの値は0.1，慣性αの値は0.9に固定した．

結果的に，図5のニューラルネットはすべての学習パターンを学習することができた．学習パターンの中には，文字表記的・品詞接続的条件が同一であっても，先行自立語列が異なるために，別の後続単語集合・非後続単語集合を出力し分けなければならない例（たとえば図4に示した例）が含まれている．こういった例が学習できたことから，ニューラルネットには与えられた文脈に応じて正しい後続単語を選択する能力があることがわかる．

次に，許容誤差および中間層のユニット数を変化させたとき，それによって学習に要するステップ数がどのような影響を受けるかを調べた．ここで1ステップとは，ニューラルネットにすべての学習パターンを1回ずつ提示する過程を意味する．リンクの初期値はランダムに設定されるため，たとえ他のパラメータが同一であっても，学習に要するステップ数は試行ごとに変化する可能性がある．そのため，中間ユニット数と許容誤差の組み合わせごとに10回ずつ測定を行い，上位5組とその平均を求めた．結果を**表1**に示す．

表1の中で学習に要する平均ステップ数が最も少ないのは，中間層のユニット数が30のときである．中間層のユニット数を10あるいは50に設定した場合は，より多くのステップ数が必要となっている．この結果

表1 フィードフォワード型ニューラルネットが学習パターンの習得に要したステップ数
Table 1 Steps required by feedforward neural networks to learn the learning patterns.

中間層中のユニット数	許容誤差	ステップ数					
		試行1	試行2	試行3	試行4	試行5	平均
10	0.1	174	159	213	191	230	193.4
	0.3	137	137	108	154	129	133.0
	0.5	93	108	106	107	106	104.0
30	0.1	86	78	91	82	93	86.0
	0.3	65	58	60	60	61	60.8
	0.5	62	66	65	56	61	62.0
50	0.1	137	167	124	148	119	139.0
	0.3	153	125	127	138	107	130.0
	0.5	123	94	125	91	113	109.2

* 後続単語集合にも非後続単語集合にも属さないということは，文字表記的・品詞接続的な制約からその単語が選択されることがありえないということを意味する．したがって，もしこれら不適切な単語に対応するユニットが1を出力することがあっても，それらはニューラルネット外の形態素解析システムが行うシンボル的なチェックで排除することが可能であるので問題は生じない．

は次のように解釈できる：

「ニューラルネットの学習は，各リンクの重みを変更することによってなされるが，中間層のユニット数が10の場合はリンクの数が相対的に少ないため，与えられたタスクを実現する際の選択肢が限られる．比喩的に言えば，持ち駒が少ないために打てる手が限定されることになる．したがって学習は難しくなり，必要とされるステップ数が増加する．

中間層のユニット数を30に増加させると，それにしたがってリンクの総数も増加する．そのために，タスクを実現する際の選択肢が増加し，学習は比較的容易になる．

中間層のユニット数をさらに増やして50とすると，選択肢はさらに増加するが，冗長なユニットおよびリンクの数が多くなり過ぎるため，全体として見ると学習は再び困難になる．そのため学習に必要なステップ数は再び増加する．」

中間層のユニット数が同一のニューラルネット同士を比較した場合は，許容誤差が小さいときほど，学習に要する平均ステップ数が多くなる傾向にあることがわかる(反例は中間層ユニット数が30で，許容誤差が0.3と0.5の場合のみである)．この結果はきわめて自然と言える．なぜなら，すべての出力ユニットの許容誤差が0.1になるまでには，それが0.3になった状態を通過する必要があり，さらにすべての許容誤差が0.3になるまでには，それが0.5の状態を通過する必要があるからである．

3.3 フィードバック型ニューラルネットによる学習

3.2節で述べたフィードフォワード型ニューラルネットにおける先行自立語列の重み付けは，試行錯誤の結果決定されたものである．したがって，4.0から始まり0.5ずつ減少する重み付けが最良のものである，という保証はまったくない．入力表現を，先行する自立語の重み付き和の形とした理由は，時間方向への広がりを持った情報を，一度にまとめて与える必要があったからである．

