

ニューラルネットに基づいた日本語文の解析

1991

高橋 直人

①

工学博士学位論文

ニューラルネットに基づいた日本語文の解析

筑波大学大学院博士課程

工学研究科電子・情報工学専攻

高橋直人

1991年

目次

1 序論	1
2 ニューラルネットと自然言語処理	3
2.1 ニューラルネットの自然言語処理への適用例	3
2.1.1 形態素解析への応用	3
2.1.2 係り受け解析への応用	4
2.1.3 統語解析への応用	4
2.1.4 その他の研究	5
2.2 相互結合型ニューラルネットの数理	7
2.2.1 相互結合型ニューラルネットのエネルギー関数	7
2.2.2 Boltzmann Machine	8
3 相互結合型ニューラルネットによる形態素解析	10
3.1 日本語形態素解析とは	10
3.2 形態素解析ネットワークの構成	12
3.3 エネルギー関数の決定	13
3.3.1 文字列の分解	13
3.3.2 文法的接続可能性	14
3.3.3 自立語数	15
3.3.4 形態素解析ネットワークのエネルギー関数	15
3.4 実験	16
3.5 考察	38
4 相互結合型ニューラルネットによる係り受け解析	40

4.1	日本語の統語構造と文節間係り受け関係	40
4.2	ネットワークでの表現方法	42
4.3	エネルギー関数の決定	43
4.3.1	係り受け関係の数	43
4.3.2	相互排他性	43
4.3.3	文節間の意味的關係	44
4.3.4	文節間の距離	45
4.3.5	係り受け解析ネットワークのエネルギー関数	46
4.4	実験	47
4.5	考察	62
4.6	単語共起頻度に関する議論	64
5	文脈情報を利用した形態素解析	66
5.1	導入	66
5.2	ニューラルネットによる文脈の表現	68
5.3	ネットワークの構成	69
5.3.1	ユニット群とその役割分担	69
5.3.2	ユニット群間リンクおよびユニット群内リンク	73
5.4	いくつかのヒューリスティクス	76
5.4.1	連想ユニット内リンクの重みの決定方法	76
5.4.2	活性伝達の範囲	77
5.4.3	ユニットの入出力関数	80
5.4.4	バックトラック時の動作	80
5.5	実験および考察	83
5.6	議論	89
5.6.1	メモリ効率とバックトラック	89
5.6.2	並列計算機によるシミュレーション	89
6	結言	94
	付録 A 単語辞書	96

付録 B 品詞接続規則	109
謝辞	127
参考文献	128
発表論文一覧	133

第 1 章

序論

従来開発されてきた自然言語解析システムは、逐次型の計算機上で実行されることを想定したものがほとんどであった。現在一般に広く利用されている計算機のほとんどすべてが逐次型であることを考えれば、これは当然とも言える。自然言語文に含まれている曖昧性を解消するためには、統語的解析のみならず、意味的・文脈的・語用論的解析が必要である。人間はこの困難な問題を極めて効率良く行なっている。一方、自然言語に関する知識やヒューリスティックスをいかに表現するかという問題にはまだ明確な解答が得られておらず、従って、計算機で自然言語を解析する場合にはしらみつぶしの方法をとらざるを得ないことが多い。すなわち、文法的に可能なすべての解析木を生成し、それらの中から与えられた評価関数を最大にするものを選択するのである。評価関数には意味的・文脈的・語用論的妥当性を反映するように構成されたものが選ばれる。このような文法的に可能なすべての解析木の妥当性を評価する方法には、大きな計算コストが必要とされる。

解析対象言語が日本語の場合、形態素解析とは文を単語単位に分割することを指す。日本語文では、通常、単語と単語の間に空白を置かないので、同一の文を単語単位に分割する方法が複数通り存在し得る。特に文が仮名のみあるいはローマ字のみで書かれている場合は分割方法が多数存在するため、組合せ的爆発が生じ得る。

また、日本語の統語的構造は文節間の係り受け関係で表現することができる。一般的に、文法的制約のみでは係り受け関係を一意に決定することはできない。ある 2 文節間の係り受け関係の成立しやすさは、それら 2 文節に含まれている単語同士の意味的關係や、その 2 文節が字面上でどれだけ離れているかといったことに影響を受ける。

ここで視点を変えてみると、形態素解析・係り受け解析の両方とも一種の最適化問題、すなわち与えられた制約条件を最もよく満たすようにコスト関数の値を最小とする問題、と考えることが

できる。Hopfield と Tank は、ある種の最適化問題は相互結合型のニューラルネットを用いることで極めて効率良く解けることを示した [Hopfield84][Hopfield85]。これは日本語形態素・係り受け解析がニューラルネット上で効率良く実行できることを示唆している。また、近年ハードウェアによるニューラルネット・シミュレータやニューラルネット・チップが開発されるようになってきており、これらを利用すれば極めて高速の自然言語解析エンジンの作成が可能になると思われる。

本論文では日本語をニューラルネットで解析する方法、および計算機上でそのシミュレーションを行った実験結果について述べる。

第2章では、まず初めにニューラルネットを自然言語処理に応用したこれまでの研究を概観し、その後で相互結合型ニューラルネットの一般的な振舞い、およびそのエネルギー関数について述べる。これは第3章および第4章で展開される理論の基礎となるものである。

第3章では日本語形態素解析を組み合わせ最適化問題として定式化し、それを相互結合型ニューラルネットを用いて解いた実験について述べる。ここでは最初に入力文字列中に含まれ得る単語を辞書引きして求め、それらを各ユニットに対応させる。組み合わせ問題と考えたときのコスト関数としては、隣接する2単語の品詞が文法的に接続可能であることのみならず、自立語数が最少になるものを選択するという語用論的制約が反映されているものを用いる。

第4章では、前章で用いた手法を日本語係り受け解析に適用した結果について述べる。ここでは係り受け関係に関する文法的な制約のほか、係り受けが生じる2文節に含まれる自立語の意味的關係、および文節間の距離がコスト関数に反映されている。システムは前処理の済んだ文節列を入力とし、それらの間で可能な係り受け関係をユニットに割り当てる。

第5章では、第3章で問題となった点の一部を解決することを目的としたネットワークを構築する。ここで用いられているネットワークは前2章のものとは異なり、相互結合型 (Hopfield 型) の形をとっていない。解くべき問題も最適化問題の一種としては定式化されておらず、むしろ超並列解析とでも言うべき形で扱われている。

最後に第6章で全体をまとめ、今後の展望について述べる。

第 2 章

ニューラルネットと自然言語処理

2.1 ニューラルネットの自然言語処理への適用例

本節では、ニューラルネットがこれまでどのように自然言語処理に応用されてきたかを概観する。とり上げる研究は本論文と関連が深いものに限定する。

2.1.1 形態素解析への応用

筆者らの研究 [高橋 90b] [高橋 90c] [高橋 91a] [高橋 91b] [高橋 91c] 以外でニューラルネットを(狭い意味での)日本語形態素解析に利用したものは見当たらないが、日本語ワードプロセッサにおける同音異義語の解消にニューラルネットを応用したものとしては [木村 90] がある。これは日本語ワードプロセッサにおける仮名漢字変換にニューラルネットの機構をとり入れることで変換精度の向上を狙ったものである。概要を簡単にまとめると以下のようになる。

1. 予め与えられた漢字仮名まじり文章(学習データ)から、語の結びつきの強さをネットワーク形式で形成する。ネットワーク中の各ノードは、仮名漢字変換辞書の各単語項目に一意に対応し、それぞれ出力値を持つ。また、ノード間のリンクは重みを持つ双方向性のリンクであり、単語間の結びつきの強さを示す。
2. 同音語選択時にある単語が確定されると、その確定情報はその単語に対応するノードに対する入力となるので、そのノードの出力値が上昇する。この出力値の上昇がリンクを伝搬することにより、そのノードに接続されているノード(連想語)の出力値が上昇する。仮名漢字変換における同音語の出力順位は、このノードの出力値の大小を比較して決定する。

ここでは入力平仮名列を単語に分割する操作はネットワーク外で行われている。

本論文の第5章では、これと似た同音異義語解消メカニズムに、入力文字列の切り分けの機能を盛り込んだニューラルネットについて述べる。

2.1.2 係り受け解析への応用

相互結合型ニューラルネットを日本語係り受け解析に応用した例としては、筆者らの一連の研究がある [高橋 90a][高橋 90b][高橋 90c][高橋 91c]。これに関しては第4章で詳しく述べる。

この他にニューラルネットを日本語係り受け解析に応用した研究としては [奥村 90], [村瀬 89] などがある。

[奥村 90] は文節間の係り受け構造を

1. 係り節の接続形
2. 受け節の叙述表現
3. 係り受け制御素性

から推定する際に3層型ニューラルネットを用いた研究である。5文節からなる文を解析するためのネットワーク(中間層ユニット数30)に317の例文を与えて学習を行なわせた後に未学習データ41文を解析したところ、その内の12文の解析に成功したと報告している。

[村瀬 89] は、係り受け解析を最適化問題と見なして Boltzmann Machine で解くという点は本研究と同じであるが、相互結合型ではなく3層構造ネットワークを用いている点が異なっている。また入力情報として音声認識装置の出力を想定している点、意味情報を考慮していない点などにおいても本研究と異なっている。

2.1.3 統語解析への応用

比較的簡単な文法規則と単語間の意味的關係をネットワークで表現し、統語的・意味的に曖昧な文を解析した研究としては [Waltz85] が有名である。同様の処理を日本語に対して行った例には [田村 87] や [森 89] 等がある。これらはいずれも、互いに整合する文法的・意味的要素を表すユニットの間には興奮性のリンクを、互いに矛盾・相反する要素を表すユニットの間には抑制性のリンクを張ることによってネットワーク全体を最適な状態に落ち着かせようとしたものである。たとえば [田村 87] では、

正月 に たこ を あげる

(2.1)

という文には以下のような語彙的曖昧性があるものとしている。

1. 「に」
 - (a) 時を示す (CaseTime)
 - (b) 受け手を示す (CaseRecipient)
 - (c) 場所を示す (CaseLocP)
 - (d) 目的地を示す (CaseLocD)
2. 「たこ」
 - (a) 空にあげるおもちゃの凧 (KITE)
 - (b) 8本足の軟体動物の蛸 (OCTOPUS)
3. 「あげる」
 - (a) 高いところに位置させる (RAISE)
 - (b) 与える (GIVE)
 - (c) 熱い油で煮る (FRY)

(2.1) では、はじめに「正月」という単語があるため、

正月 → 凧あげ → 凧 (2.2)

の連想がはたらき、「たこ」は凧の意味になり、さらに凧から「あげる」の(a)の意味が活性化される。また、「に」は「正月」と文節をなしているが、「正月」は時を表すので CaseTime とだけ興奮性リンクで結ばれて、これが選択される。このようにして語彙レベルでの曖昧性が解消される。

これらの研究は理論的に非常に興味深いものであるが、残念なことにどれも非常に限定された数の文法規則および語彙項目しか扱っていない。同様のアプローチをそのまま拡大していくと、文法規則および語彙項目が増大するにしたがってネットワークの制御が困難になるという問題が生じる。特に実用的な規模のネットワークを作ろうとした場合に、各ユニット間のリンク重みをどのように決定すれば速く、安定した曖昧性解消が可能になるかという点に関しては未知の部分が多い。

2.1.4 その他の研究

上記以外の分野で、ニューラルネットを自然言語処理に応用した研究では以下に示すものが有名である。

- Hinton らによる 文字-単語-意味素 の連想ネットワーク [Rumelhart86]:

文字列が与えられたとき、その文字からなる単語の意味素を出力するモデルである。実験に際しては実際の単語や意味素ではなく、機械的に適当に生成された記号が用いられた。

- McClelland らによる語の格決定 [Rumelhart86]:

三層構造ニューラルネットを用いて、語の意味と語順から語の文における格を導くものである。格としては Agent, Patient, Instrument, Modifier の 4 種類のみを考える。また入力文はすでにある程度解析されて動詞および名詞が取り出され、出現順にしたがって並べられているとする。このとき、ネットワークが解くべき課題は、

broke — boy window hammer (2.3)

(原文は The boy broke the window with a hammer.) のような 4 語の入力から、それぞれの名詞の格を出力することである。これは深層格の決定という、他の研究に比べてより深いレベルを取り扱おうとしたものであるが、埋め込み文に見られるような自然言語の再帰構造が考慮されていないことなどの問題点も含んでいる。

- 上村によるリカレントニューラルネットを用いた認識と生成 [上村 90]:

リカレントニューラルネットを用いた自然言語 (英語) の認識と生成の実験である。適切な数の隠れユニットが与えられればネットワークは自然言語の認識ができること、訓練のためのデータが十分に長ければネットワークは訓練用データの範囲を越えて適格文を生成することができること等が示されている。ただし、ここでは自然言語を単なる文字列として取り扱っており、文法規則や意味解析といったものは考慮されていない。

2.2 相互結合型ニューラルネットの数理

本論文の第3章および第4章では、日本語解析を組み合わせ最適化問題の一種として定式化し、それを相互結合型ニューラルネットを用いて解く方法について述べている。以下ではこれらの研究の理論的背景をなす、相互結合型ニューラルネットの数理的振舞いと、それがどのようにして組み合わせ最適化問題を解くのに利用されるかについて述べる。

2.2.1 相互結合型ニューラルネットのエネルギー関数

ニューラルネット内のユニットが階層構造のような構造を作らず、相互に結合しているようなものを相互結合型ニューラルネットと呼ぶ。原則として、あるユニットは他のすべてのユニットと結合を持ち得る。各ユニットは多入力1出力の動作規則にしたがい、適当なタイミングで状態を更新する。すべてのユニットが一斉に状態変化を行うものを同期的なネットワークと呼び、各ユニットが勝手なタイミングで状態変化を行うものを非同期的なネットワークと呼ぶ。

今、ある相互結合型ネットワーク内の各ユニット u_i ($1 \leq i \leq n$) が

$$i_i = \sum_{j=1}^n w_{ij} u_j - \theta_i \quad (2.4)$$

に対し

$$\begin{cases} i_i > 0 & \text{then } u_i = 1 \\ i_i = 0 & \text{then 変化せず} \\ i_i < 0 & \text{then } u_i = 0 \end{cases} \quad (2.5)$$

にしたがって非同期に状態遷移を行うものとする。ここで w_{ij} は u_i から u_j への結合の強さ、 θ_i は u_i のしきい値である。このとき w が $w_{ij} = w_{ji}$ かつ $w_{ii} = 0$ を満たすならば¹、以下の式 (2.6) は時間とともに単調減少し、定常状態に収束することが証明できる [麻生 88] [Hopfield84]。

$$E = -\frac{1}{2} \sum_{i=1}^n \sum_{j \neq i}^n w_{ij} u_i u_j + \sum_{i=1}^n \theta_i u_i \quad (2.6)$$

式 (2.6) をこのネットワークのエネルギー関数と呼ぶ。

¹このような相互結合型ニューラルネットを Hopfield 型ニューラルネットと呼ぶ。

見方を変えると、式(2.6)の形をした関数の極小値を求めるためには、式(2.5)のような相互結合型ニューラルネットを作成してその中の各ユニットを非同期的に動作させればよいことがわかる。したがって与えられた最適化問題と相互結合型ニューラルネットの間で

最適化問題	ニューラルネット
変数	ユニット
コスト関数	エネルギー関数

という対応づけをすることができれば、ニューラルネットを用いて最適化問題が解けることになる。ネットワークのエネルギー関数を最適化問題のコスト関数に合わせるためには、ユニット間の結合の強さおよび各ユニットのしきい値を調節すればよい。ニューラルネットを用いて組み合わせ最適化問題を解いた代表的な研究としては、Hopfield と Tank による巡回セールスマン問題の解法 [Hopfield85] が挙げられる。

2.2.2 Boltzmann Machine

相互結合型ニューラルネットを用いて最適化問題を解こうとしたときに問題となる点の1つに、式(2.6)がエネルギーの最小値以外の極小値に収束する可能性のあることが挙げられる。この場合は最適解を得ることができない。この問題点は、ネットワークとして Hopfield 型の代わりに以下に示す Boltzmann Machine を使うことで解決できる。

Farlman らは Hopfield 型ニューラルネットの動作を確率的にしたメカニズムを提案し、それを Boltzmann Machine と名付けた [Farlman83]。Boltzmann Machine のユニットは Hopfield 型のそれとは異なり、式(2.7)に示した確率で1を出力する。

$$p(u_i = 1) = \frac{1}{1 + \exp(-i_i/T)} \quad (2.7)$$

ここで i_i はユニット i への入力の総和であり、 T はネットワークの温度と呼ばれる正の数である。 T が0の極限で式(2.7)はステップ関数となり、Boltzmann Machine は Hopfield 型ネットワークと同じ動作をする。

Boltzmann Machine はユニットの動作が確率に基づいているので、ネットワークのエネルギーが増加する方向への状態変化も起こり得る。式(2.7)からわかるように、ネットワークの温度が高

いほどエネルギーが増加する方向への状態変化が生じやすい。これは、ネットワークが一旦エネルギーの極小状態に落ち入ってもそこから脱出できる可能性が高いことを示している。しかし温度が高いとネットワークの状態はなかなか収束しない。

一方温度が低い場合、ネットワークはエネルギーの局所的なくぼみから脱出するのが困難となり、最小値付近に到達するまで長い時間がかかる。局所的なくぼみを避けつつ、できるだけ早くエネルギーの最小状態に到達するようにするためには、初めのうちは温度を高くしておき、状態変化をさせながら徐々に温度を下げる疑似焼きなましの技術を使う [Kirkpatrick83]。焼きなましのスケジュールは経験的によいものを採用する。Boltzmann Machine の応用例には [Prager86] 等がある。

第3章

相互結合型ニューラルネットによる形態素解析

本章では相互結合型ニューラルネットを用いた形態素解析について述べる。まず初めに日本語における形態素解析とは何かを説明し、次にニューラルネットを用いて形態素解析を行う方法、および実際にシミュレーションを行なった結果について述べる。

3.1 日本語形態素解析とは

英語・フランス語といった言語で使用される文字数はせいぜい数十である。またこれらの言語では単語と単語の間に空白を置くことが正書法で定められている。

これに対して日本語の表記に使用される文字の種類には、平仮名、片仮名、漢字、ラテンアルファベット、その他があり、これら文字の総数は数千から数万に達する。また日本語を文字表記する際には単語と単語の間に空白を置かないのが普通である。さらに日本語は言語学的には膠着語に分類される言語であり、活用語に接尾辞が連続する際には語尾変化が生じるため、活用語の活用形まで考慮しないと単語と単語の切れ目を見つけることができない。特に入力文が平仮名(あるいは片仮名・ローマ字)のみで書かれているときは曖昧性が非常に高くなる。数十文字の平仮名からなる文が数千の構文木を持つこともある。

したがって計算機で日本語を処理する場合は、英語・フランス語等を処理する場合に比べて、入力された文字列を単語列に正しく分解する作業、すなわち形態素解析がより重要な役割を果たすことになる。

計算機を用いて日本語形態素解析を行うための方法としては

- 最長一致法
- 字種切り法

- 文節数最少法
- CYK 法

等が有名である [田中 89].

最長一致法 [西村 64] は文を左から右の向きに見たとき、辞書中で最も長い形態素を優先させて文を分割する方法である。最長一致法はアルゴリズムが簡便であるという利点があるが、経験的には約 20% の誤った分割が含まれると言われており、最近はあまり用いられなくなっている。

字種切り法 [長尾 78] は、平仮名から他の字種への変わり目、区切り符合 (句読点等) の前後、非平仮名列から数字列への変わり目、数字列から非平仮名列への変わり目に強制的に切れ目を入れる方法である。もちろんこれだけでは多くの誤った分割が含まれる。字種切り法は文字の種類のみ注目すればよいので非常に高速であるが、単独では完全な形態素解析を行うことができず、他の手法と組み合わせる必要がある。

文節数最少法 [吉村 82] は現在もっとも広く用いられている方法の 1 つである。これは入力文に含まれるあらゆる単語の組み合わせのうち、文節数が最少となる単語列を優先する方法である。文節数最少法は入力文が短い場合にはさほど有効に働かず、入力文が長くなるほど効果があるといわれている。日本語形態素解析に文節数最少法を用いた場合、90% 以上の文が正しく解析できるといわれている。

CYK 法 [Aho72] は本来、文脈自由文法¹で書かれた言語の解析を行うものであり、自然言語解析用としては不十分であるが、これを日本語形態素解析に応用する方法が杉村らによって開発された [杉村 88]。この方法を用いれば、原理的には日本語の形態素解析と構文解析を統合することが可能である。

以上のうち実際の形態素解析法として有用であると思われるものは文節数最少法と CYK 法である。これらはどちらも「最初に解となる可能性のあるすべての分割方法を生成し、その中から最善のものを選ぶ」という方法である。したがって単語数が増大したり入力文字列が長くなるにつれて組み合わせ的爆発を生じるおそれがある。

一方、別の視点から見ると日本語形態素解析は組み合わせ最適化問題の一種として定式化することができる。Hopfield と Tank は、ある種の最適化問題は相互結合型のニューラルネットを用いることで極めて効率良く解けることを示した [Hopfield84][Hopfield85]。この場合、最適化問題

¹CYK 法で使用する文脈自由文法はチョムスキー標準形で表現されたものでなければならない。また ϵ 生成規則を含んでいてはならない。

の制約条件はネットワークのエネルギー関数の形で表現される。したがって日本語形態素解析における文法的制約をニューラルネットのエネルギー関数として表現できれば、入力された日本語文を効率よく形態素解析することが可能となる。この場合はすべての可能な分割方法を生成する必要がなく、したがって組み合わせ的爆発が生じる危険もない。

次節では日本語形態素解析用ネットワークの構成と、それを使って実際に日本語文を形態素解析させてみた結果について述べる。

3.2 形態素解析ネットワークの構成

以下では説明のための例として「かれがくるまでまつ」という文字列を取り上げる。辞書項目としては図 3.1中に示した 9 単語を考える。

形態素解析ネットワークでは、解析すべき文字列の各部分にマッチする単語 1 つ 1 つにユニットを対応させる。したがって図 3.1の例ではネットワーク内のユニット数は $u_1 \sim u_9$ の 9 となる。簡単のため、以下では「ユニット u_i で表される単語」を単に「単語 u_i 」と表記する。ここでの目的は、ネットワークの活動が収束した状態で値 1 をとっている単語のみが文の正しい構成要素となるように、ユニット間の重みおよび各ユニットのしきい値を調節することである。

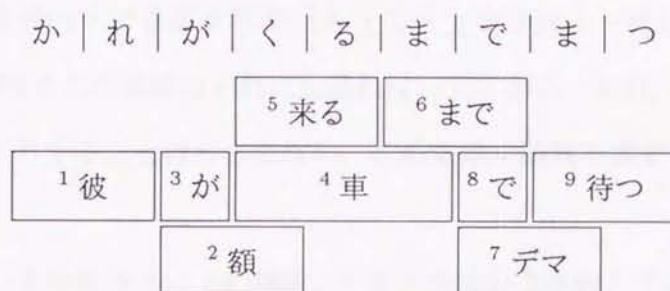


図 3.1: 入力文を単語列に分解する方法の例

3.3 エネルギー関数の決定

文を正しい単語列に分割するためには、以下の制約を満たすようなエネルギー関数を作成する必要がある。

1. 入力文中の各文字が、ちょうど1回ずつ用いられるような単語の組み合わせが選択されたときに最小値をとる
2. 隣接する単語の組み合わせがすべて文法的に接続可能な場合に最小値をとる
3. 選択された単語列に含まれる自立語の数が少ないほど小さな値をとる

最初の2条件は文法的なものである。最後の条件は、文節数最少法に類似した制限条件を与えるためのものである。上の1~3を表す関数をそれぞれ E_C , E_G , E_I とした場合、形態素解析ネットワークのエネルギー関数はこれらの一次結合

$$E = pE_C + qE_G + rE_I \quad (3.1)$$

で表わされる。ここで p , q , r は正の定数である。以下では E_C , E_G , E_I について少し詳しく説明する。

3.3.1 文字列の分解

日本語形態素解析とは入力文字列を単語列に分割することであるから、ネットワーク内で最終的に選択された単語を並べてできる文字列はもとの入力文字列と一致していなければならない。入力文字列のうち、選択された単語のどれにも現れない文字があったり、反対にある文字が2つ以上の単語の中に現れたりすることは許されない。この制限を関数で表すには以下のようにすればよい。