もしニューラルネット自体が記憶を持つならば，先行自立語列を重み付き和の形に変換する必要はなくなる．ニューラルネットに記憶を持たせるためには，その中にフィードバックループを導入すればよい．そこでフィードバック型ニューラルネットを用いて前節と同様の実験を行い，先行自立語列の重み付けに相当する機能をニューラルネット自身に獲得させることを試みた．使用したニューラルネットはElman型[7]として知られるものである(**図6**)．入力層および出力層における単語の表現には，フィードフォワード型の場合と同様，局所表現を採用した．

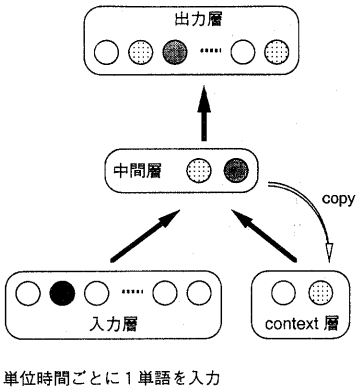


図6 フィードバック型ニューラルネットの構成
Fig. 6 Structure of a feedback neural network.

フィードバック型ニューラルネットへの入力としては，先行自立語列中の単語を先頭から順に1語ずつ，1単位時間ごとに与えた．すなわち，時刻 i においては，先行自立語列中で前から i 番目の単語に対応するユニットのみに1を入力し，それ以外のユニットには0を入力した．

中間層の出力パターンは各時刻ごとにcontext層にコピーされ，次の時刻において（入力層からの新たな出力とともに)中間層への入力となる．Context層の各ユニットの入出力関数はsigmoid関数，初期値は0.5とした．中間層からcontext層へのフィードバックループは，対応するユニットごとの1対1結合であるが，context層から中間層へのリンクは全結合である．

その他，出力層における目標出力値の設定，各ユニットの入出力関数，リンクの重みの初期値および更新の方法，学習係数 η，慣性 α，訓練に使用した学習パターン等はすべてフィードフォワード型の場合と同一である．

フィードフォワード型の場合と同様，フィードバック型のニューラルネットもすべての学習パターンを正しく学習することができた．ただし，中間層のユニット数を50に設定した場合，学習はしばしば非常に不安定になり，数万ステップが経過した後でも終了しないことがあった．このため，フィードバック型ニューラルネットに対しては，中間層のユニット数が10の場合と30の場合のみ実験を行った．

中間層のユニット数および許容誤差を変化させ，そのときに学習パターンの習得に必要とされるステップ数がどのように変化したかを**表2**に示す．フィードバック型はフィードフォワード型と比較し，学習により

表 2 フィードバック型ニューラルネットが学習パターンの習得に要したステップ数
Table 2 Steps required by feedback neural networks to learn the learning patterns.

中間層中のユニット数	許容誤差	ステップ数					
		試行1	試行2	試行3	試行4	試行5	平均
10	0.1	1051	888	1688	1559	859	1209.0
	0.3	516	770	638	798	568	658.0
	0.5	656	573	318	536	689	554.4
30	0.1	791	355	309	399	405	451.8
	0.3	453	425	219	216	409	344.4
	0.5	149	209	289	192	242	216.2

多くのステップ数を必要とすることがわかる．また，ステップ数にばらつきが大きい（すなわち学習が不安定である）ことも読み取れる．中間層のユニット数が同一ならば，許容誤差が小さい場合ほど多くのステップ数が必要となるのはフィードフォワード型の場合と同様である．

フィードバック型ニューラルネットにおける学習は，フィードフォワード型ニューラルネットにおける学習に比較すると，より難しい問題であるといえる．第1に，フィードフォワード型では先行自立語の重み付けが先見的に与えられるが，フィードバック型ではそれに相当する機能をニューラルネット自身が発見しなければならない．第2に，フィードフォワード型では中間層へ入力される個々の値の範囲が0以上4以下であるのに対し，フィードバック型のそれは0と1の間（両端を含まない）である．すなわち，フィードバック型の中間層に与えられるダイナミックレンジはより狭く，そのために各入力パターンの弁別はより困難になっている．フィードバック型ニューラルネットにおける学習の不安定さは，これらが原因であると考えられる．

4. 一般化能力のテスト

4.1 テスト用例文およびテスト手順

次に，学習の完了したニューラルネットが，単語選択に関してどのような一般化能力を示すかテストした．テスト用例文に求められる条件は，学習用例文と同一分野の文章で，しかも同一の文を含まないことである．そこで今回は，

歴史教育研究会編：世界史事典，旺文社（1992）の中から「フランス革命（p. 357）」「ナポレオン（1世）（p. 296）」（ウィーン会議（p. 47）」の3項目の解説文（漢字仮名交じり）をテスト用例文として採用した．総文字数は1902，異なり単語数は382，文の数は33である．なお，この中には，学習用例文に含まれていない単語が104語含まれている．

テストでは，このテスト用例文から作成した先行自立語列を，学習の済んだニューラルネットに入力し，そのとき正しい後続単語が出力されるか否かを調べた．詳しい実行手順は以下のとおりである．

テスト手順

[step 1] テスト用例文中の各文をあらかじめ手作業で単語列に分割しておく．

[step 2] すべてのテスト用例文に対し，step 3からstep 7を適用する．

[step 3] 選択された文の中に含まれる総単語数をNとするとき，$2 \leq n \leq N$のnに対してstep 4からstep 7を繰り返す．

[step 4] 文の先頭から解析が進み，$n-1$番目までの単語が確定したと仮定する．また，それ以降の部分に関しては文字列のみが与えられており，単語境界すらわかっていないものとする．このとき，もしn番目の単語の決定において語彙的曖昧性が生じないならば*，step 5からstep 7までをスキップして次のnに進む．そうでなければ，1番目から$n-1$番目の単語をもとに先行自立語列を作成し，学習時と同様の方法でニューラルネットに入力する．

[step 5] 中間層，出力層の順に各ユニットの出力値を計算する．

[step 6] 文字表記的・品詞接続的に$n-1$番目の単語に後続しうる単語に対応する各出力ユニット中から，出力値が最大のものを選択する．出力値が同一の場合は文字列での最長一致を優先する．

[step 7] step 6で選択されたユニットが，実際の入力文中のn番目の単語に対応したユニットであれば正解とし，そうでなければ選択誤りとする．

前節の実験で用いたものと同一の辞書（単語数6619）を使用した場合，テスト用例文中で語彙的曖昧性の解消が必要になった回数，すなわちstep 5からstep 7が実行された回数は552回であった．なお，上の手順からわかるとおり，このテストでは一度誤った単語を選択しても，それが更なる後続単語の選択に及ぼす影響については考慮していない．得られるテスト結果はあくまでも，「正しい先行自立語列が与えられた

* 文字表記的・品詞接続的チェックを行った結果，$n-1$番目の単語に後続しうる単語が辞書中に1語しか見つからなかった場合がこれに該当する．テスト用例文に含まれるすべての単語はあらかじめ辞書に登録されているので，最低1語が見つかることは保証される．学習パターン生成手順の場合と同様に，品詞接続性の検査は$n-1$番目の単語とn番目の単語のみで行っている．

表 3 テストデータに対するフィードフォワード型ニューラルネットの正答率
Table 3 Percentage of correct answers given by feedforward neural networks.

中間層中のユニット数	許容誤差	正答率（%）					
		試行1	試行2	試行3	試行4	試行5	平均
10	0.1	91.3	90.6	91.4	92.2	91.1	91.3
	0.3	91.8	92.0	90.8	90.9	91.3	91.4
	0.5	91.1	91.4	91.1	91.7	91.4	91.4
30	0.1	90.0	90.8	90.8	90.6	91.3	90.7
	0.3	90.6	90.6	90.4	91.4	90.9	90.8
	0.5	91.3	90.8	90.0	90.8	90.8	90.7
50	0.1	90.8	90.2	90.6	89.9	90.0	90.3
	0.3	89.9	89.7	90.0	90.8	91.7	90.4
	0.5	90.0	90.9	90.9	89.7	89.9	90.3