まず n をネットワーク中のユニットの総数、 N を入力文の文字数とする。このとき

$$C_{ik} = \begin{cases} 1 & \text{単語 } u_i \text{ が入力文中で } k \text{ 番目の文字を含む場合} \\ 0 & \text{otherwise} \end{cases} \quad (3.2)$$

と定義される $n \times N$ 行列 C を考える。図 3.1 の例だと

$$C = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \quad (3.3)$$

となる。この C を用いると、入力文中の各文字がちょうど1回ずつ使用されるような単語の組み合わせが選択された時に最小値0をとる関数は

$$E_C = \sum_{k=1}^N \left(\sum_{i=1}^n C_{ik} u_i - 1 \right)^2 \quad (3.4)$$

と表すことができる。

3.3.2 文法的接続可能性

式(3.4)を用いれば、選択された単語を順次連結して得られる文字列と入力文字列とを等しくすることができる。しかし、字面が同一であるからといって選択された単語列が必ずしも文法的に正しい接続になっているとは限らない。ネットワークによって選択された単語を順次連結したものが正しい日本語であるためには、すべての隣接した単語間の接続が文法的に正しくなっていないなければならない。位置的に隣接した2単語すべてが文法的に正しく接続されている場合にのみ最小値0をとる式は、以下のように記述することができる。

$$E_G = \sum_{i=1}^n \sum_{j \neq i} G_{ij} u_i u_j \quad (3.5)$$

ただし G_{ij} は

$$G_{ij} = \begin{cases} 1 & \text{単語 } u_i \text{ と単語 } u_j \text{ が入力文中で隣接しており、} \\ & \text{かつ文法的に接続可能でない場合} \\ 0 & \text{otherwise} \end{cases} \quad (3.6)$$

である。

例えば「代名詞+名詞」という接続は文法的でないとする。すると図3.1中の u_1 「彼」と u_2 「額」は、文中で隣接しておりかつ文法的に接続可能でないので $G_{12} = G_{21} = 1$ となる。

3.3.3 自立語数

分ち書きされていない日本語文を単語に分解する際の基準はいろいろ提案されているが、ここでは文節数最少法に類似した自立語数最少法を採用する。すなわち、ある文を分解する方法が複数通りある場合は、そのなかでもっとも自立語数の少ない分解方法を選択するものとする。今回の実験では1文節に含まれる自立語は1つだけとしたので、複合名詞が現れる場合を除けば自立語数最少法と文節数最少法は同じ結果を与える。

まず I_i を次のように定義する。

$$I_i = \begin{cases} 1 & \text{単語 } u_i \text{ が自立語の場合} \\ 0 & \text{その他.} \end{cases} \quad (3.7)$$

この I_i を用いると、式

$$E_I = \sum_{i=1}^n I_i u_i \quad (3.8)$$

は、選択されたユニット中に自立語が少ないときほど小さな値をとるようになる。

3.3.4 形態素解析ネットワークのエネルギー関数

以上で E_C , E_G , E_I が決定された。これより式 (3.1) と式 (2.6) を比較すると、形態素解析ネットワークにおけるユニット間の結合の強さ w_{ij} および各ユニットのしきい値 θ_i は以下のようになることがわかる。

$$\begin{cases} w_{ij} = -2 \left(p \sum_{k=1}^N C_{ik} C_{jk} + q G_{ij} \right) \\ \theta_i = -p \sum_{k=1}^N C_{ik} + r I_i. \end{cases} \quad (3.9)$$

3.4 実験

本節では、上で述べたネットワークのシミュレーションプログラムを作成し、実際に文を解析させてみた結果について述べる。用いた計算機は Sun-4/330, 言語は Allegro Common Lisp である。

解析のための入力文には「新明解国語辞典 第二版 (磁気テープ版)」[金田一 74] の語義説明文のうち、植物に関する百科事典的記述を抜き出し、平仮名べた書きに改めたものを用いた。平仮名べた書きに改めたのは同音異義語が含まれる可能性を意図的に高くし、それに対してネットワークがどの程度の曖昧性解消能力を持つか見るためである。

実験は 2 段階に分けて行なった。第 1 段階は予備実験とし、エネルギー関数各項の重み定数および焼きなましのスケジュールをいろいろ変化させてネットワークの収束の様子を観察した。第 2 段階では、第 1 段階で得られた最適な定数の組み合わせを用いて日本語文の形態素解析実験を行ない、定量的なデータを収集した。

第 1 段階の予備実験では用意した全解析対象文の約 1/4 にあたる 50 文の解析を行なった。その際、ネットワーク中の各定数をさまざまに変化させて収束の様子がどのように変化するか観察した。その結果、式 (3.9) の係数として

$$p = 4, \quad q = 2, \quad r = 3. \quad (3.10)$$

という値の組み合わせを用ると、ほとんどの場合に正しく形態素解析が行なわれ、またネットワークも比較的速く収束することがわかった。また、時刻 t のときのネットワークの温度 $T(t)$ は次式で与えられるものとした。

$$T(t) = \frac{T_0}{1 + t/\tau} \quad (3.11)$$

ただし

$$T_0 = 5, \quad \tau = 10 \quad (3.12)$$

である。この T_0 および τ も予備実験によって経験的に得られたものである。

なお、ユニットの初期状態はすべて 0 とし、状態遷移は非同期に行うこととした。

第 1 段階の実験で得られた各定数を用いて第 2 段階の実験を行なった。第 2 段階の実験結果を図 3.2 に示す。ユニットの状態遷移は非同期かつランダムに生じるので、同一条件で追試を行なった場合でも実験結果の細部は異なる可能性がある。ここではすべてのユニットが平均して 1 回ずつ発火するのに必要な時間を 1 ステップとしている。なおネットワークが与えられた条件をすべて満たした状態に収束したのは、約 200 の入力文のうち約 95% であった。

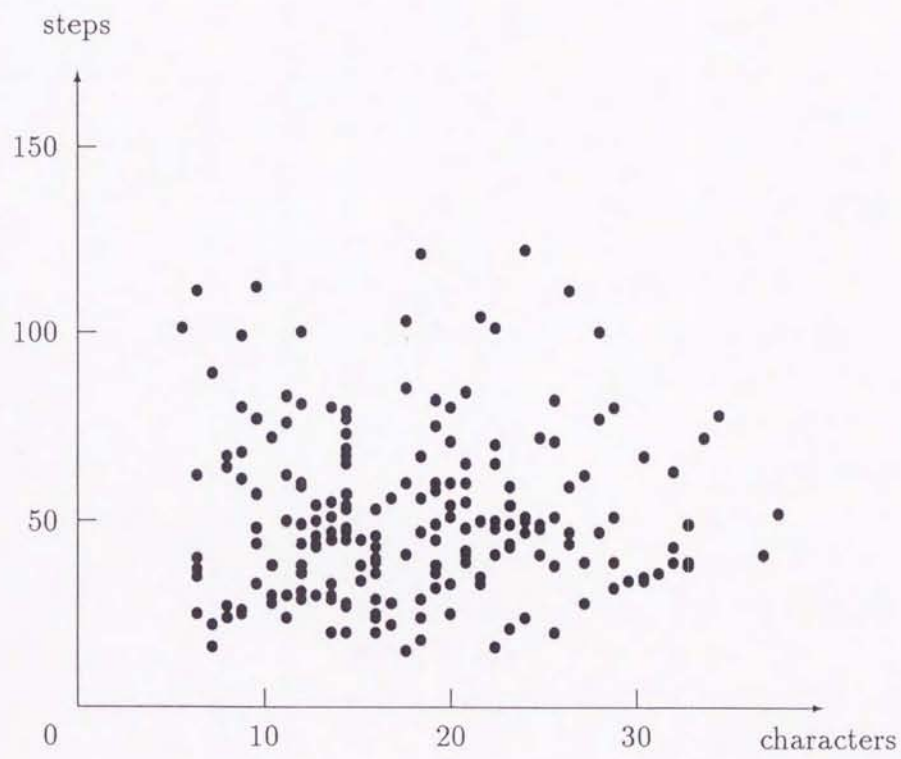


図 3.2: 入力文中の文字数と形態素解析ネットワークが収束するまでのステップ数との関係

以下、個々の入力文に対するネットワークの挙動の例をいくつか示す。

例1: 「その実から米をとる、わが国で一番大切な農作物。」

解析に成功した例である。この文を解析するためのネットワークには36個の単語が含まれている。各ユニットに割り当てられた単語は表3.1のとおり。

ユニット	開始位置	表記	品詞
0	0	その	連体詞
1	2	み, 実	名詞
2	2	み, 見	一段動詞語幹
3	3	か	終助詞
4	3	から, 殻	名詞
5	3	から	格助詞
6	5	こ, 濃	形容詞語幹
7	5	こ, 来	「来る」語幹
8	5	こめ, 米	名詞
9	7	を	格助詞
10	8	と, 取	ら行五段動詞語幹
11	8	と, 採	ら行五段動詞語幹
12	9	る	ら行五段動詞語尾
13	10	,	読点
14	11	わがくに, 我が国	名詞
15	15	で, 出	一段動詞語幹
16	15	で	格助詞
17	15	で	断定の助動詞
18	16	い, 言	わ行五段動詞語幹
19	16	い, 胃	名詞
20	16	いちばん, 一番	副詞
21	17	ち, 散	ら行五段動詞語幹
22	20	た, 田	名詞
23	20	たいせつ, 大切	名詞
24	21	い, 言	わ行五段動詞語幹
25	21	い, 胃	名詞
26	22	せ, 背	名詞
27	22	せ	「する」語幹
28	22	せ	さ変動詞語尾
29	23	つ, 付	か行五段動詞語幹
30	24	な, 名	名詞
31	24	な, 無	形容詞語幹
32	24	な, 成	ら行五段動詞語幹
33	24	な	断定の助動詞
34	25	の	助詞
35	25	のうさくもつ, 農作物	名詞
36	31	。	句点

表 3.1: 例1 のための形態素解析ネットワーク内のユニット

各ユニットのステップ毎の状態遷移の様子を図 3.3 に示す。ユニットの値が 1 となっている部分が黒い点で表されている。

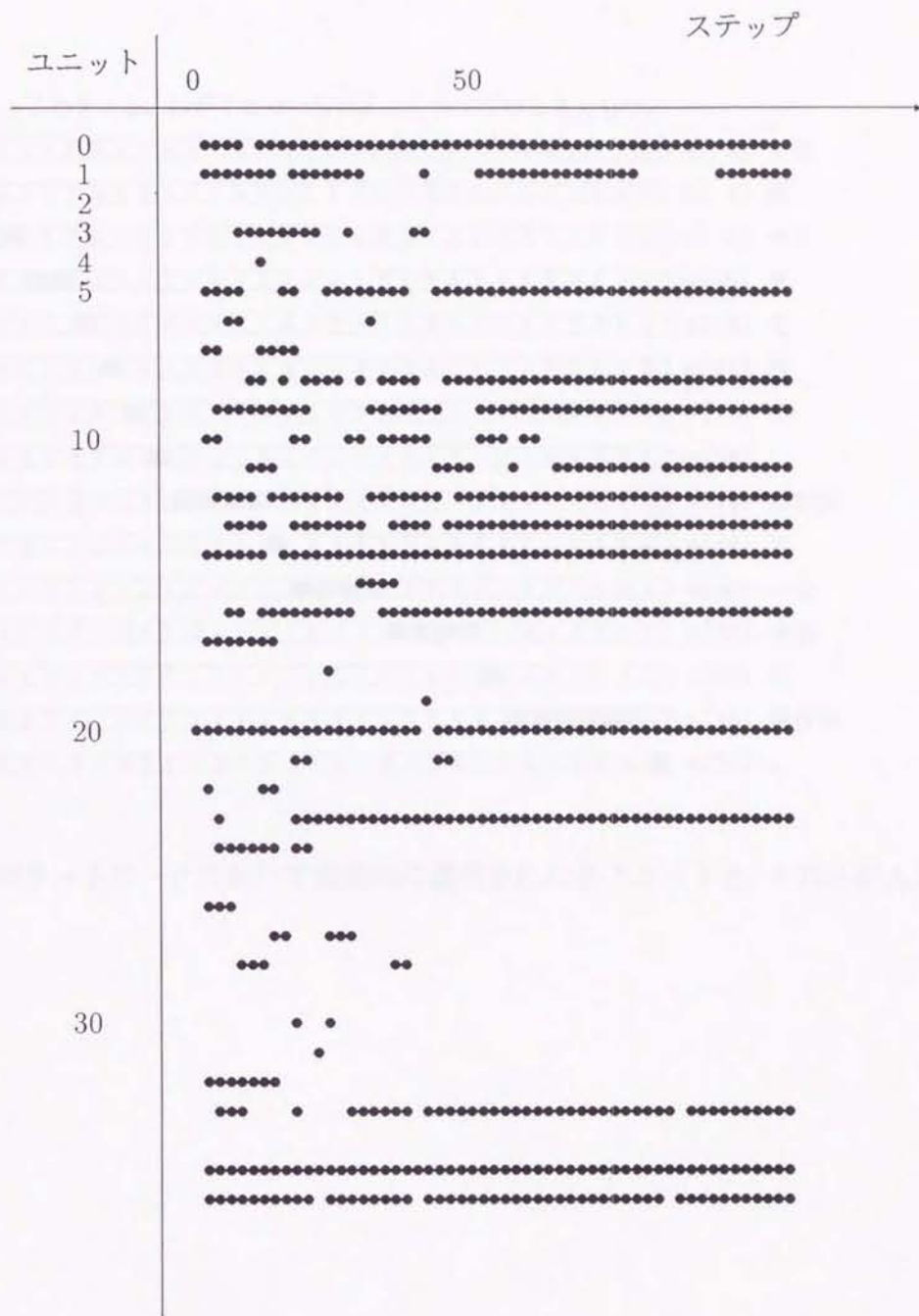


図 3.3: 例 1 のためのネットワーク内各ユニットの挙動

例2: 「はる、きいろ・十字形の花を開く一年草。」

別の例を示す。この文を解析するためのネットワーク内のユニットと、それらに割り当てられた単語は表 3.2のとおりである。

ユニット番号	単語	ユニット番号	単語
1	はる	17	花
2	きいろ	18	開く
3	十字形	19	一年草
4	の	20	。
5	花	21	。
6	開く	22	。
7	一年草	23	。
8	。	24	。
9	。	25	。
10	。	26	。
11	。	27	。
12	。	28	。
13	。	29	。
14	。	30	。
15	。	31	。
16	。	32	。

ユニット	開始位置	表記	品詞
0	0	は, 葉	名詞
1	0	は, 這	わ行五段動詞語幹
2	0	はる, 春	名詞
3	0	はる, 春	副詞
4	2	,	読点
5	3	き, 木	名詞
6	3	き, 黄	名詞
7	3	き, 来	「来る」語幹
8	3	きいろ, 黄色	名詞
9	3	きいろ, 黄色	形容詞語幹
10	4	い, 言	わ行五段動詞語幹
11	4	い, 胃	名詞
12	4	いろ, 色	名詞
13	6	.	副助詞
14	7	じゅうじがた, 十字形	名詞
15	13	の	助詞
16	14	は, 葉	名詞
17	14	は, 這	わ行五段動詞語幹
18	14	は	係助詞
19	14	はな, 花	名詞
20	15	な, 名	名詞
21	15	な, 無	形容詞語幹
22	15	な, 成	ら行五段動詞語幹
23	15	な	断定の助動詞
24	16	を	格助詞
25	17	ひら, 開	か行五段動詞語幹
26	19	く	か行五段動詞語尾
27	20	い, 言	わ行五段動詞語幹
28	20	い, 胃	名詞
29	20	いちねんそう, 一年草	名詞
30	21	ち, 散	ら行五段動詞語幹
31	26	。	句点

表 3.2: 例 2 のための形態素解析ネットワーク内のユニット

各ユニットの状態遷移は図 3.5のとおり.

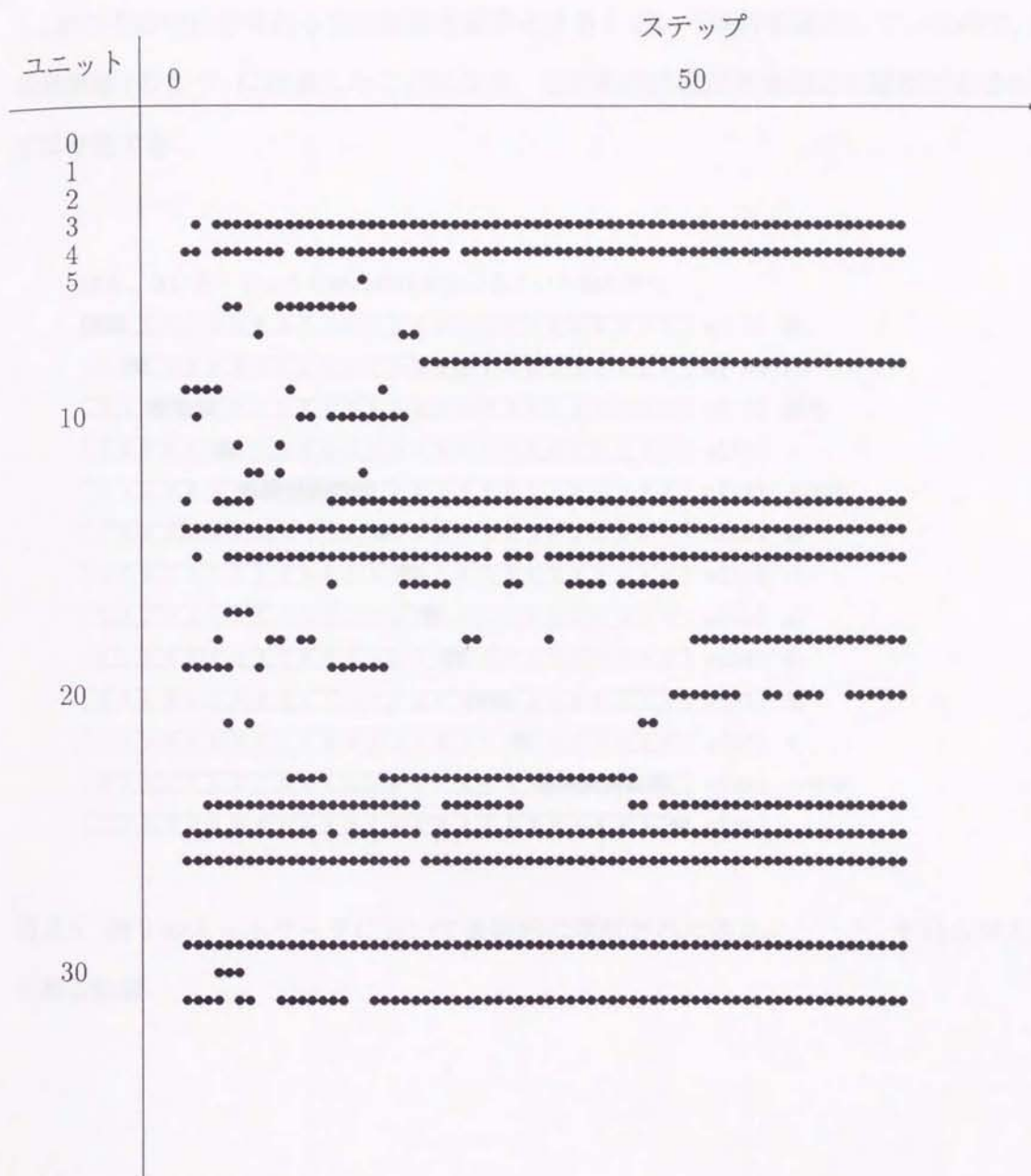


図 3.5: 例 2 のためのネットワーク内各ユニットの挙動

例3: 「秋、卵ぐらいの大きさの実がなる。」

これは自立語最少法を適用すると不適切な解析結果となる例である。入力文中に含まれる自立語の数を最少にするためには、本来ならば「大きさの実が」となるべき部分を「大きさのみが」(「大きさだけが」の意)と解析しなければならない。以下の実験結果でもそのように解析されているのがわかる。

ユニット	開始位置	表記	品詞
0	0	あ, 有	ら行五段動詞語幹
1	0	あき, 秋	副詞
2	0	あき, 秋	名詞
3	2	、	読点
4	3	た, 田	名詞
5	3	たまご, 卵	名詞
6	4	ま, 蒔	か行五段動詞語幹
7	4	ま, 巻	か行五段動詞語幹
8	6	くらい	副助詞
9	9	の	助詞
10	10	お, 尾	名詞
11	10	おお, 覆	わ行五段動詞語幹
12	10	おお, 多	形容詞語幹
13	10	おおき, 大き	形容詞語幹
14	10	おおきさ, 大きさ	名詞
15	11	お, 尾	名詞
16	12	き, 木	名詞
17	12	き, 黄	名詞
18	12	き, 来	「来る」語幹
19	13	さ, 指	さ行五段動詞語幹
20	13	さ, 咲	か行五段動詞語幹
21	13	さ, 差	さ行五段動詞語幹
22	13	さ	「する」語幹
23	13	さ	さ変動詞語尾
24	14	の	助詞
25	14	のみ	副助詞
26	15	み, 実	名詞
27	15	み, 見	一段動詞語幹
28	16	が	格助詞
29	17	な, 名	名詞
30	17	な, 無	形容詞語幹
31	17	な, 成	ら行五段動詞語幹
32	18	る	ら行五段動詞語尾
33	19	。	句点

表 3.3: 例 3 のための形態素解析ネットワーク内のユニット

各ユニットの状態遷移は図 3.7のとおり.

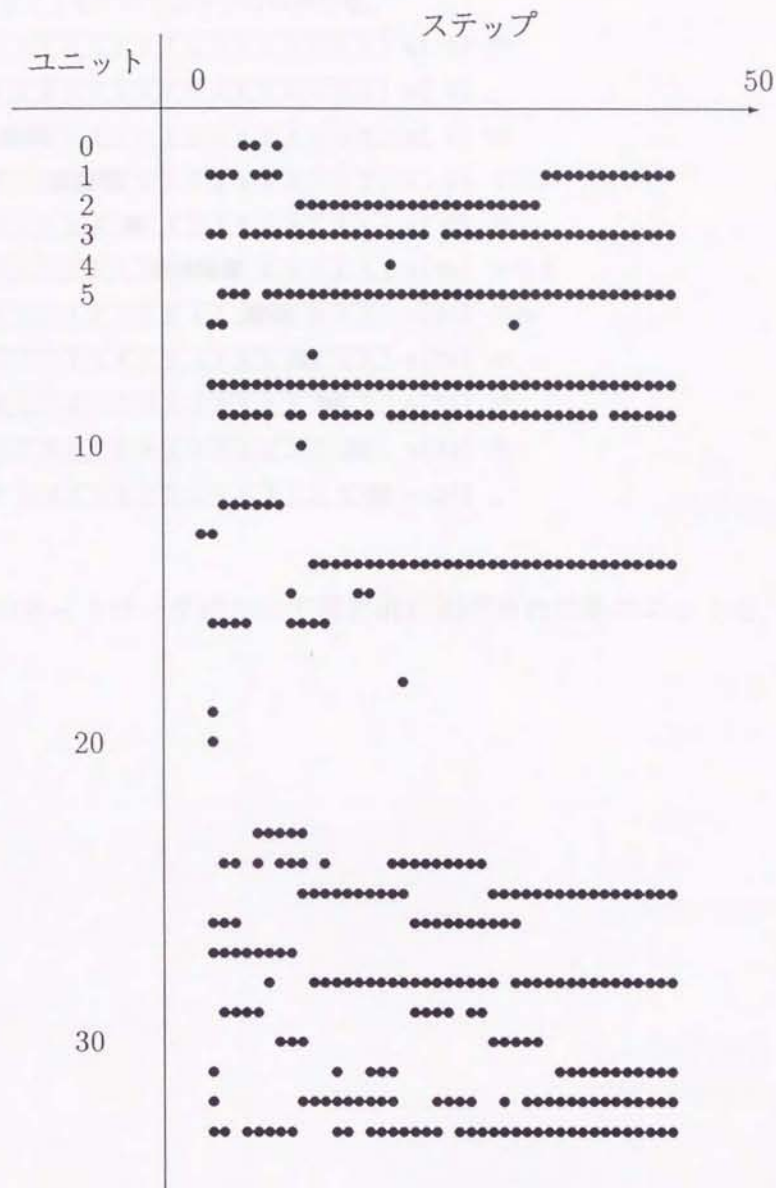


図 3.7: 例 3 のためのネットワーク内各ユニットの挙動

最終的に選択されたユニットとそれらが占める文字を図 3.8に示す.

あき、たまごぐらいのおおきさのみなる。

■■	□□□□□□□□□□□□□□□□□□□□	u[1]	秋
□□	■□□□□□□□□□□□□□□□□□□□	u[3]	,
□□□	■■■□□□□□□□□□□□□□□□□□□	u[5]	卵
□□□□□□	■■■■□□□□□□□□□□□□□□□□□□	u[8]	ぐらい
□□□□□□□□	■□□□□□□□□□□□□□□□□□□□	u[9]	の
□□□□□□□□□□	■■■■■□□□□□□□□□□□□□□□□	u[14]	大きさ
□□□□□□□□□□□□□□	■■■□□□□□□□□□□□□□□□□□□	u[25]	のみ
□□□□□□□□□□□□□□□□	■□□□□□□□□□□□□□□□□□□□	u[28]	が
□□□□□□□□□□□□□□□□□□	■□□□□□□□□□□□□□□□□□□□	u[31]	成
□□□□□□□□□□□□□□□□□□□	■□□□□□□□□□□□□□□□□□□□	u[32]	る
□□□□□□□□□□□□□□□□□□□□	■□□□□□□□□□□□□□□□□□□□	u[33]	。

図 3.8: 例 3 のネットワークにおいて最終的に選択された各ユニットと, それらが入力文字列中で占める位置.

例4: 「春、苗代に種もみを蒔き、梅雨の頃に田に植え代える。」

各ユニットとそれらに割り当てられた単語は表3.4のとおり.

ユニット	開始位置	表記	品詞
0	0	は, 葉	名詞
1	0	は, 這	わ行五段動詞語幹
2	0	はる, 春	名詞
3	0	はる, 春	副詞
4	2	、	読点
5	3	な, 名	名詞
6	3	な, 無	形容詞語幹
7	3	な, 成	ら行五段動詞語幹
8	3	なわ, 縄	名詞
9	3	なわしろ, 苗代	名詞
10	3	なわしろ	名詞
11	5	し	「する」語幹
12	5	し	さ変動詞語尾
13	5	しろ, 白	名詞
14	5	しろ, 白	形容詞語幹
15	6	ろ	終助詞
16	7	に, 似	一段動詞語幹
17	7	に	格助詞
18	8	た, 田	名詞
19	8	た	希望の助動詞
20	8	た	過去の助動詞
21	8	たね, 種	名詞
22	8	たねもみ, 種もみ	名詞
23	9	ね, 根	名詞
24	9	ね	感投詞
25	10	も, 持	た行五段動詞語幹
26	10	も	係助詞
27	11	み, 実	名詞
28	11	み, 見	一段動詞語幹
29	12	を	格助詞
30	13	ま, 蒔	か行五段動詞語幹
31	13	ま, 巻	か行五段動詞語幹
32	14	き	か行五段動詞語尾
33	15	、	読点

ユニット	開始位置	表記	品詞
34	16	つ, 付	か行五段動詞語幹
35	16	つゆ, 梅雨	名詞
36	18	の	助詞
37	19	こ, 濃	形容詞語幹
38	19	こ, 来	「来る」語幹
39	19	ころ, 頃	名詞
40	21	に, 似	一段動詞語幹
41	21	に	格助詞
42	22	た, 田	名詞
43	22	た	希望の助動詞
44	22	た	過去の助動詞
45	23	に, 似	一段動詞語幹
46	23	に	格助詞
47	24	うえ, 植え	一段動詞語幹
48	24	うえ, 上	名詞
49	24	うえかえ, 植え代え	一段動詞語幹
50	26	か	終助詞
51	28	る	一段動詞語尾
52	29	。	句点

表 3.4: 例 4 のための形態素解析ネットワーク内のユニット

このときのユニットの状態遷移は図 3.9のとおり.

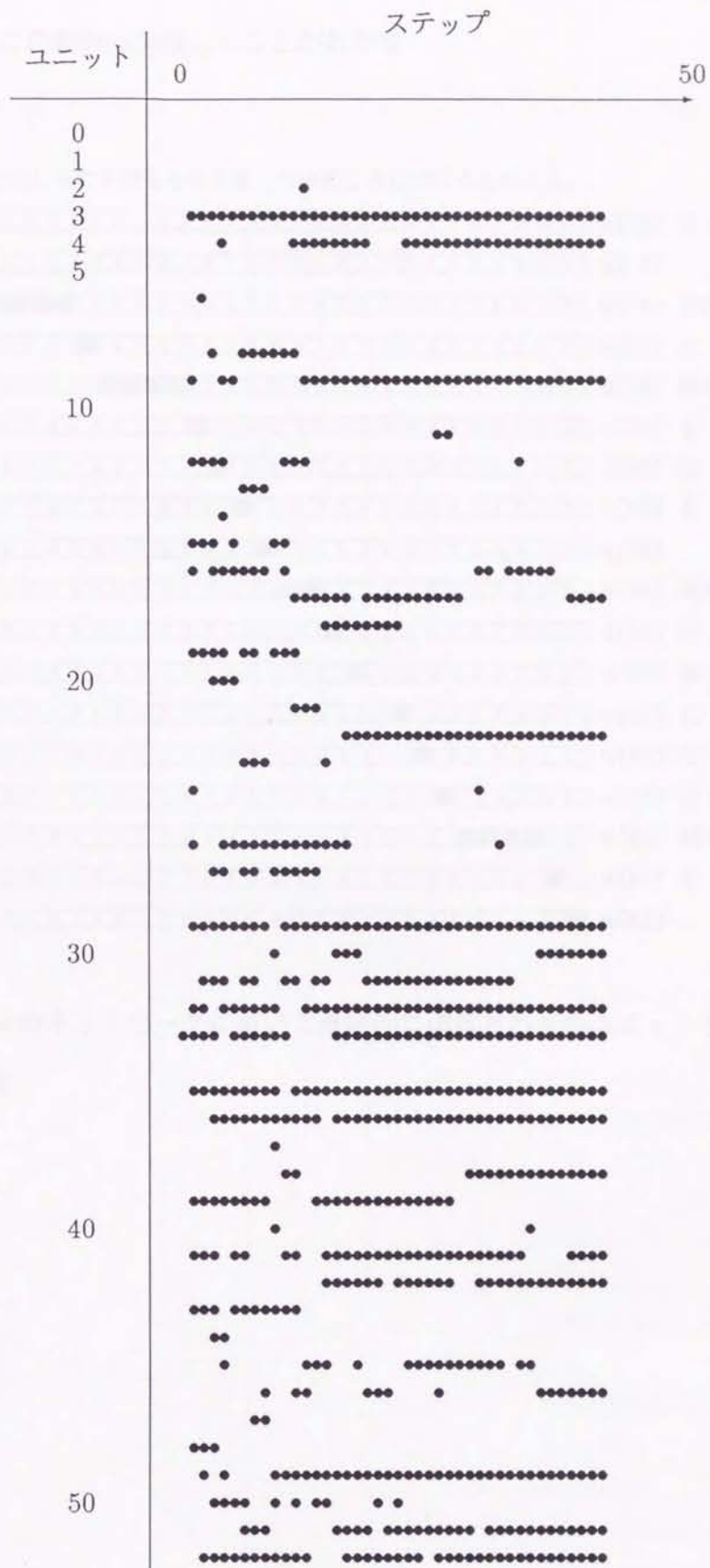


図 3.9: 例 4 のためのネットワーク内各ユニットの挙動

最終的に選択されたユニットを図 3.10に示す. 21 文字目の「ろ」がどの単語にも含まれていないことから, この解析は失敗したことがわかる.

はる、なわしろにたねもみをまき、つゆのころにたにうえかえる。

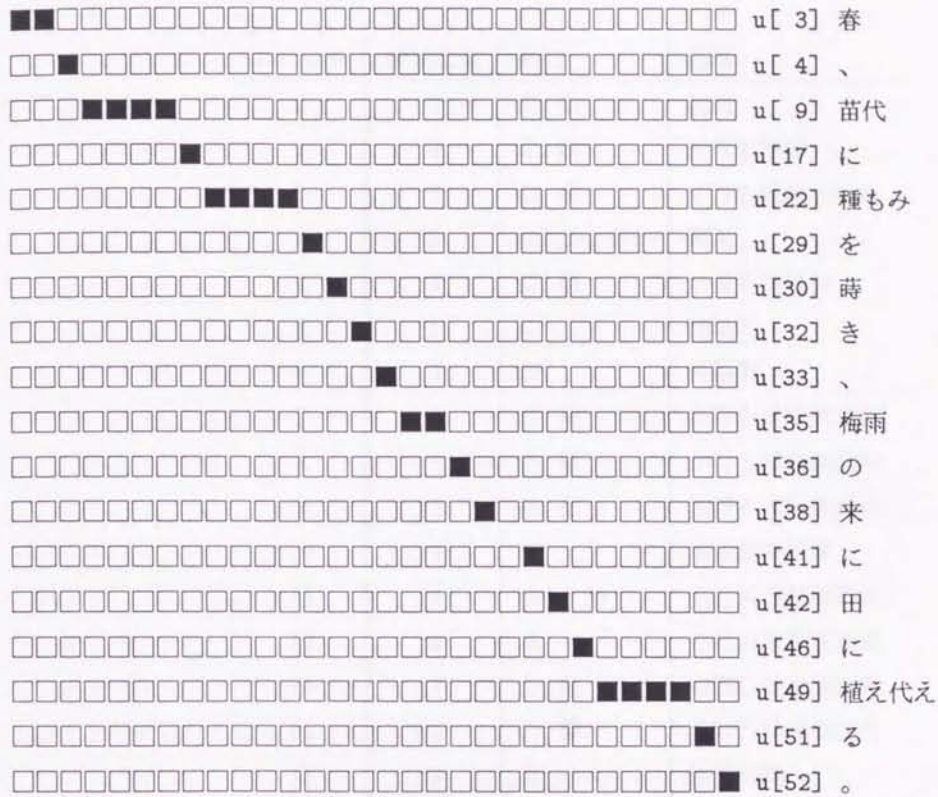


図 3.10: 例 4 のネットワークにおいて最終的に選択された各ユニットと, それらが入力文字列中で占める位置.

例5: 「何かが触ると葉を畳み、強く触れると葉の付け根から垂れ下がる。」

もう1つの失敗例である。各ユニットとそれらに割り当てられた単語は表3.5のとおり。

ユニット	開始位置	表記	品詞
0	0	な, 名	名詞
1	0	な, 無	形容詞語幹
2	0	な, 成	ら行五段動詞語幹
3	0	なにか, 何か	名詞
4	1	に, 似	一段動詞語幹
5	1	に	格助詞
6	3	が	格助詞
7	4	さ, 指	さ行五段動詞語幹
8	4	さ, 咲	か行五段動詞語幹
9	4	さ, 差	さ行五段動詞語幹
10	4	さ	さ変動詞語尾
11	4	さわ, 触	ら行五段動詞語幹
12	6	る	ら行五段動詞語尾
13	7	と, 取	ら行五段動詞語幹
14	7	と, 採	ら行五段動詞語幹
15	7	と	接続助詞
16	8	は, 葉	名詞
17	8	は, 這	わ行五段動詞語幹
18	9	を	格助詞
19	10	た, 田	名詞
20	10	たた, 畳	ま行五段動詞語幹
21	10	たたみ, 畳	名詞
22	11	た, 田	名詞
23	12	み, 実	名詞
24	12	み, 見	一段動詞語幹
25	12	み	ま行五段動詞語尾
26	13	、	読点
27	14	つ, 付	か行五段動詞語幹
28	14	つよ, 強	形容詞語幹
29	16	く	形容詞語尾
30	17	ふれ, 触れ	一段動詞語幹
31	19	る	一段動詞語尾
32	20	と, 取	ら行五段動詞語幹
33	20	と, 採	ら行五段動詞語幹

ユニット	開始位置	表記	品詞
34	20	と	接続助詞
35	21	は, 葉	名詞
36	21	は, 這	わ行五段動詞語幹
37	22	の	助詞
38	23	つ, 付	か行五段動詞語幹
39	23	つけ, 付け	一段動詞語幹
40	23	つけね, 付け根	名詞
41	24	け	か行五段動詞語尾
42	25	ね, 根	名詞
43	25	ね	否定の助動詞
44	26	か	終助詞
45	26	から, 殻	名詞
46	26	から	格助詞
47	28	た, 田	名詞
48	28	たれさが, 垂れ下が	ら行五段動詞語幹
49	32	る	ら行五段動詞語尾
50	33	。	句点

表 3.5: 例 5 のための形態素解析ネットワーク内のユニット

各ユニットの状態遷移の様子を図 3.11 に示す.

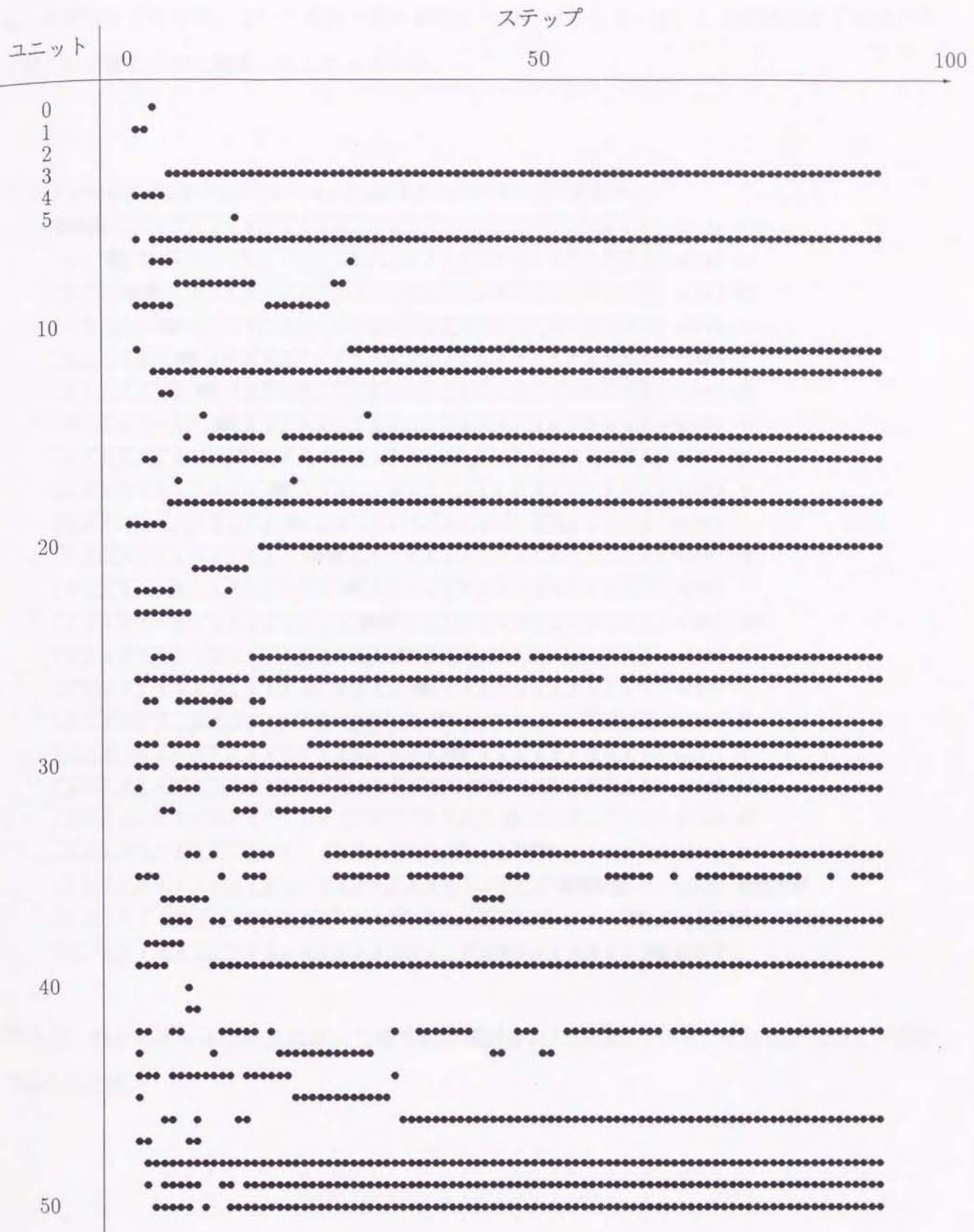


図 3.11: 例 5 のためのネットワーク内各ユニットの挙動

最終的に選択されたユニットは図 3.12のとおりである。自立語数を最少にするためには「つけね」の部分を「付け根」という名詞一語に解釈しなければならないが、この実験では「付け」と「根」の 2 語に分けて解釈してしまっている。

なにかがさわるとはをたたみ、つよくふれるとはのつけねからたれさがる。

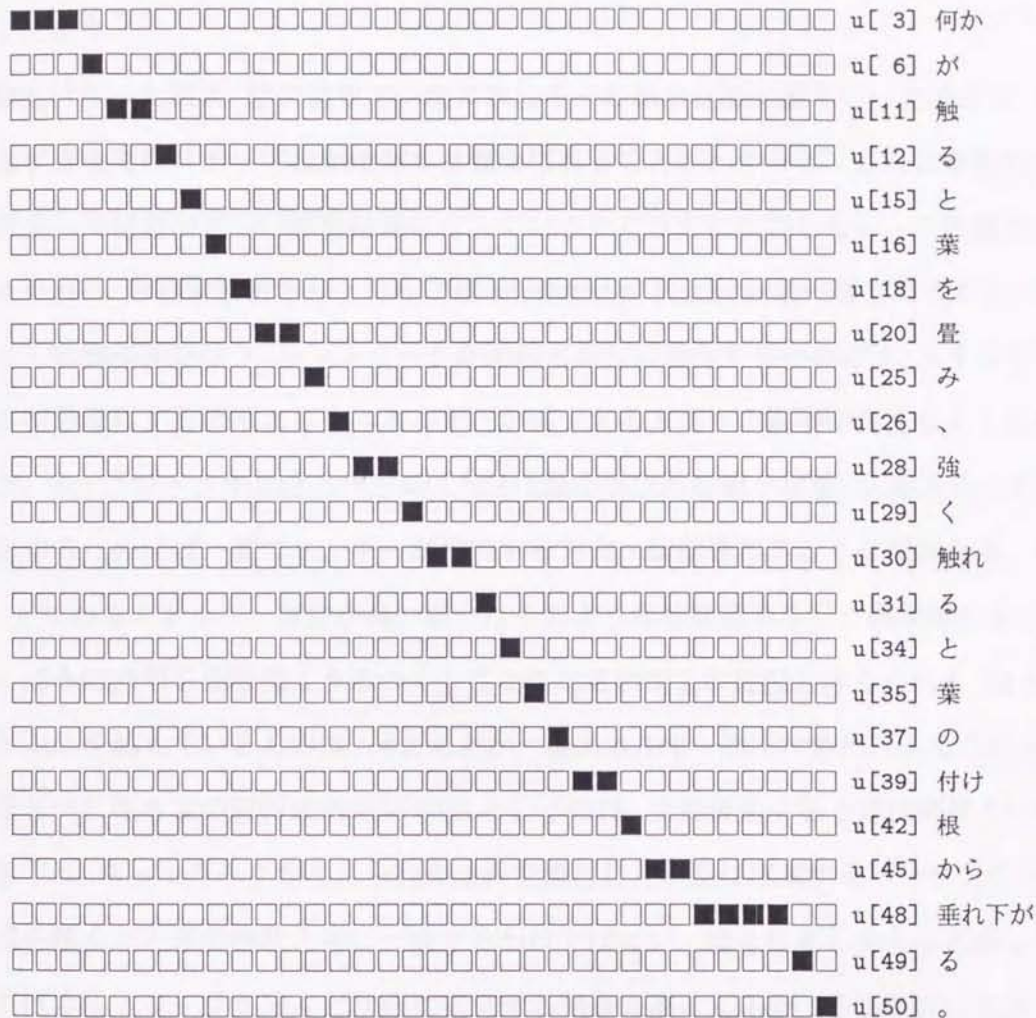


図 3.12: 例 5 のネットワークにおいて最終的に選択された各ユニットと、それらが入力文字列中で占める位置。

3.5 考察

一般に、入力文が長くなるとその中に含まれ得る単語数(すなわちネットワーク内のユニット数)は急激に増加する。しかし図 3.2を見る限り、入力文の長さが増加した場合にもネットワークが収束するまでのステップ数はさほど影響を受けないように見える。このように解析時間が入力文の長さあまり影響を受けないという点は、逐次型の自然言語解析システムにはない大きな特徴といえよう。

実験を行なった結果、解析途中で一度エネルギーの極小状態に落ちいった場合は、ネットワークが収束するまでのステップ数が増加する傾向にあることがわかった。これは最終的にネットワークの収束した状態が正しい解析結果になっているかどうかには依存しない。この理由は以下のように考えられる: 形態素解析ネットワークは Boltzmann Machine の一種であるから、ある確率にしたがって状態遷移を行う。ネットワークが期待どおりに動作するためには、エネルギーが増加する方向へ状態遷移する確率よりもエネルギーが減少する方向への確率が高くなってはならない。ところが、ネットワークがエネルギーのくぼみ(最小値以外の極小状態)に落ち込んだ場合、そこから脱出するためには一度エネルギーが増加する方向へ状態遷移を行う必要がある。くぼみが深い場合、すなわちエネルギー障壁が高い場合はそのような状態遷移を行う確率は小さく、したがってそのくぼみの内部を長時間「うろつく」ことになるのがこの原因と考えられる。図 3.2中でステップ数が 100 を越えているものは、ほとんどが一度エネルギー極小状態にとらえられたものである。

平仮名べた書き文の解析成効率が 95% というのは、文節数最少法とほぼ同等といえる。ただしこの数字は、ネットワークが与えられた条件を満たした状態に収束するパーセンテージであり、人間が文を読んだときの解釈と常に一致するわけではない。焼きなましをもっとゆっくり行うようにすればネットワークが望んだ状態に収束する割合も高くなるが、その場合にはネットワークが収束するまでにより長い時間がかかることになる。

本章で述べた日本語形態素解析用ネットワークの欠点としては以下のようなものが挙げられよう:

- 入力文中の単語間の意味的な関係が考慮されていない

形態素解析を行うための規準として、文法的な制約と自立語の数を最少にする点だけしか考慮されておらず、単語間の意味的關係が無視されている。

これに対応するにはネットワーク内の単語間の意味的結び付きの強さを表す行列を用意し、ある単語が選択されたときにはその単語と意味的に強く結び付いている単語ほど選択されやすくなる

ようにエネルギー関数を定義し直せばよい。この方法に沿ってプログラムを改良すれば前述の問題点は解消できる。また第4章で述べる係り受け解析ネットワークのエネルギー関数には、意味的な結び付きを考慮するための項がすでに導入されている。

他の問題点としては、

- 入力文ごとに異なるネットワークを作成する必要がある

本方式では、入力文が与えられるたびにその中に含まれ得る単語を調べ上げ、それらの単語をユニットとするネットワークを組み直さなければならない。ひとたびネットワークが作成されれば最適解を高速に求めることが可能であるが、解析時間全体の中でネットワークの作成に費やされる時間の割合が大きい場合にはこの高速性が生かされないおそれがある。特に入力文が短い場合には、すべての解を求めてからベストを選ぶという従来の方法よりもオーバーヘッドが大きくなる可能性がある。