ときに正しい後続単語が選択される割合」である．

4.2 フィードフォワード型ニューラルネットのテスト結果

3.2節で学習を行ったフィードフォワード型ニューラルネットに対して上記のテストを実施したとき，それぞれのニューラルネットが示した正答率，およびその平均を**表3**に示す．これを見ると，中間層のユニット数が少ないニューラルネットほど平均正答率が高く，中間層のユニット数が増加するにしたがって，わずかずつではあるが正答率が下がっているのがわかる．この理由は次のように考えられる：

「入力された文脈情報は，中間層の活性値パターンとして表現される．中間層のユニット数が少ない場合，文脈情報は効率の良い，コンパクトな形で表現されるが，ユニット数が増加した場合は，本質的でない情報を含んだ形—すなわちノイズの多い形—で表現されるようになる．そのため，未知の先行自立語列（学習パターンに含まれない先行自立語列）に対する挙動が不安定になる．」

また，正答率は中間層のユニット数のみに依存し，学習時における許容誤差には影響を受けないように見える．このことからフィードフォワード型ニューラルネットを用いた場合は，理想出力と実際の出力との許容誤差が0.5になった時点で，十分に学習が完了しているものと推察できる．

4.3 フィードバック型ニューラルネットのテスト結果

フィードフォワード型ニューラルネット上で行ったのと同様のテストを，フィードバック型ニューラルネットに対しても行った．このときの結果を**表4**に示す．フィードフォワード型での実験結果とは異なり，学習時の許容誤差のみならず，中間層のユニット数も正答率に影響を与えていないように見える．

表 4 テストデータに対するフィードバック型ニューラルネットの正答率
Table 4 Percentage of correct answers given by feedback neural networks.

中間層中のユニット数	許容誤差	正答率（%）					
		試行1	試行2	試行3	試行4	試行5	平均
10	0.1	92.8	92.4	91.5	92.0	92.4	92.2
	0.3	92.0	92.2	91.7	92.3	91.8	92.1
	0.5	92.0	90.2	91.3	91.8	91.7	91.4
30	0.1	91.8	92.0	92.4	92.2	91.7	92.0
	0.3	92.3	92.8	92.4	91.8	91.8	92.3
	0.5	92.4	92.8	91.8	90.8	90.2	91.6

フィードフォワード型とフィードバック型との平均正答率を比較してみると，わずかながらフィードバック型の方が高くなっていることがわかる．この結果を見る限り，ニューラルネットに先行自立語列を入力する際には，重み付け和の形で与えるよりも，時系列そのままの形で与えた方が，一般化能力の点で有利であると言える．

4.4 未学習のニューラルネットを用いた場合のテスト結果

対照実験として，学習をまったく行っていない状態のニューラルネットに対して同様のテストを行った．未学習状態のニューラルネットが示す正答率は，中間層のユニット数，あるいはニューラルネットの構造（フィードフォワード型であるかフィードバック型であるか）にかかわらず一定である．なぜなら未学習状態のリンクの重みはすべて0である*ために，各出力ユニットへの入力の和もすべて0になり，したがってそれら出力ユニットからの出力は，中間層のユニット数あるいはニューラルネットの構造にかかわらず常に0.5となるからである．

学習をまったく行っていない状態のニューラルネットの正答率は38.9%であった．これは，学習済みのニューラルネットの平均正答率と比較してはるかに低い数字である．このことから，学習済みのニューラルネットが示した語彙的曖昧性解消能力（4.2節および4.3節の結果）は，ニューラルネットという処理形態そのものから生じたのではなく，学習を通じて獲得されたものであることが確認できる．

また，テスト用データ中に学習用データと同一の文

* 3.2節および3.3節で行った学習実験の中でリンクの重みの初期値を[−0.01, +0.01]に設定しているのは，対称性を破壊するためのテクニックに過ぎない[5]．