- 最終的な解が1つしか得られない

ネットワークによって表されるのはただ1つの状態であるため、同じ程度にもっともらしい単語分割の方法が複数通りあった場合でも、それらのうちの1つしか表現することができない。

等がある。第5章では本章とは異なった構成のネットワークを用いることで、これらの問題点の解消を図っている。

第4章

相互結合型ニューラルネットによる係り受け解析

4.1 日本語の統語構造と文節間係り受け関係

計算機による自然言語解析では、まず形態素解析を実行し、その次に統語解析を行なって与えられた文法に従った統語解析木を作成するという手順をとることが多い。前章で形態素解析について述べたので、本章では日本語統語解析(すぐ下にあるように、ここでは係り受け解析を意味する。)を相互結合型のニューラルネットを用いて解く方法および実験結果について説明する。

橋本文法(いわゆる学校文法)では、日本語の統語的構造は文節間の係り受け関係という形で表現されるとしている[橋本 46]。橋本文法の欠点は従来からいろいろ指摘されているが¹、統語構造を簡便に表現できるために現在でも広く用いられている。本研究では日本語の統語構造の表現にこの文節間係り受け関係を用いることとする。

日本語における文節間係り受け関係とは、2文節の間の修飾・被修飾関係をさす。文節Aが文節Bを修飾するとき、「AはBに係る」「BはAを受ける」という。日本語で書かれた文あるいは句の統語的構造は、この係り受け関係を用いて記述することができる。文節間係り受けの条件としては以下の4つが必要十分であると考えられる[尾関 86]。

- R1. 最後の文節以外の文節は、自分よりも後方にある文節のどれか1つに係る
- R2. 係り受け関係は互いに交差しない
- R3. 文節中の語の意味によって係り受けの成立しやすさが影響を受ける
- R4. 位置的に近い文節間ほど係り受けが成立し易い。

¹たとえば[郡司 86]等を参照。

R1 より係り受けの総数は構文にかかわらず

$$(文節数) - 1 \quad (4.1)$$

となることがわかる。R2 は非交差条件として有名である。R3 は、例えば図 4.1 のような文においては係り受け関係 4 よりも係り受け関係 3 の方が成立し易い、ということを表している。R4 は語用論的な制約であり、統語的あるいは意味的な制約ではない。

計算機による係り受け解析では、まず入力された文節列に対してすべての可能な組み合わせを求め、次にそれらの候補の中で最も意味的整合性の高い(と思われる)ものを最終結果として出力するというステップを踏むものが多い。しかし、1つの文節列が与えられたとき、その文節列上の可能な係り受け関係の組み合わせの数は文節列長に対して指数関係以上のオーダーで増加する[尾関 86] ので、入力文節列長が長いときこの方法は計算コストの点で問題がある。

一方、係り受け解析も形態素解析と同様に組み合わせ最適化問題として定式化できるので、相互結合型ニューラルネットを用いれば長い文節列でも効率良く係り受け解析を実行することが可能である。以下では係り受け解析用ネットワークの構成と、それを用いて行なったシミュレーションの結果について述べる。

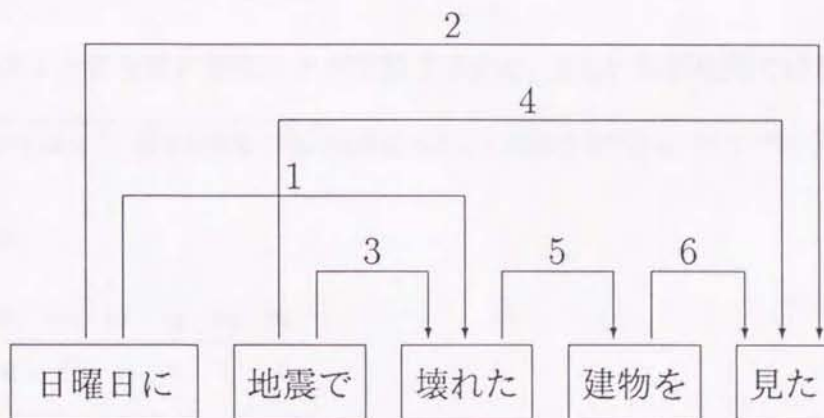


図 4.1: 文節間係り受け関係の例 (1)

4.2 ネットワークでの表現方法

係り受け解析ネットワークがなすべきことは、すべての可能な係り受け関係の中から必要十分な数の関係のみを選択することである。最終的に選択された係り受け関係は前節の条件 R1 ~ R4 を満たしている必要がある。組み合わせる対象はあくまでも「係り受け関係」であるので、係り受け解析ネットワークでは成立する可能性のある「係り受け関係」を各ユニットに割り当てる。文節そのものをユニットに割り当てるのでない点に注意されたい²。

今、「日曜日に地震で壊れた建物を見た」という文が入力されたとしよう。この場合は図 4.1 に示したように全部で 6 個のユニットが存在することになる。

係り受け解析ネットワークを定められたエネルギー関数にしたがって動作させ、それがエネルギー最小状態に収束して停止したとき値 1 をとっているユニットが最終的に選択された係り受け関係を表す。例えば図 4.1 の文には解となり得る状態が全部で 3 通りあるが、そのなかの 1 つ

unit	u_1	u_2	u_3	u_4	u_5	u_6	(4.2)
value	1	0	1	0	1	1	

は、「日曜日に」と「地震で」とが「壊れた」に、「壊れた」が「建物を」に、そして「建物を」が「見た」に係っている状態を表している³。

一方、状態

unit	u_1	u_2	u_3	u_4	u_5	u_6	(4.5)
value	1	0	0	1	1	1	

は係り受け関係 1 と係り受け関係 4 とが交差するため、正しい解析結果ではない。

²要素そのものではなく、要素の要素の間関係をユニットに割り当てるというアプローチは [Takefuji89] にも見られる。

³他の 2 状態は

unit	u_1	u_2	u_3	u_4	u_5	u_6	(4.3)
value	0	1	1	0	1	1	

および

unit	u_1	u_2	u_3	u_4	u_5	u_6	(4.4)
value	0	1	0	1	1	1	

である。

4.3 エネルギー関数の決定

条件 R1 ~ R4 を 4 つの関数 $E_1 \sim E_4$ で表現したとき, 係り受け解析ネットワークのエネルギー関数は

$$E = aE_1 + bE_2 + cE_3 + dE_4 \quad (4.6)$$

(ただし a, b, c, d は正の定数) と表される. 以下に $E_1 \sim E_4$ の定義を与える.

4.3.1 係り受け関係の数

最終的に値 1 をとるべきユニットの数は, 式 (4.1) より (文節数 - 1) となる. n をユニットの総数, m を最終的に 1 となるべきユニットの数とすると, 望ましい個数のユニットが 1 となったときに最小値をとる関数は以下のように表すことができる:

$$E_1 = \left(\sum_{i=1}^n u_i - m \right)^2. \quad (4.7)$$

4.3.2 相互排他性

各文節の係り先はただ一つである. 従って, 係り元が同一の文節であるような係り受け関係が複数個ある場合は, それらのうちの 1 つだけが値 1 をとるようにしなければならない. また, 非交差条件より互いに交差するような係り受け関係の両方が 1 となることは許されない.

このような係り受け関係間の相互排他性を表すために, 相互排他行列 X を導入する. X は $n \times n$ 行列 (n は成立する可能性のある係り受け関係の総数) であり, その (i, j) 要素の値は

$$X_{ij} = \begin{cases} 1 & \text{係り受け関係 } i \text{ と係り受け関係 } j \text{ が相互排他的である場合} \\ 0 & \text{otherwise} \end{cases} \quad (4.8)$$

である. 相互排他行列は明らかに対称行列になる. 例えば図 4.1 の文に対する相互排他行列は

$$X = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (4.9)$$

となる。

この相互排他行列 X を用いると、値 1 をとるユニット中のどの 2 つも相互排他的でない場合に最小値をとる式は

$$E_2 = \sum_{i=1}^n \sum_{j=1}^n X_{ij} u_i u_j \quad (4.10)$$

と表される。

4.3.3 文節間の意味的關係

ある 2 文節間に係り受け關係が成立するか否かは、係り側の文節と受け側の文節の文法的關係および意味的關係に依存する。係り受け解析ネットワークにおける各ユニットは、もともと文法的に成立し得る係り受け關係のみを表しているので、ここでは文節間の意味的關係を係り受けの成立し易さに反映させる手段を考えればよい。

今、 i を係り受け關係とし、 Y_i を i という係り受け關係が意味的に成立し易いほど小さい値をとる数列とする。図 4.1 の例では「地震で見た」という係り受けよりも「地震で壊れた」という係り受けの方が意味的整合性が高いと考えられるので、 $Y_3 < Y_4$ となるであろう。この Y_i を用いると、入力文全体として意味的整合性の高い係り受け關係が多く選ばれるほど小さな値をとるような式は

$$E_3 = \sum_{i=1}^n Y_i u_i \quad (4.11)$$

と表すことができる。

この Y_i を如何に設定するかは重要な問題である。今回の実験では約 500 の学習用例文から得られた単語間係り受け共起頻度に、式 (4.12) で示す操作を施したものを Y_i として用いた。

$$Y_i = \begin{cases} 0 & \xi_i = 0 \text{ のとき} \\ -1 & \xi_{max} = \xi_{min} \text{ のとき} \\ -\frac{1}{2} \left(\frac{\xi_i - \xi_{min}}{\xi_{max} - \xi_{min}} + 1 \right) & \text{その他の場合} \end{cases} \quad (4.12)$$

ただし ξ_i は i 番目のユニットによって示される係り受け関係の単語間係り受け共起頻度であり、

$$\begin{aligned}\xi_{max} &= \max\{\xi_i \mid i = 1, \dots, n\} \\ \xi_{min} &= \min\{\xi_i > 0 \mid i = 1, \dots, n\}\end{aligned}\tag{4.13}$$

である。

また単語間係り受け共起頻度とは、

文節 A が文節 B に係っているとき、文節 A 中の自立語と文節 B 中の自立語の単語間係り受け共起頻度を 1 増やす

という操作を学習用例文全体にわたって施した結果である。

式(4.12) は、1) 単語間係り受け共起頻度の符号を反転し、さらに、2) その絶対値が $[0, 1]$ の間におさまるような変形を行う。符号を反転するのは、単語間係り受け共起頻度は意味的整合性が高いほど大きな値をとるのに対し、 Y_i は意味的整合性が高いほど小さい値をとる必要があるためである。

また絶対値を一定の範囲におさめるのは以下の理由による：入力文が異なればその中で用いられる単語も異なる。用いられる単語が異なれば単語間係り受け共起頻度のとる値の範囲も変化する。入力文ごとに意味的整合性を表す値の範囲が変化するのでは安定したパラメータを得ることが困難であるので、どのような入力文に対しても一定の範囲内におさまるようにする必要がある。

式(4.12) は 0 以外の値をとる単語間係り受け共起頻度を $[-1.0, -0.5]$ におさまるよう正規化する。ただし、単語間係り受け共起頻度が 0 のときはそのまま変化させない。0 の場合を特別扱ったのは、学習用例文中で一度も係り受け関係を生じなかった単語の組み合わせをそうでないものから差別化し、それによって入力文中の単語間係り受け共起頻度の範囲が広い場合の解析成功率を向上させることを目的としたためである。

4.3.4 文節間の距離

日本語では一般的に、文節間の距離が小さいほど係り受けが生じやすいという傾向がある。係り受け関係 i の係り元を k_i 番目の文節、係り先を l_i 番目の文節とすると、文節間の距離が短い係り受け関係が選ばれるほど小さな値を取る式は

$$E_4 = \sum_{i=1}^n Z_i u_i\tag{4.14}$$

と表すことができる。ただし、

$$Z_i = l_i - k_i \quad (4.15)$$

とする。

4.3.5 係り受け解析ネットワークのエネルギー関数

以上により決定された $E_1 \sim E_4$ を式 (4.6) に代入して整理し、式 (2.6) の係数を比較すると、ユニット間の結合の強さ w_{ij} および各ユニットのしきい値 θ_i は

$$\begin{cases} w_{ij} = -2(a + bX_{ij}) \\ \theta_i = (1 - 2m)a + cY_i + dZ_i \end{cases} \quad (4.16)$$

とすればよいことがわかる。

4.4 実験

形態素解析の場合と同様、定量的なデータを収集する前に、全体の約1/4の入力データを使って予備実験を行ない、それによってネットワーク内の各定数を手作業で決定した。

まず、式(4.6)の a, b, c, d は以下のように決定した。

$$a = 6, \quad b = 5, \quad c = 15, \quad d = 1. \quad (4.17)$$

時刻 t のときのネットワークの温度 $T(t)$ は次の式で与えられるものとした。

$$T(t) = \frac{T_0}{1 + t/\tau}. \quad (4.18)$$

ただし

$$T_0 = 10, \quad \tau = 10 \quad (4.19)$$

である。

また、ユニットの初期状態はすべて0とし、状態遷移は非同期に行なった。計算機環境は形態素解析の場合と同一である。

係り受け解析ネットワークへ入力するデータは、すでに形態素解析が済んで切り分けが終了した文節列である。各文節には、

- その文節内の自立語に関する情報
- 連体修飾句・連用修飾句のように、その文節がどのような文節を修飾するかを示す情報

が付加されている。たとえば図4.1の例に対応した示した入力はおおむね表4.1のようになる。

ユニット	文節内自立語	自立語の品詞	修飾機能
1	日曜日	名詞	連用修飾
2	地震	名詞	連用修飾
3	壊れた	動詞	連体修飾
4	建物	名詞	連用修飾
5	見た	動詞	—

表 4.1: 図 4.1 に対応する入力文節列

解析対象の文としては、単語間係り受け共起頻度の学習用例文中に現れた単語のみから構成されているものを選んだ。解析対象文・学習用例文はともに文献 [金田一 74] 中の植物に関する語義説明文を用いた。今回の実験では解析対象文の一部をそのまま学習用例文として用いたが、このように学習用例文中の係り受け関係と解析対象文中の係り受け関係とが類似した傾向のものである場合は、入力した 200 文のうち 95% 以上の文において正しい係り受け関係が得られることがわかった。解析した文全体の約 5% は、エネルギーの極小値に捕まって脱出できず、誤った最終状態に収束した。ユニットの状態遷移は非同期かつランダムに生じるので、同じ文を解析しても異なる結果が導かれる場合もある。

本実験で係り受け解析ネットワークが安定状態に収束するまでに必要とされたステップ数を図 4.2 に示す。ここではすべてのユニットが平均して 1 回ずつ発火するのに要する時間を 1 ステップとしている。前述のとおり状態遷移は非同期かつランダムに生じるので、同じ文を解析しても収束までのステップ数が同一になるとは限らないが、数十回繰り返した実験のいずれの場合においても図 4.2 と大差ない結果が得られた。

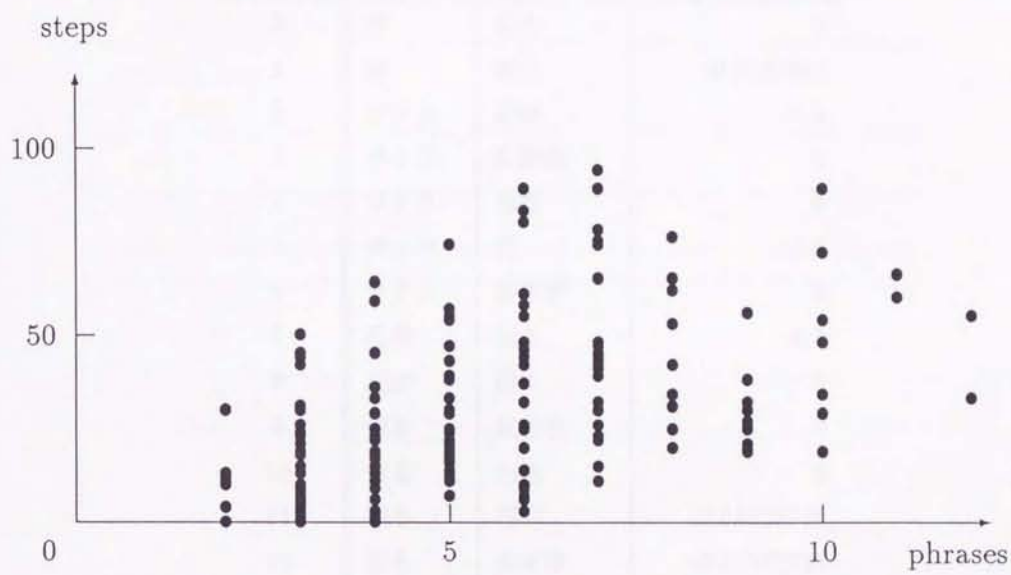


図 4.2: 入力文中の文節数と係り受け解析ネットワークが収束するまでのステップ数との関係

以下に個々の入力に対しての解析結果をいくつか示す。

例1: 「春、 / サクラの / 花卉に / 似た / 紅紫色・ / 白色の / 花を / 開く / 多年草。」

/ は文節の切れ目を示す。この例文には9個の文節が含まれており、それらの間の可能な係り受け関係は以下の19個がある。ネットワーク内のそれぞれのユニットはこの係り受け関係に対応する。

ユニット	係る文節	受ける文節	意味的整合性 Y_i
0	春	似る	0
1	春	開く	-0.6136364
2	サクラ	花卉	-0.5
3	サクラ	紅紫色	0
4	サクラ	白色	0
5	サクラ	花	-0.5
6	サクラ	多年草	0
7	花卉	似る	-0.5
8	花卉	開く	0
9	似る	紅紫色	0
10	似る	白色	0
11	似る	花	-0.58522725
12	似る	多年草	-0.52272725
13	紅紫色	白色	0
14	紅紫色	花	-0.51704544
15	紅紫色	多年草	0
16	白色	花	-0.57954544
17	白色	多年草	0
18	花	開く	-0.9943182
19	開く	多年草	-0.54545456

表 4.2: 例1のための係り受け解析ネットワーク内のユニット

各ユニットの状態遷移の様子を図 4.3 に示す。ユニットが値 1 をとっている部分が黒い点で表されている。

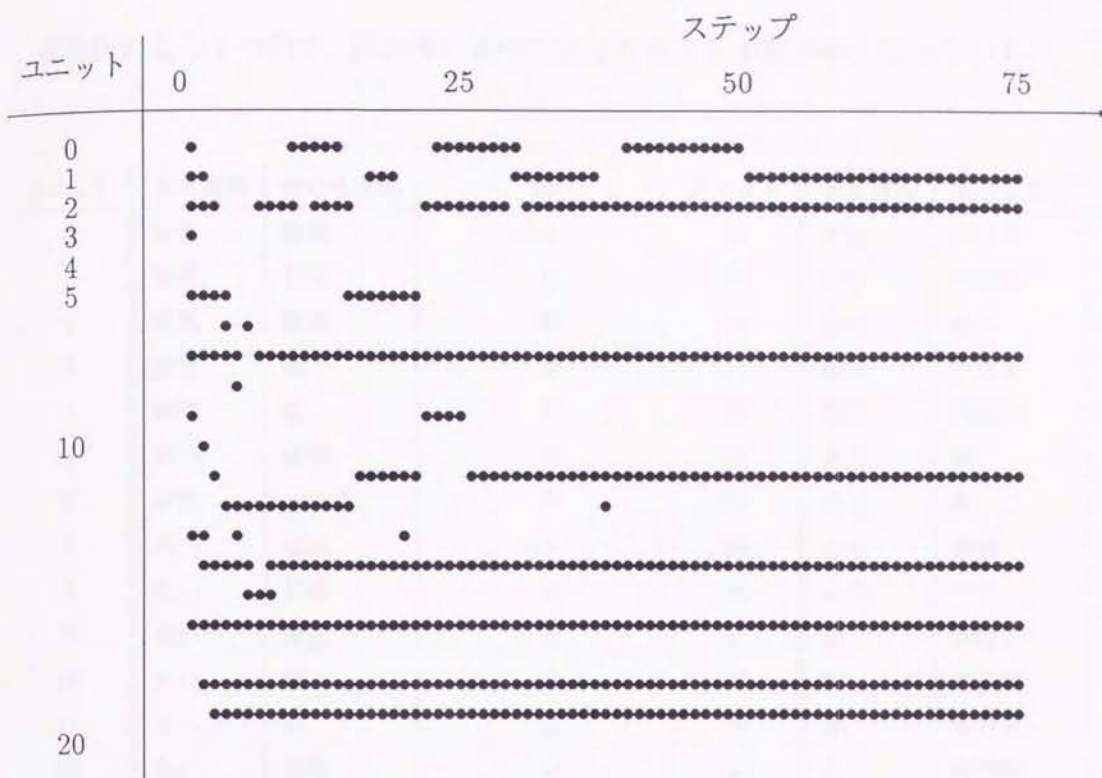


図 4.3: 例 1 のためのネットワーク内各ユニットの挙動

最終的に選択されたユニットの番号と、そのユニットが表す係り受け関係は図 4.4 のとおり。この係り受け解析が成功していることがわかる。

- u[1] : 春 → 開く
- u[2] : サクラ → 花卉
- u[7] : 花卉 → 似る
- u[11] : 似る → 花
- u[14] : 紅紫色 → 花
- u[16] : 白色 → 花
- u[18] : 花 → 開く
- u[19] : 開く → 多年草

図 4.4: 例 1 のネットワーク内で最終的に選択された各ユニットと、それらが表す係り受け関係。

例2: 「黄色で / 丸い / 果実は / 甘味・ / 酸味が / あり、 / 秋から / 冬に / かけての / 代表的
な / 果物の / 一つ。」

成功例をもう1つ示す. 前の例に比べてかなりユニット数が増えている.

ユニット	係る文節	受ける文節	Y_i	ユニット	係る文節	受ける文節	Y_i
0	黄色	果実	-0.5	18	甘味	かける	0
1	黄色	甘味	0	19	甘味	代表的	0
2	黄色	酸味	0	20	酸味	ある	-0.5833333
3	黄色	秋	0	21	酸味	かける	0
4	黄色	冬	0	22	酸味	代表的	0
5	黄色	果物	0	23	ある	秋	0
6	黄色	一つ	0	24	ある	冬	0
7	丸い	果実	-0.5	25	ある	果物	0
8	丸い	甘味	0	26	ある	一つ	0
9	丸い	酸味	0	27	秋	かける	-0.9166667
10	丸い	秋	0	28	秋	代表的	0
11	丸い	冬	0	29	冬	かける	-0.5
12	丸い	果物	0	30	冬	代表的	0
13	丸い	一つ	0	31	かける	果物	0
14	果実	ある	-0.5833333	32	かける	一つ	0
15	果実	かける	0	33	代表的	果物	-0.5
16	果実	代表的	0	34	代表的	一つ	0
17	甘味	ある	-0.5	35	果物	一つ	-0.5

表 4.3: 例2 のための係り受け解析ネットワーク内のユニット

各ユニットの状態遷移の様子を図 4.5 に示す。例 1 に比してユニット数がかかなり多くなっているが、ネットワークが収束するまでのステップ数にはあまり変化がないことがわかる。

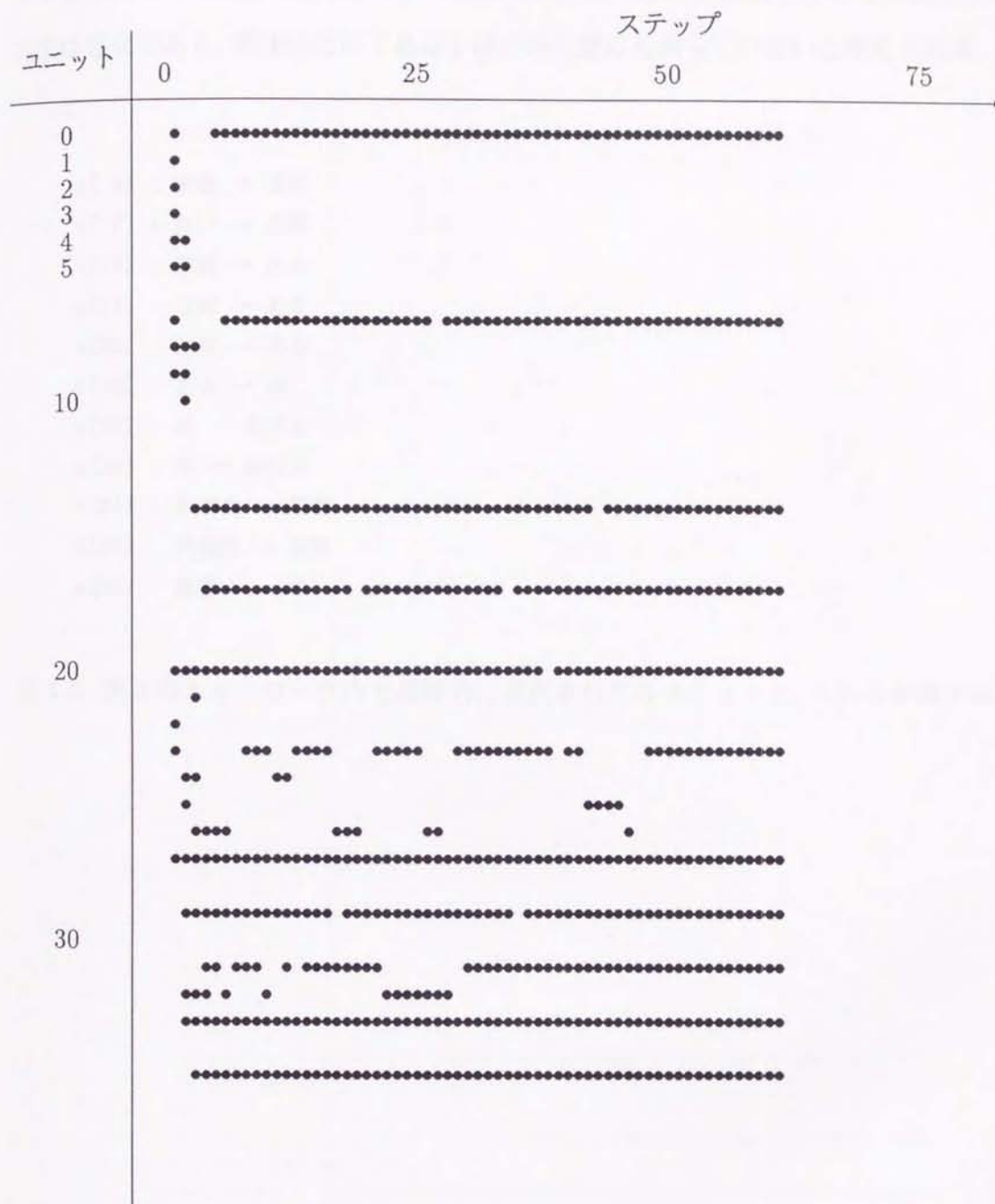


図 4.5: 例 2 のためのネットワーク内各ユニットの挙動

最終的に選択されたユニットの番号と、そのユニットが表す係り受け関係は図 4.6のとおり。「ある」の係り先が「秋」になっているのは不自然だが、これは「すべての文節は自分よりも後方にあるどれか1つの文節に係る」という制限を満足させるためにはしかたない結果である。(この入力文は重文であり、意味的には「ある」はどの文節にも係っていないと考えられる。)

- u[0] : 黄色 → 果実
- u[7] : 丸い → 果実
- u[14] : 果実 → ある
- u[17] : 甘味 → ある
- u[20] : 酸味 → ある
- u[23] : ある → 秋
- u[27] : 秋 → かける
- u[29] : 冬 → かける
- u[31] : かける → 果物
- u[33] : 代表的 → 果物
- u[35] : 果物 → 一つ

図 4.6: 例 2 のネットワーク内で最終的に選択された各ユニットと、それらが表す係り受け関係.

例3: 「夏から / 秋に / かけて / 薄い / 黄白色の / 花を / 開く。」

ネットワーク内の各ユニットは以下のとおり.

ユニット	係る文節	受ける文節	意味的整合性 Y_i
0	夏	かける	-0.52840906
1	夏	薄い	0
2	夏	開く	-0.59659094
3	秋	かける	-0.52840906
4	秋	薄い	0
5	秋	開く	-0.53977275
6	かける	薄い	0
7	かける	開く	-0.5
8	薄い	黄白色	-0.5
9	薄い	花	0
10	黄白色	花	-0.51704544
11	花	開く	-0.9943182

表 4.4: 例3のための係り受け解析ネットワーク内のユニット

各ユニットの状態遷移の様子を図 4.7に示す.

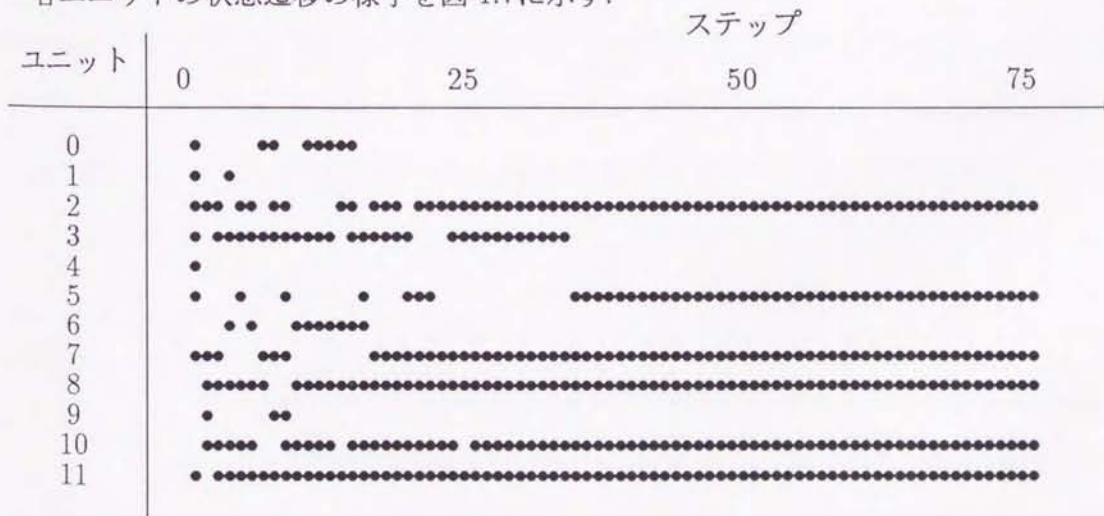


図 4.7: 例3のためのネットワーク内各ユニットの挙動

最終的に選択されたユニットは図 4.8のとおり. $u[2]$ の係り受けの意味的整合性が $u[0]$ のそれよりも大きい結果, 「夏」が「開く」に係るという意味的には誤った解析結果となっている. 「秋」の係り先に関しても同様である. ただしネットワークとしてはエネルギーが最も小さい状態に収束しており, 純粋に最適化問題と考えた場合には正解を得たことになる. この例から係り受けの意味的整合性の設定が重要であることがわかる.

- $u[2]$: 夏 → 開く
- $u[5]$: 秋 → 開く
- $u[7]$: かける → 開く
- $u[8]$: 薄い → 黄白色
- $u[10]$: 黄白色 → 花
- $u[11]$: 花 → 開く

図 4.8: 例 3 のネットワーク内で最終的に選択された各ユニットと, それらが表す係り受け関係.

例4: 「夏から / 秋に / かけて / 赤紫の / 色の / 花を / つける。」

失敗した例を示す. この入力に対するネットワークの各ユニットは以下のとおりであるが, u[7] (「色」と「花」の係り受け) の意味的整合性が0であるのは常識的に考えて少しおかしい. 単語間係り受け共起頻度を作成する際の例文が偏っていたと思われる.

ユニット	係る文節	受ける文節	意味的整合性 Y_i
0	夏	かける	-0.85714287
1	夏	つける	-0.5
2	秋	かける	-0.85714287
3	秋	つける	-0.64285713
4	かける	つける	-0.64285713
5	赤紫	色	-0.5
6	赤紫	花	-0.5
7	色	花	0
8	花	つける	-0.9285714

表 4.5: 例4のための係り受け解析ネットワーク内のユニット

状態遷移の様子を図 4.9に示す. 他の例に比べてネットワークが収束するまでに非常に長い時間がかかっていることがわかる (横軸のスケールが2倍になっていることに注意). この例に限らず, 解析が失敗した場合は収束までに時間がかかることが多い. これは前章の考察の節でも述べたように, ネットワークがエネルギーの極小状態に捕まってしまった結果と考えられる.

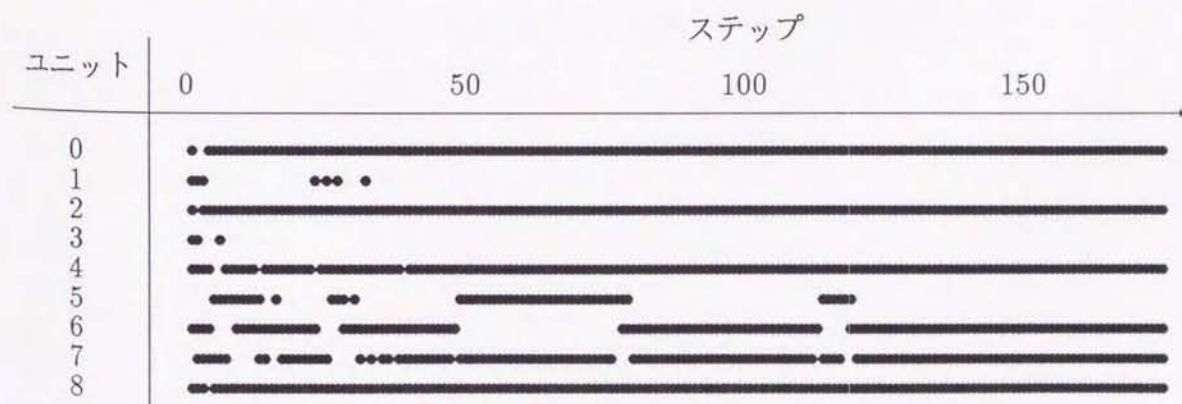


図 4.9: 例4のためのネットワーク内各ユニットの挙動

最終的に選択されたユニットの番号と、そのユニットが表す係り受け関係は図 4.10のとおり。「赤紫」の係り先として「花」と「色」の2つが可能であるが、どちらの係り受けの意味的整合性も -0.5 と等しい。この場合は文節間の距離が短い方、すなわち「色」に係るべきであるが、ネットワークはその解を見つけられなかった。

- u[0] : 夏 → かける
- u[2] : 秋 → かける
- u[4] : かける → つける
- u[6] : 赤紫 → 花
- u[7] : 色 → 花
- u[8] : 花 → つける

図 4.10: 例 4 のネットワーク内で最終的に選択された各ユニットと、それらが表す係り受け関係。

例5: 「細長い / さやの / 中に / 黒味を / 帯びた / 赤色で / 小形の / 種を / 結ぶ。」

もう1つの失敗例である. ネットワーク内の各ユニットは以下のとおり.

ユニット	係る文節	受ける文節	意味的整合性 Y_i
0	細長い	さや	-0.5
1	細長い	中	0
2	細長い	黒味	0
3	細長い	赤色	0
4	細長い	小形	0
5	細長い	種	-0.5
6	さや	中	-0.5
7	さや	黒味	0
8	さや	赤色	0
9	さや	小形	0
10	さや	種	0
11	中	帯びる	0
12	中	結ぶ	-0.5
13	黒味	帯びる	-0.75
14	黒味	結ぶ	0
15	帯びる	赤色	-0.5
16	帯びる	小形	0
17	帯びる	種	0
18	赤色	小形	0
19	赤色	種	-0.5
20	小形	種	-0.5
21	種	結ぶ	-0.5

表 4.6: 例5のための係り受け解析ネットワーク内のユニット

状態遷移の様子を図 4.11に示す.

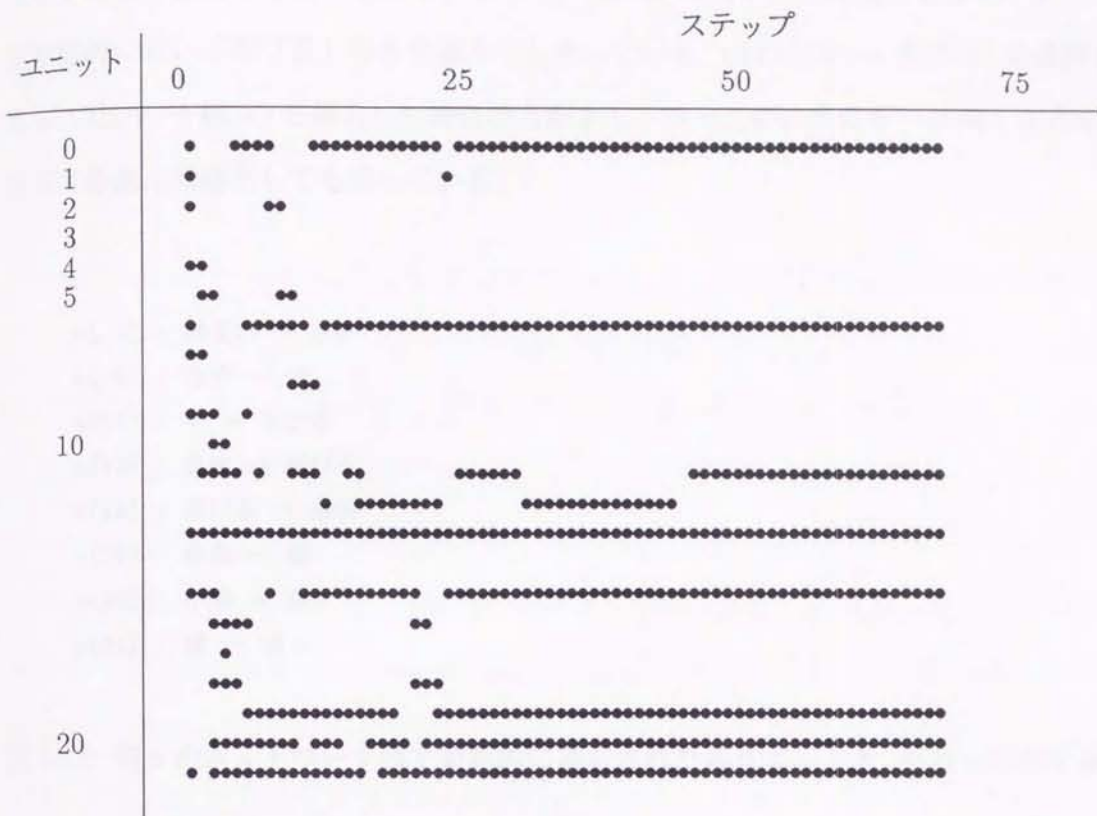


図 4.11: 例 5 のためのネットワーク内各ユニットの挙動

最終的に選択されたユニットの番号と、そのユニットが表す係り受け関係は図 4.12のとおり。「中」の係り先としては「帯びる」よりも「結ぶ」の方が意味的整合性が高いが、ネットワークは位置的に近い「帯びる」の方を選んでしまっている。u[11](中 → 帯びる)を選択した場合よりもu[12](中 → 結ぶ)を選択した場合の方がネットワークのエネルギーは低くなるので、この解析結果は最適化問題としても誤っている。

- u[0] : 細長い → さや
- u[6] : さや → 中
- u[11] : 中 → 帯びる
- u[13] : 黒味 → 帯びる
- u[15] : 帯びる → 赤色
- u[19] : 赤色 → 種
- u[20] : 小形 → 種
- u[21] : 種 → 結ぶ

図 4.12: 例 5 のネットワーク内で最終的に選択された各ユニットと、それらが表す係り受け関係。

4.5 考察

一般に逐次型のアルゴリズムを用いて CFG の解析を行なった場合は、最善でも文の長さの 3 乗に比例した時間がかかることがわかっている。ところが図 4.2 では文節数の増加に比してステップ数はそれほど増加しないように見える⁴。このことは係り受け解析にネットワークを用いる大きなメリットと言えるだろう。

一般にネットワーク内のユニット数が少なければ少ないほど解くべき問題は簡単であり、したがってネットワークはより少ないステップ数で収束するものと考えられる。しかし図 4.2 の中には、文節数 = 2、すなわち係り受け関係が 1 つしかないにもかかわらず収束までに数十ステップを必要とした例が見られる。これはネットワークの初期温度が比較的高く、そのため動作開始からしばらくの間はエネルギーが増加する方向への状態遷移が生じやすいためと考えられる。ネットワークの初期温度 T_0 を小さくとればこのようなむだな状態遷移は生じにくくなるが、温度が低い状態でエネルギーの極小状態から脱出するには長時間を要する。これは現在よりもゆっくりと焼きなましを行なわなければならないことを意味し、ひいては解析全体のステップ数の増加につながる。たとえむだな状態遷移があったとしても、ネットワークが収束するまでのステップ数が少ない方が工学的には有効であると言えよう。

式 (4.16) のネットワークに式 (4.17) のパラメータを与えた場合に、エネルギー最小の状態が常に文法的に正しい係り受けの組を表すとは限らない。例えば、図 4.13 に示すような文節数 = 3、係り受け数 = 3 の入力があったとする。また、このとき意味的整合性を表す数値である Y_i は、

$$Y_1 = -1, Y_2 = 0, Y_3 = -1 \quad (4.20)$$

で与えられるものとする。このときエネルギー最小となるのは

$$u_1 = 1, u_2 = 0, u_3 = 1 \quad (4.21)$$

という状態であるが、これは文法的に正しい係り受けの組ではない。

それにもかかわらず、図 4.2 の実験結果においては文法的に誤った解析結果となったものは 1 つもなかった。この理由は以下のように考えられる：今回の実験では係り受け共起頻度の学習に用いた例文と解析用入力文とが同一分野の文であった。したがって意味的整合性の高い係り受けの組み合わせを選択することがそのまま文法的にも正しい組み合わせを選択することにつながると考えられる。

⁴もっとも図 4.2 に示された情報だけから、文節数と収束までのステップ数の間に強い相関関係があるとは言えない。

また予備実験の結果得られた一連の定数(式(4.17))を見ると、係り受けの意味的整合性を表す関数 E_3 の係数 c が、文法的正当性を表す関数 E_1 および E_2 の係数 a および b よりも大きくなっているが、これも学習用例文と解析用入力文が同一分野であったためと考えられる。

ちなみに単語間係り受け共起頻度をまったく考慮しなかった場合、すなわち式(4.17)の c を0に変更し、

$$a = 6, b = 5, c = 0, d = 1 \quad (4.22)$$

とした場合の解析成功率は44%に留まった。誤った解析結果のほとんどでは、位置的にもっとも近い文節が係り先として選ばれていた。これは、単語間係り受け共起頻度という意味的な制約がなくなり、位置的に近い係り受けほど成立しやすいという語用論的な制約のみが残っている状態から容易に予測される結果である。なお、この場合も文法的に誤った解析結果は得られなかったが、これは文法的正当性を表す関数 E_1 および E_2 の係数 a および b が、係り受けの距離を反映する関数 E_4 の係数 d に比して大きいためと考えられる。

本章で述べた係り受け解析方法は、前章の形態素解析と対になる方法である。したがって前章で指摘した問題点のうちいくつかはそのまま本章にもあてはまる：まず第1に入力文ごとに異なったネットワークを作成しなくてはならない点であるが、係り受け解析では辞書引き操作が不要であるため、前章の形態素解析ネットワークに比べてネットワーク自体の作成が容易である。またネットワーク内のユニット数は入力文節列の統語的情報のみに依存し、解析が可能な語彙数とは関係がないという点も前章の形態素解析ネットワークに比べて有利な点である。第2に最終的な解析結果が1つだけしか得られない点であるが、これは相互結合型ネットワークで組み合わせ問題を解く場合に共通した問題であり、この枠組の中では解決できない。これを解決するにはネットワークの外部に何らかの監視・調節用システムを用意する必要があるだろう。



図 4.13: 文節間係り受け関係の例(2)

4.6 単語共起頻度に関する議論

一般に日本語文の意味解析には、格文法(あるいは結合価文法)と意味素性を組み合わせて用いることが多い。すなわち、動詞を文の意味的中心と考え、その動詞の格要素としてどのような意味素性を持つ語が許されるかの記述をもとに入力文の意味的整合性を求める方法である。これに対して本研究では文節と文節との意味的關係度を単語間係り受け共起頻度のみから計算している。これに対する批判としては以下の2つが予想できる。

1. (意味素性を支持する立場からの批判)

個々の単語間の共起頻度を考えるのでは、単語に共通する意味的特長を把握することはできないであろう。意味素性を用いて名詞のグループ化を図るべきだ。

2. (格文法・結合価文法を支持する立場からの批判)

実際の日本語文においては係り受け関係にある2文節の間には何らかの統語的關係、たとえば一方が動詞でもう一方がその格要素であるといった関係、が存在している。単語間係り受け関係にはこのような統語的情報が含まれておらず、したがって十分な解析ができないはずだ。

第1の批判に対しては、まず「意味素性を用いる方法と単語自体の共起頻度を用いる方法に本質的な違いはない」という点を指摘しておきたい。なぜなら、より精密な解析を行うためにはより多くの意味素性を設定する必要があるが、これを押し進めていくと最終的には1単語1意味素性となって個々の単語の間を関係を考えるのと同じになるからである。また意味素性としてどのようなものを設定すべきかという問題には研究者の間でも合意がとれておらず、それぞれが独自に設定した意味素性に基づいて研究を行なっているのが現状である。また人間がシステムに対してトップダウンに与えた意味素性が実際の入力文に対してうまく機能するかどうかの保証はない。これに対して個々の単語の間の共起頻度を利用する方法は現実の例文に基づいてボトムアップにデータを収集するのでデータ自体の精度が高く、したがって精密な意味解析が期待できる。

単語間共起頻度を利用する際の問題点としては、精密なデータを用意するためには多数の高品位な例文を準備する必要があるという点が挙げられる。また多数の例文から個々の単語間の共起頻度を抽出するための効率的なツールも必要である。これらに関しては、近い将来に大規模なコーパスおよび各種ユーティリティソフトウェアが公開されることを期待したい。単語間共起頻度を日本語解析に応用した研究は最近盛んになってきており、[高橋 88][高橋 89a][高橋 89b]な

どのほか、田中らの一連の研究 [田中・吉田 88][田中・吉田 89](その他多数), [松本 89], [荻野 89], [松川 89] などがある。

第2の批判はもっともなものであるが、統語的關係をも盛り込んだ単語間共起頻度情報を作成するためには非常に大きなテーブルを用意しなくてはならず、効率の良い情報圧縮技術と圧縮されたデータの検索技術が不可欠である。また現状では用意することのできる単語共起頻度情報そのものがさほど多くないために、統語的關係を考慮したデータテーブルを作成してもほとんどのエントリは空白のままになるであろう。白井らは単語共起頻度を日本語係り受け解析に応用したが、そこでは自立語用共起マトリクスと付属語用共起マトリクスを別々に持たせることで情報の圧縮を図っている [鈴木 86][白井 86]。統語情報と共起頻度情報の両者を本格的に統合して利用するためには、さらに工夫が必要と思われる。

第5章

文脈情報を利用した形態素解析

5.1 導入

本章ではまず最初にニューラルネットを用いて文脈を表現する方法について述べ、次いでその表現された文脈情報を形態素解析に利用することを試みる。

第3章では形態素解析を組み合わせ最適化問題に帰着させ、それを相互結合型ニューラルネットを用いて解く方法について述べた。そこでは入力文の構成要素となる可能性のあるすべての単語を用意しておく必要があった。そのためにはあらかじめ入力文字列を部分文字列に分解し、各部分文字列を表記として持つ単語が辞書中に存在するかどうかを調べなければならないが、これをソフトウェアで実行するには大きな計算コストが必要とされる¹。ひとたびすべての単語候補が得られれば、それらの組み合わせの中から最適なものを選び出すことは高速に行なえるが、人間が文を読む場合にこれと同様の操作をしているとは考えにくい。

一方、人間が文を読み進めて行く際には先行単語によって後続する単語の範囲を予測し、それによって探索空間を狭めていると考えられる。これを裏付ける証拠の1つとして semantic garden path sentences の存在が挙げられる。たとえば