は含まれていない．それにもかかわらず，学習パターン習得後のニューラルネットは，テスト用データに対して90%を越える平均正答率を示した．このことから，ニューラルネットは学習パターンを単に丸暗記したのではなく，その中に含まれる構造を抽出してそれをテスト用データに適用したのだと考えることができる．

なお，4.1節で説明した「テスト手順」のstep 6からわかるように，出力ユニットからの出力値が等しい場合は，最長一致が優先される．したがって未学習のニューラルネットが示す正答率は，最長一致法を用いた場合のそれと等しいことを指摘しておく．

5. 考　　察

本章では拡張性に関する議論を行う．すなわち，前節までで述べたニューラルネットを拡張して実用的な形態素解析システムに組み込む際，問題となりそうな点について考察する．

5.1 より大きな辞書の使用

今回の実験では単語数6619の辞書を使用したが，実用的な形態素解析を行うためには，数万～数十万語の辞書が必要であると言われている．本稿のニューラルネットでは，1単語1ユニット方式を採用しているので，実用的な形態素解析のためには，入力層および出力層の各々に，数万～数十万のユニットを用意しなければならないことになる*．この場合の問題点は，第1に1単語1ユニット方式のままで実用的な規模の辞書が扱い切れるかどうかであり，第2に，もし扱い切れるのなら，そのときのニューラルネットが，どのような学習能力および一般化能力を示すかである．

まず第1の点に関してであるが，1単語1ユニット方式のままでも，実用的な規模の辞書を扱うことは十分可能であると考えられる．根拠としては，文献3)で述べられている仮名漢字変換システムが，既に実用化されていることが挙げられる．文献3)で使用されているニューラルネットは，本稿と違って相互結合型であるが，1単語1ユニット方式という点，および日本語を対象としている点で本稿と共通である．このようなシステムが実用化されているという事実は，1単語1ユニット方式のままでも，実用的な規模の辞書の扱いに問題がないことの証左であると言える*．

次に第2の問題点，すなわち拡張されたニューラルネットの学習能力・一般化能力について考察する．

学習用例文は第3節で用いたもののまま変化させず，辞書だけを大きくした状態で，3つ組の学習パターンを作り直すとする．この場合，先行自立語列および後続単語集合は現在のものと同じで変化しない．一方，非後続単語集合に属する単語の数は増加する可能性がある．このことは，出力層において明示的に0を出力しなくてはならないユニットの数が増加することを意味する．満たすべき制約が増えるので，一般に学習はより困難になる．

さて，このより困難になった問題をニューラルネットが学習できるかどうかであるが，もし十分な数の中間層ユニットが存在するなら，原理的に学習可能であることは明らかである．また実際問題として，もし第3章の学習パターンおよび第4章のテストパターンに対して現在のニューラルネットと同等の出力結果を得ることだけが目的であるならば，わざわざ初期状態から学習させなくても，以下のようにして新しいニューラルネットを作成することができる．

[step 1] 第3章の実験の結果得られたニューラルネット（単語数6619）を，任意に1つ選択する．

[step 2] 新しく加えたい単語に対応するユニットを出力層に追加する．これらユニットのバイアス値は絶対値の十分に大きい負の値とする．

[step 3] 中間層の各ユニットと，出力層に追加された各ユニットとの間を，重み0のリンクで結ぶ．

このようにすることで，以前から存在していたユニットからの出力を変更することなく，新しく加えられた単語に対応するユニットからの出力のみを常に十分小さく抑えることができる．出力が十分に小さければ，それらのユニットに対応する単語が後続単語として選択されることはない．

以上の考察により，たとえ辞書が実用規模まで大きくなっても，第3章の学習パターンおよび第4章のテストパターンに対する出力結果が，現在のニューラルネットより悪くならないようにすることが可能である