The astronomer married the star. (5.1)

という文を最初読んだときは、先行する astronomer という単語のために star の意味として「星」が想起され奇妙な感じを与える。その後ややあって to marry という動詞の目的語は人間でなければならないことから star は「人間の映画スター」であるという解釈に落ち着く [Charniak83]。

このように単語範囲を限定することが文の認識を困難にする場合もあるが、大抵の場合はこれ

¹文字列中に含まれるすべての単語を専用ハードウェアで検索する研究には [福島91] 等がある。

とは反対に探索空間を狭めることで認識時間が短縮でき、また認識精度が向上するものと期待できる。実際、筆者らは以前この手法を計算機による日本語解析に応用したシステムを開発している[高橋 88] [高橋 89a]。また[潤潟 90]は音声認識において先行単語と共起頻度の高い単語を優先的に検索した結果、認識精度が向上したと報告している。しかし、これらのシステムはいずれも逐次型計算機上での実行を念頭に置いており、先行単語と関連の深い単語の範囲を限定するための計算効率はあまりよいとは言えない。

一方、ニューラルネットを用いて同音異義語の解消を行なった研究には、[Waltz85]、[田村 87]、[木村 90]などがある。これらはいずれもすでに形態素解析が済んだ後の単語列を入力として受け取るようになっている。本章ではさらに一步進んで、同音異義語の解消のみならず形態素解析まで含めた解析を行うニューラルネットを構築する。このニューラルネットは与えられた文字列にマッチする単語を探す際、過去に読み込んだ単語と意味的なつながりの強いものほど優先的に出力するようになっている。またその場合の単語検索は並列に行われるので、検索時間が辞書の大きさに影響を受けないという特長を持つ。また、このネットワークモデルを逐次型の計算機上でシミュレートした結果について述べ、さらに並列計算機上でのシミュレーションについて考察する。

5.2 ニューラルネットによる文脈の表現

本節ではニューラルネットを用いて文脈を表現する方法について述べる。

文脈とは、ある特定の単語または概念の集合が、他よりも強く意識されている状態ととらえることができる。文を読み進むにつれて文脈は刻々と変化して行く。新たに単語が読まれると、その単語の表す概念だけでなく、その概念と関係の深い概念の注目度(意識されている度合)も上昇する。また一度上昇した注目度は時間がたつにつれて減少する。

今、ネットワーク内の各ノードに個々の単語または概念を割り合て、注目度をノードの出力値で表現することにすると、文脈はこれらのノードの出力値パターンとして表現できる。この場合、各ノードの出力値は

1. あるノードの出力値が上昇すると、そのノードと関係の深いノードの出力値も上昇する(活性の伝搬)
2. 一度上昇した出力値は時間とともに下降する(時間による減衰)

という性質を備えたものでなければならない。

第一の性質(活性の伝搬)を実現するには、関連あるノード同士を興奮性のリンクで結ばばよい。これは基本的に Quillian の memory model [Quillian67] や Hirst の marker passing [Hirst87] などと同じ手法である。同様のメカニズムは [Waltz85] [田村 87] [木村 90] などでも採用されている。本章のネットワークでは、単語間の共起頻度をもとにノード間のリンクの重みを決定することとする(詳しくは後述)。

一般に、ノードからノードへ活性の伝搬を行うシステムでは heat death (または false positive) と呼ばれる問題が生じ得る [Waltz85]。これは本来期待していなかった範囲にまで活性が伝搬されてしまうことをいう。本ニューラルネットに関していえば、ある概念が活性化されたとき、それとリンクで接続された他の概念が活性化され、それらがまた別の概念を活性化し…、といった動作が連続して起きた結果、ほとんどすべての概念が活性となってしまうことが heat death に当たる。Heat death を防ぐためにはノードの活性値を下げる方向に働くメカニズムがあればよい。本ニューラルネットでは、各ユニットの出力値を決定する関数の中に時間とともに減衰する項を付け加えることでこれを実現している。またこの機能は、上で述べた第二の性質(時間による減衰)をも同時に満足するものとなっている。さらに活性の伝搬形態に制限を加えることで heat death を防ぐ工夫をしている(後述)。

5.3 ネットワークの構成

本節では実験に用いたニューラルネットの構成について説明する。

外部から本ニューラルネットに与えられる情報は、

1. 解析すべき文字列
2. 直前の単語の品詞
3. それまでに認識された単語

の3種類である。これらの情報から文字列の先頭部分にマッチする単語が1つ選出され、その単語に対応する出力ユニットが活性化する。以下では本ニューラルネット内のユニット群とそれらの間のリンクについて説明する。

5.3.1 ユニット群とその役割分担

本ニューラルネット内のユニットは以下の5群に分かれている。

1. 文字ユニット群：解析すべき入力文字とその位置を示すユニットである(図5.1)。基本的には[Rumelhart86]の単語認識ネットワークで用いられたものと同様の構造を持つ。ただし、日本語は英語に比べると文字の種類が多いため、それぞれの文字を表すユニットを各位置毎に用意することは難しい。本ニューラルネットでは日本語1文字をまずJISコード2バイトで表し、1バイト目のコードと2バイト目のコードをそれぞれ独立した位置における文字とみなすことによってユニット数の削減を図っている。例として「はな」という入力文字列が与えられた場合を考える。「は」のJISコードは16進数で $24_{(16)}$, $4F_{(16)}$ の2バイト、同様に「な」の方は $24_{(16)}$, $4A_{(16)}$ の2バイトである。この場合、 $24/1$ (1番目の文字を表わすユニットのうち文字コードが $24_{(16)}$ であるもの。以下同様)、 $4F/2$, $24/3$, $4A/4$ の4個のユニットが値1を出力し、それ以外のユニットは値0を出力する。ただし、図が繁雑になるのを避けるために図5.1中ではそれぞれの日本語1文字を表すユニットが位置毎に用意されているように記してある。
2. 先行品詞ユニット群：現在解析しようとしている単語の直前の単語の品詞を表現するためのユニット群である(図5.1)。入力はニューラルネットの外部から与えられる。直前の単語の品詞を表わすユニットは値1を出力し、それ以外のユニットは値0を出力する。

3. 後続品詞ユニット群：現在解析しようとしている単語の品詞として許されているものを示すためのユニット群である(図5.1)。「後続」というのは先行品詞の直後に続くという意味であり、現在解析している語の後に続くという意味ではないので注意されたい。先行品詞ユニット群から値1の入力を受けたユニットのみが値1を出力ユニット群へ伝える。このユニット群を経由せず、先行品詞ユニット群から出力ユニット群へ直接リンクを張って出力ユニット群を制御することも可能であるが、一般にある1つの品詞に後続し得る品詞は複数存在するため、そのような方法ではリンクの数が不必要に多くなってしまう。ここに示したように、一度先行品詞から後続品詞を選択し、その後に各単語を制御するようにした方がリンクの数を少なく押えることができる。

4. 連想ユニット群：先行単語から後続単語を連想するためのユニット群であり、単語1個につきユニット1個が用意されている(図5.2)。システムによってある単語が認識されると、その単語を表すユニットに外部から入力を与えられる。この入力によってユニットは活性化するが、この活性値は時間とともに一定の時定数で減衰する。また、意味的に密接な関係にある単語を表すユニットの間は強い興奮性のリンクで結ばれており、活性値はこのリンクによって他のユニットへと伝搬する。これによって先行する単語と関連の深い単語を表すユニットの活性値が高くなる。

なお、連想ユニット群の内部における各ユニットの活性値は文章中のトピックの移り変わりを反映するだけであり、具体的な入力文字や直前の品詞等の文法的制限は受けない。

5. 出力ユニット群：連想ユニット群の各ユニットと一対一で対応している(図5.1, 5.2)。文字ユニット群、後続品詞ユニット群、および連想ユニット群から受ける入力をもとにして、入力文字列にマッチし、直前の単語と文法的に接続可能で、かつ現在までに確定された単語と最も意味的に関連の深い単語を表すユニット1個のみが活性化するように調整が行なわれる。

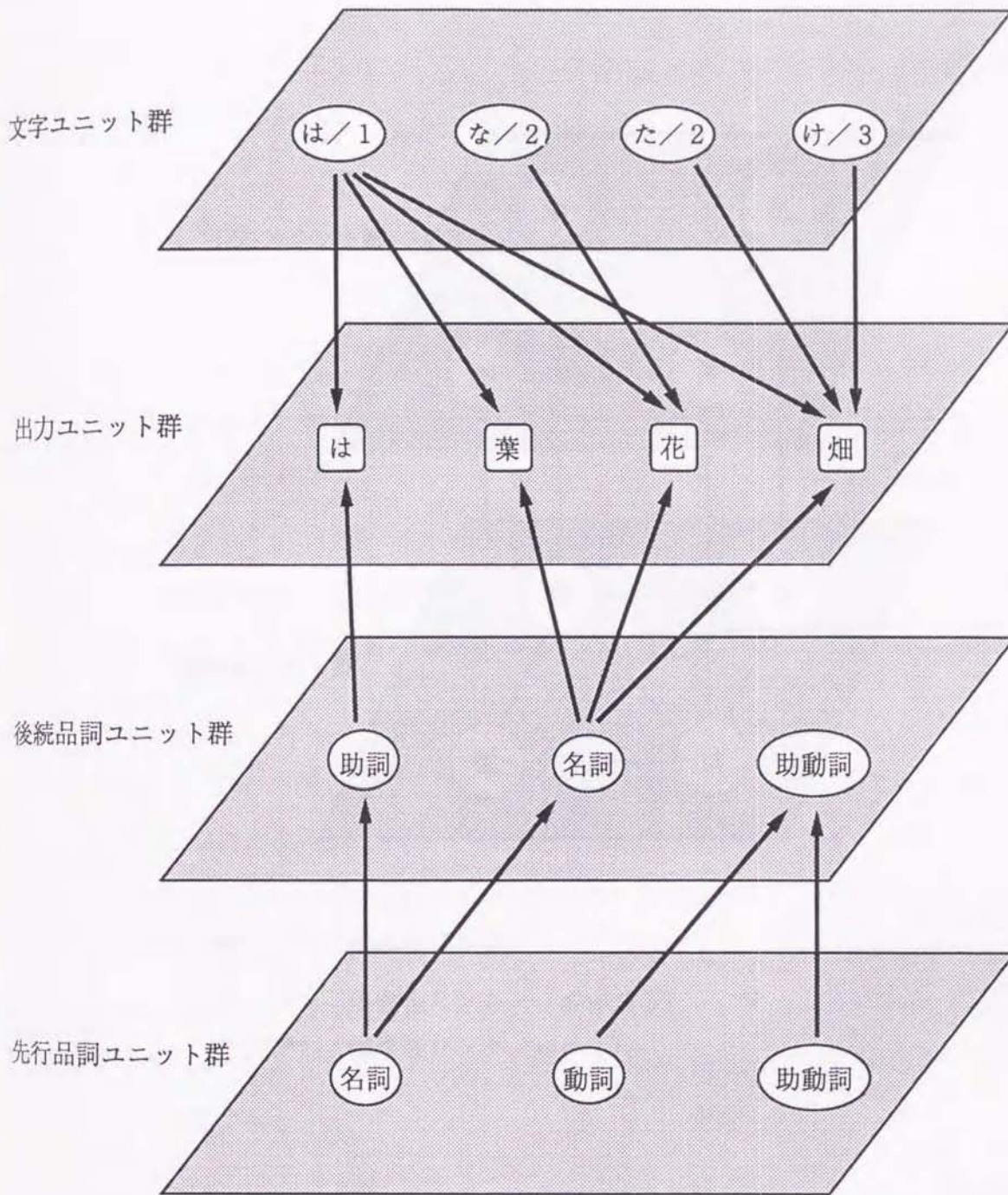


図 5.1: 各ユニット群 (連想ユニット群は除く) とそれらの間のリンク

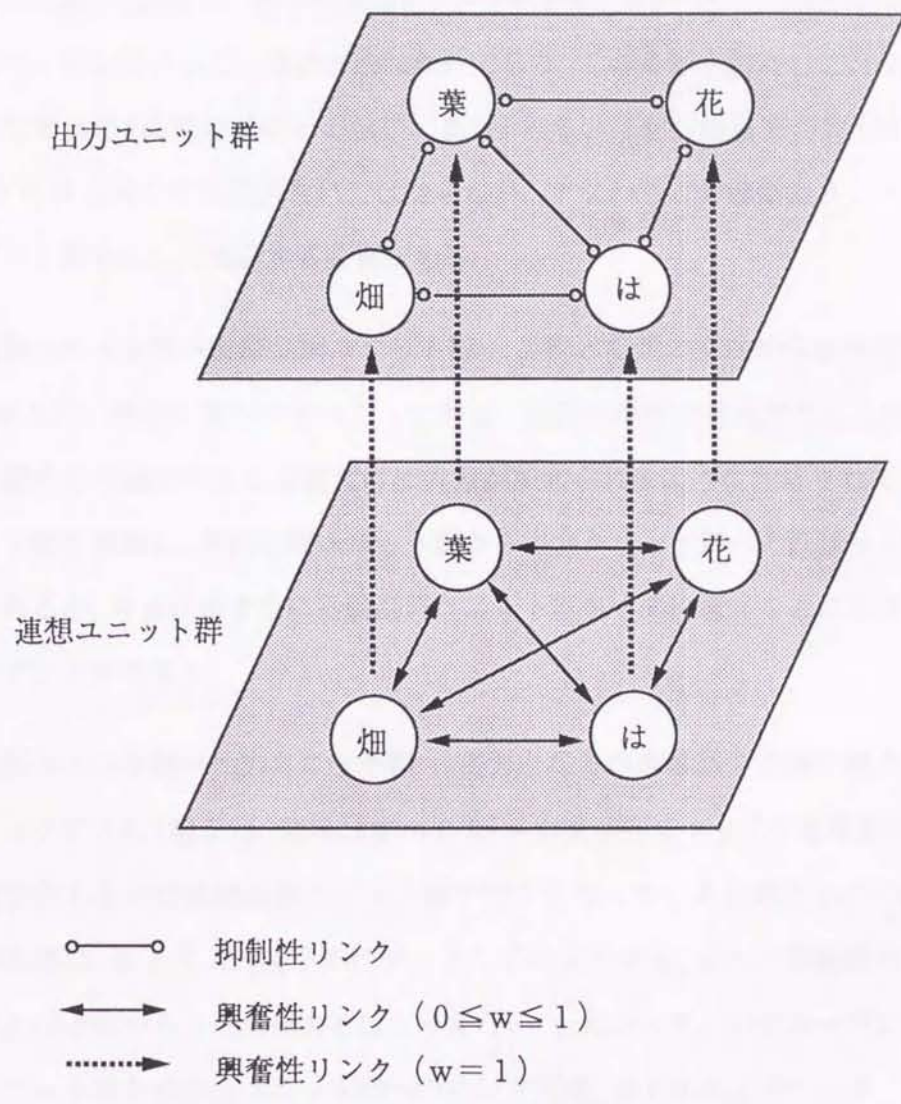


図 5.2: 連想ユニット群と出力ユニット群, およびそれらの間のリンク

5.3.2 ユニット群間リンクおよびユニット群内リンク

ネットワーク内には以下に示すようなリンクが存在する。

1. 文字ユニット群→出力ユニット群：入力文字列にマッチする単語のみを選択するように働くリンクである(図 5.1)。結合の強さ(リンクの重み)はすべて1。単語の文字表記が入力文字列にマッチしていることは絶対的な条件であり、この条件を満たしていない単語はたとえ文法的な妥当性(直前の単語との接続可能性)あるいは意味的な妥当性(以前の単語との関連)がどれほど高くても選択されてはならない。すなわちこの種類のリンクは出力ユニットへのゲート素子として機能する必要がある。
2. 先行品詞ユニット群→後続品詞ユニット群：品詞と品詞の文法的な接続可能性を表わすリンク(図 5.1)。結合の重みはすべて1である。直前の単語の品詞が与えられた時に、その直後に位置する可能性のある品詞だけに入力が増えられるように配線されている。後続品詞ユニット群を省略し、先行品詞ユニット群から出力ユニット群へと直接リンクを張ることも可能であるが、本方式のように後続品詞ユニット群を中間に置くことにより全体のリンク数を減らすことができる。
3. 後続品詞ユニット群→出力ユニット群：出力となるべき単語を品詞の観点から制限するためのリンクである(図 5.1)。重みはすべて1。入力文字列にマッチする単語のうち、選択される可能性があるのは後続品詞ユニット群中で1となっている品詞のもののみである。それ以外の単語は、たとえ入力文字列にマッチしていようとも、また以前解析された単語との関連がどれほど深かろうとも出力とはなりえない。したがってこのグループに属するリンクも文字ユニット群から出力ユニット群へのリンク同様、出力ユニットへのゲート素子として機能する必要がある。
4. 連想ユニット群内：先行単語から後続単語を連想するためリンクである。[木村 90]と同様、実際の解析に先だって学習用の例文を統計的に処理し、その結果によってリンクの重みを決定する。重みの範囲は $[0, 1]$ であり、意味的に密接な関係にある単語を表すユニットの間に張られているリンクほど大きな値をとるようになっている。連想ユニット群内のリンクの重み決定に用いた具体的な方法に関しては後述する。

一般的に言って、文の中で位置的に近い単語同士は意味的に密接な関係を持つことが多い。したがって、文を読み進む場合は最近に認識された単語ほど後続単語の予測に強い影響を与

えるものと考えられる。逆の見方をすれば、より遠い過去に認識された単語ほど後続単語への影響が小さいと言える。このように単語の影響力が次第に薄れてゆく現象は、ユニットの活性化レベルを時間と共に減衰させることで表現できる [木村 90] [田村 87]。したがって連想ユニット群内のユニットは、リンクを介して結合している他のユニットに興奮を伝えるだけでなく、自分自身の出力値も一定の時定数で減少するようなものでなければならない。

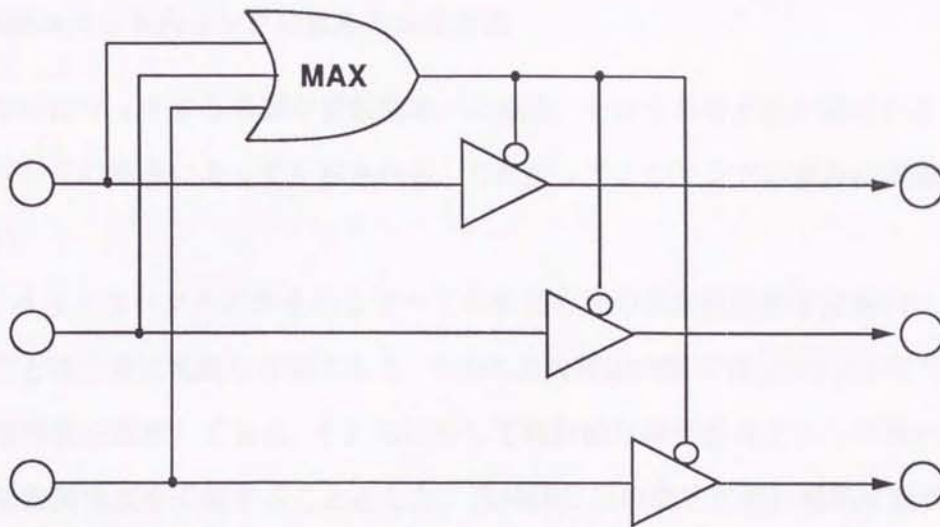
5. 連想ユニット群→出力ユニット群： 現在どの単語がどの程度注目されているかを示すためのリンクである。連想ユニット群内の各ユニットと、それに対応している出力ユニット群内の各ユニットを一对一で結合しており、連想ユニットの活性化レベルをそのまま出力ユニットに伝える働きをしている (図 5.1)。

一見するとこのリンクを省略して連想ユニットに出力ユニットの役割を兼ねさせることが可能であるように思われる。しかし、出力ユニット群の中では最終的にただ1つのユニットのみが値1を出力することを許され、それ以外のユニットは値0を出力しなければならない。このようにただ1つのユニットのみが1であるようなユニット群を連想に用いた場合、2つ以上前の単語の影響を考慮することができないという不都合が生じる。したがって一見冗長なようでも連想に用いるための連想ユニット群と、その中から1つだけを選択するための出力ユニット群の2つが必要となる。

6. 出力ユニット群内： 前述のように、出力ユニット群内では連想ユニットからの入力最大のもののみが1となり、それ以外のユニットは0となる。このような最大値検出機能をニューラルネットで実現するためにはすべてのユニット間を抑制性のリンクで接続すれば良い ([Waltz85] [田村 87]) が、これをそのまま実現すると単語の数が多くなるにしたがってリンクの数が爆発的に増大してしまう。実際には文字ユニット群からの入力がゲート素子的に働くため、同時に0以外の入力を受けるユニットは同一の部分文字列を共有する単語に限られる。よって抑制性リンクを張る必要があるのは同一の部分文字列を共有するユニット間のみでよい。より厳密に表現すると、単語Aの長さを n 、単語Bの長さを m 、また $n > m$ とするとき、単語Aの最初から m 番目までの部分文字列と単語Bの文字列全体が等しいときのみAを表すユニットとBを表すユニットとの間に抑制性リンクを張れば良い。図 5.2の出力ユニット群内で「花」と「畑」の間にリンクがないのはこのためである。

しかしこれでもなおリンクの数は膨大なものとなる。このネットワークモデルを計算機上に実装する場合は、まず連想ユニット群から出力ユニット群への入力の最大値を求め、次に

その最大値と各出力ユニットへの入力とを比較してそれらが等しい場合のみ1を出力する
 ような構成をとる方がより現実的であろう (図 5.3). 後述するシミュレーションの実験でも
 そのようになっている.



連想ユニット群

出力ユニット群

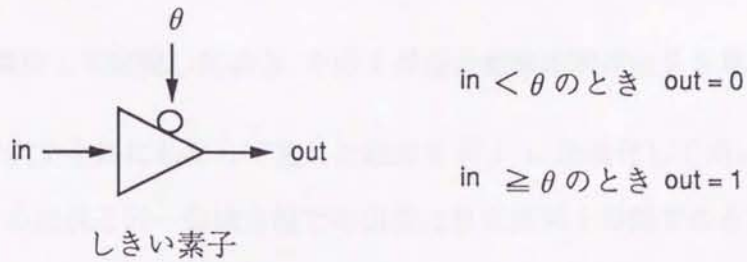


図 5.3: 最大値検出のための回路

5.4 いくつかのヒューリスティクス

本節では実際にネットワークを構成する際に用いたいくつかのヒューリスティクスについて述べる。これらはいずれも経験的に利用価値が高いと判断されたものであり、理論的な裏付けに基づいて導入されたわけではない。理論的な解析は今後の課題である。

5.4.1 連想ユニット内リンクの重みの決定方法

入力文字列にマッチする単語が複数個あった場合、そのうちのどれが選ばれるかは連想ユニット群内のリンクの重みによって左右される。したがってこのリンクの重みは慎重に決定しなければならない。

しかし、ネットワーク内に含まれるすべての単語の間の意味的関係を定量的に評価して数値で表現することは非常に困難な作業である。そのため今回の実験では[木村 90]などと同様にあらかじめ学習用例文を用意しておき、それらに対して統計的な操作をほどこして得た数値をもって単語間の意味的関係度を近似することとした。具体的には位置的に近い場所に現れることの多い2単語は意味的にも強く結び付いていることが多いと仮定し、以下の比で重みづけをした数値を用いた。

$$(\text{同一文中}) : (\text{句読点間}) : (\text{直接隣接}) = 1 : 2 : 5 \quad (5.2)$$

より正確には、ある特定の2単語が

1. 同一文中に出現した場合、その2単語の意味的関連度を1単位増加させる
2. 同一の句読点から句読点ではさまれた区間に出現した場合、その2単語の意味的関連度を2単位増加させる
3. 位置的に直接隣接して出現した場合、その2単語の意味的関連度を5単位増加させる

という操作を学習用例文全体にわたって施した結果を $[0,1]$ に正規化して用いた。最初の2項目、すなわち同一文中での出現と同一句読点間での出現は自立語同士の間でのみ計数した。最後の項目、すなわち直接隣接した出現に関しては自立語と付属語の両方を計数の対象とした。

上の比率には理論的裏付けがあるわけではなく、よい結果が得られるように実験を通じて求めたものに過ぎない。学習用例文、あるいは解析対象の文によって数値は変化する可能性がある。また、意味的関連性を文章上の位置的距離で近似しているため、望ましくない関連情報(ノイズ)の

混入は避けられない。ただし[木村90]では、学習用例文の数を増やすことによって実用上問題のない程度までノイズを低減させることが可能であると報告されていることから、統計的な出現頻度情報で意味的関連性を近似することはある程度まで可能であると考えられる。

5.4.2 活性伝達の範囲

ネットワーク全体を heat death から守るため、連想ユニット群内の活性伝搬に以下のような制限を加えることにした。

活性伝搬の制限：

連想ユニット群内で活性を伝搬させるのは、形態素として確定したユニット—すなわち出力ユニット群内で1を出力したユニットに対応する連想群内ユニット—と直接接続しているユニットのみである。それ以外のユニットはリンクを通じた活性伝搬を行わない。

例を図5.4および図5.5に示す。図5.4は入力された文字列にマッチする単語を決定する直前の状態である。連想ユニット群内に示した数値はこの時点でのそれぞれのユニットの活性値を表す。また出力ユニット群内の「葉」および「花」を指している矢印は、この2つのユニットのみが後続品詞ユニット群および文字ユニット群から値1の入力を受けていることを示す。

一方図5.5は、連想ユニット群・後続品詞ユニット群・文字ユニット群からの入力によって出力ユニット群の出力が決定された直後の状態である。出力ユニットが値1を出力するためには、後続品詞ユニット群と文字ユニット群からの入力が1であることが絶対的な条件である。この条件を満たしており、かつ連想ユニット群の中の対応するユニットの活性値が最も大きいもののみが1を出力し、それ以外のユニットは0を出力する(前述)。図5.4で後続品詞ユニット群および文字ユニット群から入力1を受けているユニットは「葉」と「花」だけであり、その中で対応する連想ユニット群内ユニットの活性値が最も大きいのは「花」である。したがって「花」だけが1を出力し、それ以外のユニットは0を出力することになる。つまり、現在与えられている文字列にマッチする単語として「花」が選ばれたことになる。

連想ユニット群内でこの次の時刻に活性が伝搬されるのは「花」とリンクで直接結ばれているユニットのみである。「は」「葉」「畑」の各ユニットとリンクで結ばれているユニットには活性は伝搬されない²。このように活性を伝搬されるユニットに制限を加えたのは、出力ユニット群

²もちろん「葉」と「花」の両方からのリンクを持つようなユニットには活性伝搬がなされる。

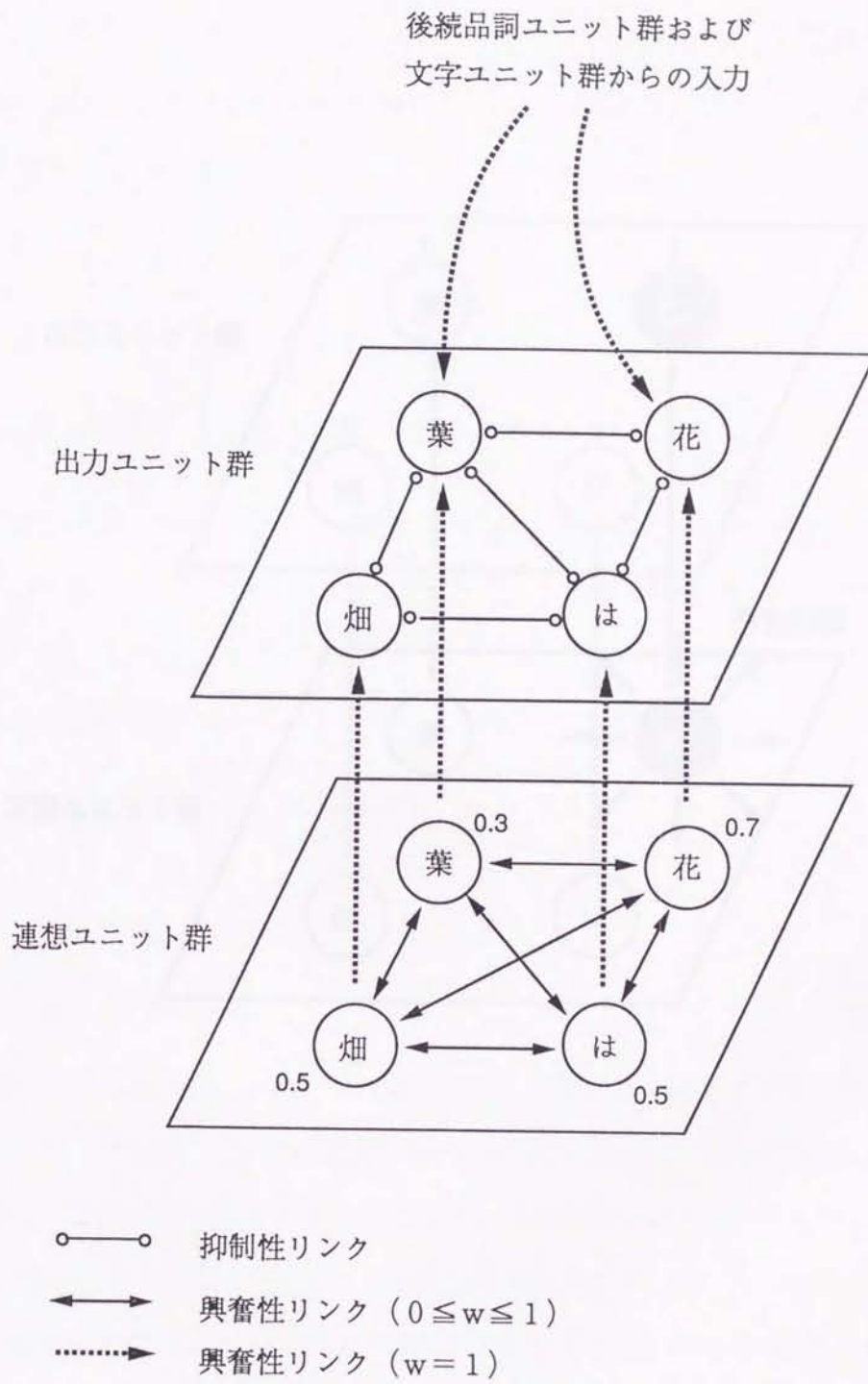


図 5.4: 活性伝搬の制限：出力ユニット決定前

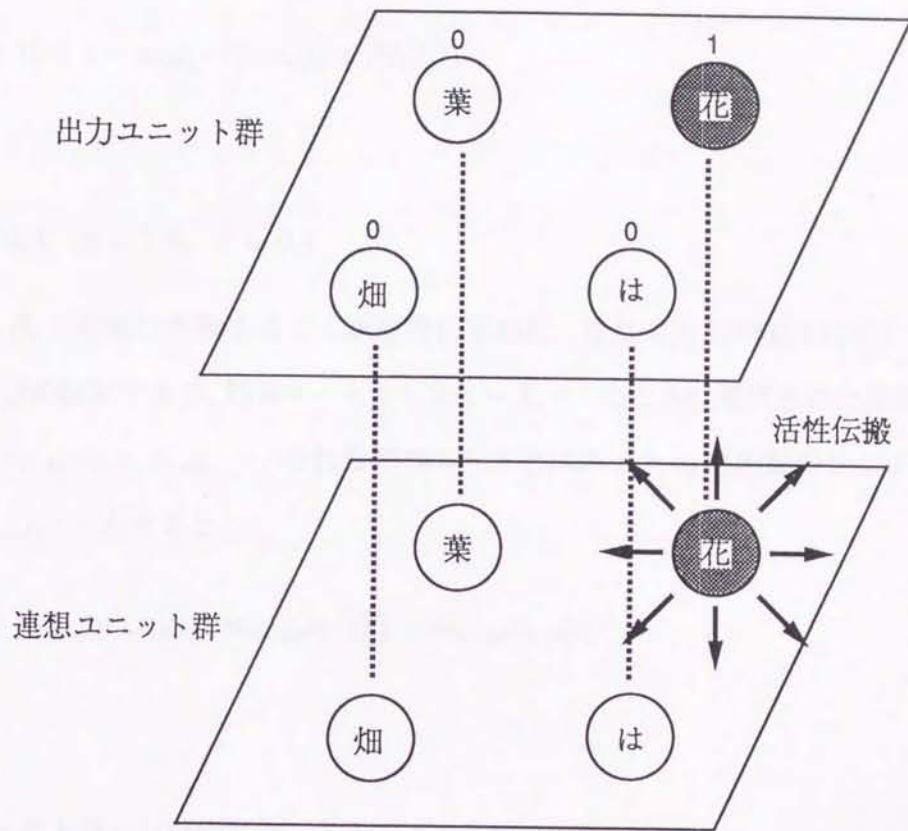


図 5.5: 活性伝搬の制限：出力ユニット決定後

内で選択されなかった単語, すなわち文の構成要素ではない単語, と意味的関連のある語の注目度を上げる必要はないと判断したからである.

5.4.3 ユニットの入出力関数

活性伝搬を上のように決定した結果, 連想ユニット u_i の時刻 $t+1$ における出力値は式 (5.3) で示したようなものとなる.

$$u_i[t+1] = 1 - \exp(-\tau(\alpha u_i[t] + \beta I_i[t])) \quad (5.3)$$

ただし

$$\alpha = 0.3, \quad \beta = 1.0, \quad \tau = 0.1 \quad (5.4)$$

である. α, β, τ の値は実験を通じて経験的に求めた. また $I_i[t]$ は時刻 t においてユニット u_i が受け取る入力総和であり, 時刻 $t-1, t-2, t-3, \dots$ のときに選択された単語を表すユニットをそれぞれ $v_{t-1}, v_{t-2}, v_{t-3}, \dots$, それらのユニットとユニット u_i との間のリンクの重みを $w_{v_{t-1}, i}, w_{v_{t-2}, i}, w_{v_{t-3}, i}, \dots$ とすると,

$$I_i[t] = w_{v_{t-1}, i} v_{t-1}[t] + w_{v_{t-2}, i} v_{t-2}[t] + w_{v_{t-3}, i} v_{t-3}[t] \dots \quad (5.5)$$

である.

5.4.4 バックトラック時の動作

本章で説明しているネットワークは, 与えられた入力文字列に最もよくマッチする単語を並列探索する. しかし, これはあくまでも単語の選択を並列に行うだけで, 解析方法自体は深さ優先の縦型探索である. 複数の構文解析木を同時に作成して保持しているわけではないから解析が袋小路に入ってしまった場合はバックトラックを行う必要があるが, 単なる縦型探索とは異なり形態素解析を進めていく各段階において最も尤度の高いと判断された単語を選択していくわけであるから, バックトラックの回数は小さく抑えられると期待できる.

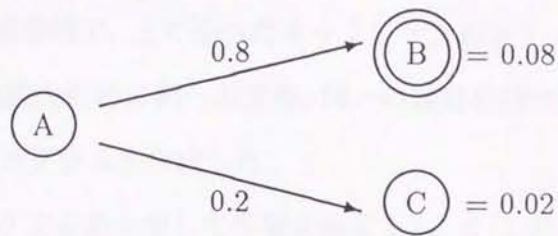
しかし次の単語が完全に予測できるのでない限り, バックトラックの必要が生じる可能性は0にはならない. Marcus の決定性統語解析パーザ [Marcus80] は先読みを行ってバックトラックを防ごうとするものであるが, 本章のネットワークはこれとは異なり, 「バックトラックは必要悪」との思想に基づいたものである.

実際にバックトラックの必要が生じた時は、前回誤って選択されてしまったユニット、すなわち連想ユニット群内で最も大きな値を持つユニット³の値を強制的に引き下げなければならない。単純にネットワークを誤った選択を行ったときの状態にもどしただけでは、また前回と同じユニットが選択されてしまうからである。

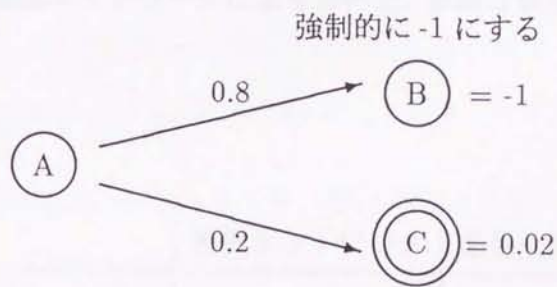
例えば、ユニット A とユニット B の間のリンクの重みが 0.8 で、ユニット A とユニット C の間のリンクの重みが 0.2 であるとする。今、ネットワークが単語 A (ユニット A で表される単語のこと。以下同じ。) を認識し、その次の単語の候補として B と C が見つかったとする (図 5.6a)。また A 以外のユニットの活性値はすべて 0 で、ユニット B およびユニット C は A 以外のユニットとはリンクで接続されていないものとする。このとき次の単語としては B が選択される。解析がさらに先まで進んだ結果不都合が生じ、単語 B を選択した時点までバックトラックが生じたとして、このときは前回選択された単語 B を表すユニットの値を強制的に -1 に設定する⁴。こうすることで次の候補 C が選択されるようになる。もし C を選択してもさらに不都合が生じてここまでバックトラックしてきた場合は、今選択されたユニット C の値をも強制的に -1 する。図 5.6 の例では B, C 以外には単語の候補がなく、しかもどちらのユニットを選択しても解析に失敗することがわかったので、その前の単語 A の選択に誤りがあったと考え、さらにもう一段階遡って解析をやり直す。

³より厳密には、「後続品詞ユニット群および文字ユニット群の両方から入力 1 を受けている出力ユニット」に対応している連想ユニット中で最も大きな値を持つもの」である。

⁴式 (5.3) からわかるように、通常の活性伝搬においてはユニットは正の値しかとらない。ユニット値が負であることは、そのユニットの表す単語の選択が既に一度失敗していることの標識になっている。



a. 最初の選択



b. バックトラック時の選択

図 5.6: バックトラック時の動作

5.5 実験および考察

本節では、上で述べたネットワークのシミュレーションプログラムを作成し、実際に日本語文を解析させてみた結果について述べる⁵。

解析のための入力文には高等学校で用いられている世界史の教科書 [村川 82] から、フランス革命に関する部分の記述を取り出して使用した。第 3 章での形態素解析の実験とは異なり、平仮名への変換はしていない。なぜなら、本章の目的は人間が実際に文脈を利用して探索空間を狭めているのと同様の手法を用いたネットワークを構築することであり、その有効性を確かめるためには人間が普通に読むのと同様の日本語文を解析させる必要があると考えたからである。

第 3 章、第 4 章と同一の計算機環境で、上で述べたネットワークのシミュレーションプログラムを作成した。またそれ以外に比較のために同一の文法、同一の辞書を持つ単純な最長一致法 (縦型探索) に基づいた形態素解析プログラムを作成した。

前述の解析用入力文のうち 50 文を取り出して学習用例文とし、式 (5.2) にしたがって連想ユニット群内のリンクの重みを決定した。また辞書としては [金田一 74] で最重要語・重要語とマークされている単語に、選ばれた入力文を解析するのに必要な固有名詞等を追加したもの (単語数 6572、見出し語数 13333) を用いた。

入力文 100 に対する本章の連想ネットワークによる解析と、単純な最長一致法とによる解析の結果を表 5.1 にまとめる。

	連想ネットワーク	最長一致法
バックトラックの回数	64	280
解析成功率	92%	21%

表 5.1: 連想ネットワークと最長一致法との比較

実験に用いた文の数が少ないので定量的な評価を下すには慎重になる必要があるが、連想ネットワークを用いたことでバックトラックの回数が 1/4 以下に抑えられたことがわかる。また単純

⁵本章のネットワークに関してはまだ十分な実験が行われていない。式 (5.2) の定数値にも、もっとよい選択肢がある可能性がある。

な最長一致法の解析の成功率が21%であるのに対し、連想ネットワークを用いた場合の成功率は92%と大きく向上した。

最長一致法における解析誤りの多くは、付属語が連続している部分を名詞の平仮名表記と認識したものであった。たとえば文末の「～した」を「～舌」と誤る場合などがこれに当たる。この種の誤りが多発したのは、

1. 入力に用いた原テキスト中では和語動詞が平仮名書きされることが多い。
2. 辞書中のすべての単語には漢字表記だけでなく平仮名表記が付与されている。
3. 文法規則として体言止め(文の最後が名詞で終わること)を許している。
4. 文法規則として名詞の連続を許している。

等が相互に関連し合った結果と考えられる。

最長一致法では失敗するが連想ネットワークを用いた場合には成功する例には以下のようなものがある。

1. 平仮名表記された同音異義語の区別：

単純な最長一致法では「起こる」と「怒る」のように平仮名で表記すると同じになる2語を区別することはできない。連想ネットワークを用いた場合は先行単語との関連度にしたがって解析が進むので、このような2語を区別することが可能である。

2. 選択すべき語よりも長い範囲で一致する文字列が他に存在する場合：

最長一致法はその名のとおり辞書の中にある単語で入力文字列と最も長く一致するものを選択する方法であるから、たとえば「特例として」という入力があった場合は「として」の部分を「都市で」のように解析してしまう可能性がある。一方連想ネットワークを用いた場合はたとえ他により長く入力文字列とマッチする単語があったとしても、正しい語を選択することが可能である。

以上述べた連想ネットワークの利点を十分に引き出すためには、ネットワーク内のリンクの重みおよび各ユニットの入出力関数が適切に設定されていなければならない。今回の実験で生じた解析の失敗例のいくつかは、これらの設定が最適でないことを暗示している。以下に解析結果のいくつかを示す。失敗例のほとんどは、学習用例文中に出てこなかった単語の組み合わせが入力された際に生じたものである。

成功例1. 「フランスでは、18世紀後半になると絶対主義の弊害がしだいにあらわれてきた。」

解析結果:

M- ふらんす - フランス
KAKU-DE- で
KAKARI-HA- は
TOUTEN-、
M- 1
M- 8
M- せいき - 世紀
M- こうはん - 後半
KAKU-NI- に
D5R- な - 成
D5R-U- る
SETUZOKU-TO- と
M- ぜったい - 絶対
M- しゅぎ - 主義
NO-NO- の
M- へいがい - 弊害
KAKU-GA- が
H- しだいに - 次第に
D1- あらわれ - 現れ
SETUZOKU-TE- て
DK-KI- き - 来
KAKO-TA- た
KUTEN-。

解析に成功した例である。選択された単語を文頭から順に1行ずつ表示してある。M- で始まる語は名詞を表し、KAKU- で始まる語は格助詞を表す。それ以外の品詞略号については付録6を参照のこと。

失敗例 1. 「また当時支配的であった啓蒙主義の思想は、不合理な社会を打破し、自由な生活を営みたいという気分をゆきわたらせた。」

解析結果:

.....
(前半省略)
.....
TOUTEN-、
M- じゅう - 自由
DANTEI-NA- な
M- せいかつ - 生活
KAKU-WO- を
D5M- いとな - 営
D5M-I- み
KIBOU-TAI- たい
SETUZOKU-TO- と
D5W- い - 言
D5W-U- う
M- きぶん - 気分
KAKU-WO- を
D5R- ゆきわた - 行き渡
D5R-A- ら
SIEKI-SE- せ
KAKO-TA- た
KUTEN-。

「営みたいという」の部分が「営みたいと言う」のように解析されてしまった。「～という」はよく用いられる言いまわしであるが、その前の部分には非常に広い範囲の句が現れ得るので、先行単語との関連だけからこれを正確に解析するのは難しい。「という」を1語として辞書に登録してしまうというのが現実的な解決方法であろう。

失敗例2. 「ルイ14世の晩年以來、フランスの国家財政は窮迫し、立て直しの試みもかいなく、ルイ16世のときになって財政政策はまったくゆきづまった。」

解析結果:

M- るい-ルイ

M- 1

M- 4

M- せい-世

NO-NO- の

M- ばんねん-晩年

H- いらい-以來

TOUTEN-

M- ぶんす-フランス

NO-NO- の

M- こっか-国家

M- ざいせい-財政

KAKARI-HA- は

M- きゅうはく-窮迫

DS-SI- し

TOUTEN-

M- たてなおし-立て直し

NO-NO- の

M- こころみ-試み

KAKARI-MO- も

M- かい-回

K-NAKU-なく-無く

TOUTEN-

.....

(後半省略)

.....

「甲斐なく」となるべき部分が「回なく」になってしまっている。これは学習用例文中に「甲斐」という語が出てこなかったために、「かいなく…」を解析する際に「甲斐」と「回」の尤度がどちらも0で等しくなり、たまたま辞書中のエントリで先に見つかった「回」が選択された結果である。

失敗例3. 「当時国民議会議を指導したミラボーやラファイエットなどは立憲君主主義者で、王政変革の意志はなく、そのことは革命の旗として制定された三色旗の白色がブルボン家をあらわすことに象徴された。」

解析結果:

- M-とうじ-当時
- M-こくみん-国民
- M-ぎかい-議会
- KAKU-WO-を
- M-しどう-指導
- DS-SI-し
- KAKO-TA-た
- M-みらぼー-ミラボー
- HUKU-YA-や
- M-らふあいえっと-ラファイエット
- HUKU-NADO-など
- KAKARI-HA-は
- M-りっけん-立憲
- M-くんしゅ-君主
- M-しゅぎしゃ-主義者
- KAKU-DE-で
- TOUTEN-、
-
- (後半省略)
-

文頭、副詞となるべき「当時」が名詞と解析されてしまっている(“M-”はその語が名詞であることを示す記号である)。実験で用いた文法は名詞-名詞接続を許しているなのでこの「当時」は名詞であっても副詞であっても問題を生じないが、この場合は意味的に考えて副詞とすべきであろう。ネットワークは文法的な矛盾が生じたときしかバックトラックを行なわないので、「国民」を読んで「当時-国民」という名詞連続が意味的に不適切であることがわかってそのまま解析を続ける。

5.6 議論

5.6.1 メモリ効率とバックトラック

一般的に言って、縦型探索 (depth first) は使用するメモリが少なく済むという利点があるが、探索に失敗した場合はバックトラックを行う必要がある。反対に横型探索 (breadth first) は複数の選択肢を並列的に処理するため多くのメモリを必要とするが、バックトラックが生じることはない。すでに述べたように、本章の連想ネットワークを用いた形態素解析では以前出現した単語と意味的に関連の強い語を優先的に探してバックトラックの回数を減らしている。また探索方法自体は縦型であるので使用するメモリは少なく済む。このように、縦型探索と連想ネットワークを組み合わせることによって、省メモリという利点を活かしたままバックトラックの回数を最少限に抑えることが可能となった。以上の探索方法をまとめると表 5.2 のようになる。

	横型探索	縦型探索	縦型 + 連想ネットワーク
使用メモリ	多い	少ない	少ない
バックトラック	不必要	多い	少ない

表 5.2: 横型探索, 縦型探索, 連想ネットワークの比較

5.6.2 並列計算機によるシミュレーション

次に本ネットワークの並列性について考察する。本ネットワーク内における活性伝搬の様子は図 5.7 のように図示される。それぞれのユニット群内の各ユニットは並列に状態遷移を行うことができる。またユニット群間の活性伝搬も各々のリンクを通じて並列に行える。したがって入力された文字列の先頭部分にマッチする単語のうち、それまでに出現した単語との意味的関連が最も深いものを常に一定時間内で検索できることになる。しかもこの検索時間は辞書の大きさに影響を受けない。

ただし以上の並列検索を実際実現するためには、ユニットとリンクをハードウェアで実装する必要がある。理想的には専用の並列日本語形態素解析マシンを製作すればよいが、ネットワークの有効性を調べる実験のために専用マシンを製作するのはコストの面で問題がある。今回の実