---

* 入力層に関する限り，実際に用意すべきユニットは，学習パターンの先行自立語列中に現れる単語に対応するものだけで事足りる．もし仮に，辞書中の全単語に対応するユニットを用意したとしても，先行自立語列中に現れない単語に対応するユニットからの出力は常に0であるため，それらのユニットから中間層へと伸びているリンクの重みは，決して更新されることなく，初期値の0のまま残る．リンクの重みが0であるということは，学習終了後にどのような入力パターンが来ようとも，それらのリンクを通じて中間層に渡される値が，常に0であることを意味する．したがって，工学的には，これらのユニットおよびリンクは省略できる．

* もっとも，一般化能力という観点からは，1単語1ユニット方式の局所表現よりも，複数のユニットで複数の単語を表す分散表現の方が有利であると言われている．この場合，どのような分散表現を採用するかが問題であるが，ニューラルネット自身に単語の分散表現を獲得させる方法として，文献8）が興味深いアイデアを提供している．

ことがわかる．

5.2 未知語の問題

第4章で行った一般化能力のテストにおいては，入力文中に含まれる単語はすべて辞書に登録されているものと仮定した．しかし，実際の形態素解析においては，未知語の存在を避けて通ることはできない．特に重要で，また困難な点は，未知語を未知語として認識することである．

今，入力文の i 文字目から未知語が始まるとしよう．このとき，使用する辞書によって次の2つの場合がありうる：

a) 入力文の i 文字目以降にマッチする単語が辞書中に存在しない．

b) 入力文の i 文字目以降にマッチする単語が辞書中に存在する．

a) の場合はすぐに未知語であることがわかるので，それなりの対応をとることができる．一方 b) の場合は，単語境界を誤り，間違った単語を切り出してしまうことになるので問題が生じる．この場合を，さらに次の2種類に分けて考える：

b-1) 誤って切り出された単語が，たまたまそのときの文脈と関連の強い場合．

b-2) 誤って切り出された単語が，そのときの文脈と関連の弱い場合．

本稿のニューラルネットは，先行する文脈と関連の強い単語を優先的に見つけることを目的としている．そのため，b-1)の場合において，誤って切り出された単語を後続単語と認識してしまうのは仕方ないと言えよう．多くの場合は，解析がさらに先に進み矛盾が発見された時点で誤りに気付くことになる．

一方 b-2) の場合は，「切り出された単語に対応するユニットからの出力値をもとに，その解の信頼度を推察する」というヒューリスティックスが使える可能性がある．もしある出力ユニットからの出力値が 0.5 未満であったなら，そのユニットに対応する単語は真の後続単語でない可能性が高い．なぜなら 0.5 未満というのは，本来ならば非後続単語集合の要素が出力すべき値だからである．

たとえば，一般化能力のテストにおける試行の1つ(フィードフォワード型，中間ユニット数 30，許容誤差 0.1 の第5試行)において，ニューラルネットが選択した単語に対応するユニットの出力値が 0.5 未満であった回数は 22 回であったが，そのうち選択された単語が正解であったのは，わずか 10 回にとどまった．

つまり，切り出された単語に対応するユニットからの最大出力値が 0.5 未満である場合，その中に正解が含まれる可能性は低い．すなわち，そこに未知語がある可能性が高いことがわかる．

このように，ニューラルネットを用いることで，未知語の扱いに関しても有利な情報が得られると期待できる．

5.3 前方での解析誤りが後方に与える影響

第4章で行った一般化能力のテストの結果は，1度誤った単語を選択しても，それが更なる後続単語の選択に悪影響を及ぼすことはない，と仮定した場合のものであった．しかし実際の形態素解析においては，1度単語選択に失敗すると，それがその後ろの部分の解析に悪影響を及ぼすことが多い．完全横型の探索を行えば，前方での誤りが後方に影響を与えることはなくなるが，それでは本稿の主旨に反する．

第4章の実験において単語選択に失敗した状況を分析してみると，その多くの場合，正解を第2候補としていることがわかった*．たとえば上でも取り上げた，フィードフォワード型，中間ユニット数 30，許容誤差 0.1 の場合の第5試行では，単語選択に失敗した 48 例のうち，41 例において正解が第2候補となっていた．つまり，この試行においてベスト2に正解が含まれる割合は $(504+41)/552=98.7\%$ であった．

このことから，前方での誤りに対して頑強性を増すためには，最大値を出力しているユニット1つのみを選択するのではなく，最大値に近い値を出力しているユニットいくつかを並列に扱うようなメカニズム（ビームサーチに類似したメカニズム）が有効であると思われる．ただし，あまり多くのユニットを並列に扱おうとすると，「完全並列探索を避け，見込みのある単語から順に調べることで効率向上を狙う」という本来の目標がぼやけてしまうおそれがあるので注意が必要である．

6. おわりに

文脈情報を利用した語彙的曖昧性解消の可能性を探るための基礎的実験を行った．実験には入出力層の各ユニットに個々の単語を割り当てた階層型ニューラルネットを使用し，文脈情報として先行する自立語の系列を与えた．実際の出版物に現れた 108 文を用いてニューラルネットを訓練した後，別の出版物に現れた 33 文を用いて語彙的曖昧性の解消を行わせた結果，全体の 90%以上の局面で正しい単語が選択された．

---

* 単語 x が第2候補であるとは，文字表記的・品詞接続的に選択されうる全単語に対応するユニット中，x に対応するものからの出力値が上から2番目に大きい値であったということである．

今回の実験では，同一の後続単語集合に含まれる各単語は，与えられた文脈においてみな同一の関連度を示すものと仮定した．しかし実際には，たとえ同一の後続単語集合に含まれていても，文脈との関連度は単語ごとに異なるはずである．各出力ユニットからの出力値が，入力層に与えられた文脈と，その出力ユニットに対応した単語との関連度を反映するようにすれば，形態素解析の補助機能としてより有効になると思われる．

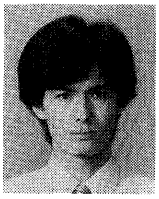
今後は上記の点に配慮しつつ，より広範囲の文脈・意味情報を反映した単語選択について研究を進めて行く予定である．

謝辞 本研究は，リアルワールドコンピューティング（RWC）研究計画の一環として行われた．研究を進めるにあたり，貴重なアドバイスをいただいた電子技術総合研究所 自然言語研究室ならびに同 新情報計画室の諸氏に感謝いたします．

参 考 文 献

1) Walts, D. L. and Pollack, J. B.: Massively Parallel Parsing: A Strongly Interactive Model of Natural Language Interpretation, *Cognitive Science*, Vol. 9, pp. 51-74 (1985).
2) 田村, 安西：Connectionist Model を用いた自然言語処理システム，情報処理学会論文誌, Vol. 28, No. 2, pp. 202-210 (1987).
3) 木村, 鈴岡, 伊藤, 天野：神経回路網の連想機能を用いたかな漢字変換システム—ニューロワープロの実験試作—，第 4 回人工知能学会全国大会 9-3, pp. 301-304 (1990).
4) 高橋, 板橋：ニューラルネットワークを用いた日本語解析の試み，情報処理学会論文誌, Vol. 32, No. 10, pp. 1330-1337 (1991).
5) Rumelhart, D. E., Hinton, G. E. and Williams, R. J.: Learning Internal Representations by Error Propergation, in Rumelhart, D. E., McClelland, J. L. and the PDP Research Group: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 1, pp. 318-362, The MIT Press, Massachusetts (1986).
6) 高橋：後続部分予測機能を持つ日本語解析システム，信学技報 NLC 92-40 (1992).
7) Elman, J. L.: Finding Structure in Time, *Cognitive Science*, Vol. 9, pp. 179-211 (1990).
8) Miikkulainen, R.: *Subsymbolic Natural Language Processing*, pp. 47-83, The MIT Press, Massachusetts (1993).

（平成 5 年 11 月 5 日受付）
（平成 7 年 6 月 12 日採録）



高橋 直人（正会員）

1987 年筑波大学第三学群情報学類卒業．1992 年同大学大学院博士課程工学研究科修了．博士（工学）．同年電子技術総合研究所入所．自然言語処理，マルチリンガル情報システムの研究等に従事．人工知能学会，言語処理学会各会員．