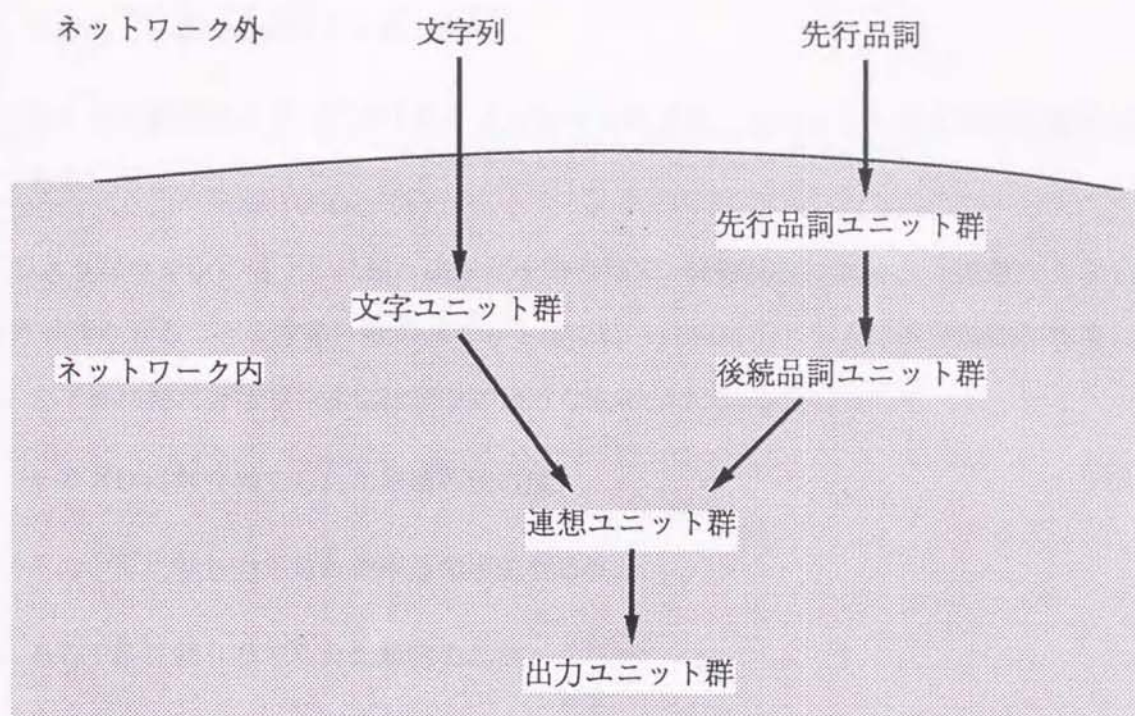


図 5.7: ネットワーク内における情報の流れ

験は通常の逐次型計算機(ワークステーション)上でシミュレーションを行ったが、各ユニットの状態遷移およびリンクを介しての活性伝搬をシミュレートするには多大な計算時間が必要とされる。

これらのことを考慮すると、本ネットワークは Connection Machine[Hillis85]のような並列計算機上でシミュレートするのに適していると考えられる⁶。その場合は、上のネットワークの構成をそのままシミュレートするのではなく、計算機の構造がうまく利用できるようなネットワークの構成を変更する必要があるだろう。以下では仮想的な並列計算機を想定し、その上で本章のネットワークを動作させることを考える⁷。想定する並列計算機は Connection Machine と同様、以下に示すような特性を備えているものとする。

⁶実際に並列計算機を用いて音声認識装置から入力される自然言語の統語解析を行う研究としては [Higuchi91a] [Higuchi91b] 等がある。

⁷以下で述べる並列計算機上でのシミュレーションは実際にはまだ実装されていない。

1. SIMD 型のマシンであり、個々の PE (processing element) がローカルなメモリを持つ。
2. 任意の PE 間での通信が可能である。
3. ホスト計算機からすべての PE にメッセージのブロードキャストおよび個別通信が可能である。

このようなマシンの各 PE に個々の単語を割り当て、後続品詞ユニットと連想ユニットの働きを合わせ持たせることとする。他のユニット群の行う仕事はホストとの通信機能を利用して処理する。各 PE に持たせるデータには以下のようなものがある：

1. その PE に割り当てられた単語の注目度。
2. その PE に割り当てられた単語の表記文字列。
3. その PE に割り当てられた単語の品詞。
4. 他の単語との意味的関連度。

1 はユニットの活性値 (出力値) に相当する。2 および 3 が意味するものは自明であろう。4 の “他の単語との意味的関連度” は、前述のネットワークでいえば 5.3.2 節の連想ユニット群内リンクに相当する情報である。PE 上のデータとして表現する場合は、

$$\langle j_1, w_{ij_1} \rangle, \langle j_2, w_{ij_2} \rangle, \dots, \langle j_n, w_{ij_n} \rangle \quad (5.6)$$

というリスト形式を用いるのが便利であろう。ただし、 i はこのデータが置かれる PE の PE 番号、 j_1, \dots, j_n はこのユニットが表す単語との間が 0 より大きな重みのリンクで結ばれている単語⁸を表している PE の PE 番号、 $w_{ij_1}, \dots, w_{ij_n}$ はそれぞれ i と j_1, \dots, j_n との間のリンクの重みである。

各 PE に以上のようなデータを持たせたとき、ホスト計算機と PE の基本動作サイクルは図 5.8 のようになる⁹。where 文は if 文と似ているが、括弧内の条件が満たされた PE 上でのみ then 以下が実行されることを表す命令とする。

⁸0 よりも大きな重みのリンクで結ばれている単語のみをリストで表現する代わりに、すべての単語とのリンクの重みをベクトルで表現することも可能である。ただし、一般に 1 つの単語が意味的関連を持つ単語の集合は辞書中の全単語に比べてかなり小さい範囲に留まる [田中・吉田 88] ので、そのようなベクトルの大部分は 0 で占められることになる。PE 数の多い並列計算機では各々の PE にそれほど多くのメモリを割り当てることができないので、本文にあるようなむだな情報を省いたリスト表現の方が有利であろう。

⁹繁雑さを避けるため、バックトラック時の動作に関する記述は省略する。

図5.8の中ではステップ9の最大値検出に必要なとされる計算時間が気になるが、これに関しては Connection Machine と同様、各 PE から返される値の最大値を検出する専用命令を備えており効率よく実行できるものとする。

一般にこの種の並列計算機では PE がハイパーキューブ状に並んでおり、上下左右に隣接した PE との間の通信は高速に実行できるが、ランダムな 2PE 間の通信はそれに比べてかなりコストがかかる。このため上の動作サイクルではステップ12の活性伝搬のフェーズがボトルネックになる可能性がある。しかしすでに5.4節で述べたように、本章の連想ネットワークでは実際に単語として認定されたものだけが周囲に活性伝搬を行う。よって [木村 90] のネットワークなどとは比べて PE 間の通信は少なく、計算機の構造に由来する通信時間の遅れは小さく抑えることが可能であると思われる。

1. [PE] 他の PE からの入力と現在の自分自身の活性値から新しい活性値を算出する;
2. [HOST] 入力文字列を全 PE にブロードキャストする;
3. [HOST] 先行単語品詞から後続単語品詞として許されるもののリストを作成し、
各々をブロードキャストする;
4. [PE] where (ホストからブロードキャストされた入力文字列が
自分の持つ単語の表記文字列と等しい
5. and
6. ホストからブロードキャストされた後続品詞列のどれかが
自分の持つ単語の品詞と等しい)
7. then ホストに自分の出力値を返す;
8. else ホストに 0 を返す;
9. [HOST] 各 PE の中から最大の値を返したものを選出する;
10. [HOST] 10 で選出された PE にその PE が選択されたことを知らせる;
11. [PE] where (ホストから選択通知があった)
12. then 他の PE に、その PE の表す単語との意味的関連度に応じて
活性を伝搬させる;
14. 1 に戻る;

図 5.8: 並列計算機上での検索アルゴリズム

第 6 章

結言

本論文で行った研究は、

1. 相互結合型ニューラルネットを用いた日本語形態素解析
2. 相互結合型ニューラルネットを用いた日本語係り受け解析
3. 文脈情報を利用した形態素解析

の3つに大きく分けることができる。本章ではまず最初に上記のそれぞれについて個別にまとめる。その後で全体を通したまとめを行い、最後に今後の展望を述べる。

第3章では相互結合型ニューラルネットを用いて日本語の形態素解析を試みた。ネットワークのユニットには入力文字列の構成要素となり得る単語を割り当てた。またエネルギー関数は、自立語数を可能な限り少なくするという条件と、隣接する単語同士の文法的接続可能性とをもとにして作成した。約200の平仮名べた書き文を解析させたところ、その中の95%の文が期待通りに解析された。実験を行った範囲においては、形態素解析ネットワーク収束するまでのステップ数は入力文字列の長さに無関係であるように見えた。

第4章では相互結合型ニューラルネットを用いて日本語の係り受け解析を試みた。システムは前処理の済んだ文節列を入力とし、それらの間で可能な係り受け関係をユニットに割り当てる。ネットワークのエネルギー関数は、文法的な制約、係り側の文節に含まれる語と受け側の文節に含まれる語の意味的關係、係り側の文節と受け側の文節の距離、の3つをもとにして作成した。語と語の意味的關係は約500の学習用例文から抽出した単語間係り受け共起頻度を正規化したもので近似した。学習用例文と同一分野の約200文を文節列に変換してこのネットワークに係り受け解

析を行わせたところ、そのうちの約95%の係り受けを正しく解析することができた。入力される文節数が増加しても、係り受け解析ネットワークが収束するまでのステップ数はさほど増加しなかった。

第5章では、第3章とは異なるアプローチで形態素解析を試みた。ここでは、入力文字列の先頭部分にマッチする単語を検索するときに文脈に沿った単語を優先的に選択するようにした。具体的には、単語をノードとし単語間の意味的関連度をノード間リンクの重みとするようなネットワークを用い、それ以前に現れた単語と意味的に関連の深い単語を並列検索するメカニズムを使用した。また直前の単語とこれから選択しようとする単語の文法的接続可能性の判定にもネットワークを用い、意味的関連性の判定と同時に実行できるようにした。単語間の意味的関連性はテキスト中における同時出現頻度をもとに推定した。

このネットワークを用いて約100文の形態素解析を実行したときに生じたバックトラックの回数は64回だった。この回数は同一の文を単純な最長一致法で形態素解析した場合に生じたバックトラックの回数の1/4以下であった。また形態素解析の成功率は92%であった。

この方式では単語候補をすべて並列に探索する。したがって並列計算機上で実行した場合は辞書中の語彙項目が増加しても探索に必要な時間は一定のままである。

ニューラルネットを利用した自然言語解析システムは、従来のシステムに比べてハードウェア化が容易であると思われる。また上に述べたように、入力文の長さが長くなっても、解析に要する時間はそれほど増加しない。従ってある特定の分野における比較的複雑な文を専門に解析する自然言語解析エンジン等としての利用が考えられる。分野を限定すれば単語間の意味的関係を表す関数も比較的容易に作成できよう。

今後の課題としては、解析率の向上、形態素解析と係り受け解析の両方を実行する統合されたネットワークの作成等が残されている。また、エネルギー関数の定数決定に Boltzmann Machine の学習アルゴリズムを適用することの可能性についても考察を進める必要がある。

付録 A 単語辞書

第3章および第5章の形態素解析に用いた辞書の一部を以下に示す。自立語は辞書項目の一部のみを、付属語はそのすべてを示す。助動詞はすべての活用型を見出し語として登録してある。

(pos XXX spell YYY)

という形は、品詞が XXX で文字表記が YYY という語を表す。

;;; 一段動詞

(POS D1 SPELL ("あき" "飽き"))
(POS D1 SPELL ("あきらめ" "諦め"))
(POS D1 SPELL ("あきれ" "呆れ"))
(POS D1 SPELL ("あけ" "明け"))
(POS D1 SPELL ("あげ" "上げ"))
(POS D1 SPELL ("あこがれ" "憧れ"))
(POS D1 SPELL ("あずけ" "預け"))
(POS D1 SPELL ("あせ" "褪せ"))
(POS D1 SPELL ("あたえ" "与え"))
(POS D1 SPELL ("あたため" "暖め"))

;;; 五段動詞

(POS D5B SPELL ("あそ" "遊"))
(POS D5B SPELL ("うか" "浮か" "浮"))
(POS D5B SPELL ("えら" "選"))
(POS D5B SPELL ("およ" "及"))
(POS D5B SPELL ("ころ" "転"))
(POS D5B SPELL ("さけ" "叫"))
(POS D5B SPELL ("しの" "憫"))
(POS D5B SPELL ("しの" "忍"))

(POS D5B SPELL ("たっと" "尊"))

(POS D5B SPELL ("と" "飛"))

(POS D5G SPELL ("あお" "仰"))

(POS D5G SPELL ("あお" "扇"))

(POS D5G SPELL ("いそ" "急"))

(POS D5G SPELL ("うすら" "薄ら"))

(POS D5G SPELL ("およ" "泳"))

(POS D5G SPELL ("か" "嗅"))

(POS D5G SPELL ("かせ" "稼"))

(POS D5G SPELL ("かつ" "担"))

(POS D5G SPELL ("くつろ" "寛"))

(POS D5G SPELL ("こ" "漕"))

(POS D5K SPELL ("うご" "動"))

(POS D5K SPELL ("お" "置"))

(POS D5K SPELL ("か" "書"))

(POS D5K SPELL ("か" "搔"))

(POS D5K SPELL ("かがや" "輝"))

(POS D5K SPELL ("かたづ" "片付"))

(POS D5K SPELL ("かたむ" "傾"))

(POS D5K SPELL ("かわ" "乾"))

(POS D5K SPELL ("き" "聞"))

(POS D5K SPELL ("き" "利"))

(POS D5M SPELL ("あ" "編"))

(POS D5M SPELL ("あやし" "怪し"))

(POS D5M SPELL ("いとな" "営"))

(POS D5M SPELL ("いりく" "入り組" "入組"))

(POS D5M SPELL ("う" "生"))

(POS D5M SPELL ("うちこ" "打ち込" "打込"))

(POS D5M SPELL ("うら" "恨"))

(POS D5M SPELL ("うらや" "羨"))

(POS D5M SPELL ("うりこ" "売り込" "売込"))

(POS D5M SPELL ("おが" "拝"))

(POS D5N SPELL ("し" "死"))

(POS D5R SPELL ("あ"))

(POS D5R SPELL ("あお" "煽"))

(POS D5R SPELL ("あが" "上"))

(POS D5R SPELL ("あが" "上が" "上"))

(POS D5R SPELL ("あが" "揚が"))

(POS D5R SPELL ("あずか" "預か"))

(POS D5R SPELL ("あせ" "焦"))

(POS D5R SPELL ("あた" "当た" "当"))

(POS D5R SPELL ("あたたま" "暖ま"))

(POS D5R SPELL ("あつま" "集ま" "集"))

(POS D5S SPELL ("あい" "愛"))

(POS D5S SPELL ("あか" "明か" "明"))

(POS D5S SPELL ("あけはな" "明け放" "明放"))

(POS D5S SPELL ("あま" "余"))

(POS D5S SPELL ("あまやか" "甘やか"))

(POS D5S SPELL ("あら" "荒ら" "荒"))

(POS D5S SPELL ("あらわ" "現わ" "現"))

(POS D5S SPELL ("あらわ" "現わ"))

(POS D5S SPELL ("あらわ" "表"))

(POS D5S SPELL ("いいあらわ" "言い表わ" "言い表" "言表わ" "言表"))

(POS D5T SPELL ("う" "打"))

(POS D5T SPELL ("うけも" "受け持" "受持"))

(POS D5T SPELL ("か" "勝"))

(POS D5T SPELL ("さきだ" "先立"))

(POS D5T SPELL ("そだ" "育"))

(POS D5T SPELL ("た" "断"))

(POS D5T SPELL ("た" "立"))

(POS D5T SPELL ("たも" "保"))

(POS D5T SPELL ("なりた" "成り立" "成立"))

(POS D5T SPELL ("ま" "待"))

(POS D5W SPELL ("あ" "会"))

(POS D5W SPELL ("あ " "合"))
(POS D5W SPELL ("あしら"))
(POS D5W SPELL ("あじわ " "味わ " "味"))
(POS D5W SPELL ("あつか " "扱"))
(POS D5W SPELL ("あら " "洗"))
(POS D5W SPELL ("あらそ " "争"))
(POS D5W SPELL ("い " "云"))
(POS D5W SPELL ("い " "言"))
(POS D5W SPELL ("いわ " "祝"))

;;; カ変動詞

(SPELL ("き " "来") POS DK-KI))
(SPELL ("こ " "来") POS DK-KO))
(SPELL ("くれ " "来れ") POS DK-KURE))
(SPELL ("くる " "来る") POS DK-KURU))

; サ変動詞

(SPELL ("さ ") POS SAHEN-SA))
(SPELL ("せ ") POS SAHEN-SE))
(SPELL ("し ") POS SAHEN-SI))
(SPELL ("すれ ") POS SAHEN-SURE))
(SPELL ("する ") POS SAHEN-SURU))

;;; 形容詞

(POS K SPELL ("あお " "青"))
(POS K SPELL ("あか " "赤"))
(POS K SPELL ("あかる " "明かる " "明る"))
(POS K SPELL ("あさ " "浅"))
(POS K SPELL ("あさまし " "浅まし"))
(POS K SPELL ("あじけな " "味気無"))
(POS K SPELL ("あたたか " "暖 " "暖か"))
(POS K SPELL ("あたらし " "新し"))
(POS K SPELL ("あつ " "厚"))
(POS K SPELL ("あつ " "暑"))

(SPELL ("ない " "無い") POS K-NAI))

(SPELL ("なから" "無から") POS K-NAKARA))

(SPELL ("なかろ" "無かろ") POS K-NAKARO))

(SPELL ("なかる" "無かる") POS K-NAKARU))

(SPELL ("なかっ" "無かっ") POS K-NAKATU))

(SPELL ("なけれ" "無けれ") POS K-NAKERE))

(SPELL ("なく" "無く") POS K-NAKU))

(SPELL ("なさ" "無さ") POS K-NASA))

(SPELL ("おなじ" "同じ") POS K-ONAZI))

(SPELL ("おなじから" "同じから") POS K-ONAZIKARA))

;;; 副詞

(POS H SPELL ("ああ"))

(POS H SPELL ("あいかわらず" "相変わらず" "相変らず"))

(POS H SPELL ("あえて" "敢えて" "敢て"))

(POS H SPELL ("あくまで" "飽く迄" "飽迄"))

(POS H SPELL ("あらたに" "新たに"))

(POS H SPELL ("あらためて" "改めて"))

(POS H SPELL ("あるいは" "或は"))

(POS H SPELL ("あんまり"))

(POS H SPELL ("いかにも"))

(POS H SPELL ("いきなり"))

;;; 連体詞

(POS R SPELL ("あかすの" "開かすの"))

(POS R SPELL ("あくる" "明くる"))

(POS R SPELL ("あけやらぬ" "明けやらぬ"))

(POS R SPELL ("あたる" "当たる"))

(POS R SPELL ("あの" "あの"))

(POS R SPELL ("あらくれた" "荒くれた"))

(POS R SPELL ("あらぬ" "あらぬ"))

(POS R SPELL ("あらゆる" "あらゆる"))

(POS R SPELL ("ありし" "在りし"))

(POS R SPELL ("ありとあらゆる" "ありとあらゆる"))

;;; 接続詞

(POS S SPELL ("あるいは"))
(POS S SPELL ("いか" "以下"))
(POS S SPELL ("いっぽう" "一方"))
(POS S SPELL ("いらっしゃい"))
(POS S SPELL ("おいて" "於て" "於いて"))
(POS S SPELL ("おける" "於ける"))
(POS S SPELL ("および" "及" "及び"))
(POS S SPELL ("および" "及び"))
(POS S SPELL ("かかわらず"))
(POS S SPELL ("かぎり" "限" "限り"))

;;; 句読点

(SPELL ("。") POS KUTEN))
(SPELL ("、") POS TOUTEN))

;;; 助詞

; 格助詞

(SPELL ("で") POS KAKU-DE))
(SPELL ("が") POS KAKU-GA))
(SPELL ("へ") POS KAKU-HE))
(SPELL ("から") POS KAKU-KARA))
(SPELL ("に") POS KAKU-NI))
(SPELL ("と") POS KAKU-TO))
(SPELL ("を") POS KAKU-WO))
(SPELL ("より") POS KAKU-YORI))

; 「の」

(SPELL ("の") POS NO-NO))

; 副助詞

(SPELL ("ばかり") POS HUKU-BAKARI))
(SPELL ("だけ") POS HUKU-DAKE))
(SPELL ("のみ") POS HUKU-DAKE))
(SPELL ("ほど") POS HUKU-HODO))

(SPELL ("ころ ") POS HUKU-KORO))
(SPELL ("ごろ ") POS HUKU-KORO))
(SPELL ("頃 ") POS HUKU-KORO))
(SPELL ("くらい ") POS HUKU-KURAI))
(SPELL ("ぐらい ") POS HUKU-KURAI))
(SPELL ("まで ") POS HUKU-MADE))
(SPELL ("なり ") POS HUKU-NA))
(SPELL ("など ") POS HUKU-NADO))
(SPELL ("等 ") POS HUKU-NADO))
(SPELL ("等々 ") POS HUKU-NADO))
(SPELL ("ながら ") POS HUKU-NAGARA))
(SPELL ("すら ") POS HUKU-SURA))
(SPELL ("・ ") POS HUKU-TEN))
(SPELL ("とか ") POS HUKU-TOKA))
(SPELL ("つつ ") POS HUKU-TUTU))
(SPELL ("や ") POS HUKU-YA))
(SPELL ("やら ") POS HUKU-YA))

; 係助詞

(SPELL ("どころか ") POS KAKARI-DOKOROKA))
(SPELL ("は ") POS KAKARI-HA))
(SPELL ("きり ") POS KAKARI-KIRI))
(SPELL ("こそ ") POS KAKARI-KOSO))
(SPELL ("も ") POS KAKARI-MO))
(SPELL ("さえ ") POS KAKARI-SAE))
(SPELL ("しか ") POS KAKARI-SIKA))
(SPELL ("とて ") POS KAKARI-TOTE))

; 接続助詞

(SPELL ("から ") POS SETUZOKU-1))
(SPELL ("が ") POS SETUZOKU-1))
(SPELL ("けれど ") POS SETUZOKU-1))
(SPELL ("けれども ") POS SETUZOKU-1))
(SPELL ("し ") POS SETUZOKU-1))
(SPELL ("ので ") POS SETUZOKU-2))
(SPELL ("のに ") POS SETUZOKU-2))

(SPELL ("ば ") POS SETUZOKU-BA))
(SPELL ("で ") POS SETUZOKU-DE))
(SPELL ("ものなら ") POS SETUZOKU-MONONARA))
(SPELL ("ものの ") POS SETUZOKU-MONONO))
(SPELL ("て ") POS SETUZOKU-TE))
(SPELL ("と ") POS SETUZOKU-TO))
(SPELL ("とも ") POS SETUZOKU-TOMO))
(SPELL ("とて ") POS SETUZOKU-TOTE))

; 終助詞

(SPELL ("い ") POS SYUU-I))
(SPELL ("か ") POS SYUU-KA))
(SPELL ("ものか ") POS SYUU-MONOKA))
(SPELL ("な ") POS SYUU-NA))
(SPELL ("ろ ") POS SYUU-RO))
(SPELL ("よ ") POS SYUU-YO))

; 間投助詞

(SPELL ("な ") POS KANTOU-NA))
(SPELL ("ね ") POS KANTOU-NE))
(SPELL ("よ ") POS KANTOU-YO))

::: 助動詞

; 断定

(SPELL ("あら ") POS DANTEI-ARA))
(SPELL ("あれ ") POS DANTEI-ARE))
(SPELL ("あり ") POS DANTEI-ARI))
(SPELL ("あろ ") POS DANTEI-ARO))
(SPELL ("ある ") POS DANTEI-ARU))
(SPELL ("あっ ") POS DANTEI-ATU))
(SPELL ("だ ") POS DANTEI-DA))
(SPELL ("だろ ") POS DANTEI-DARO))
(SPELL ("だっ ") POS DANTEI-DATU))
(SPELL ("で ") POS DANTEI-DE))
(SPELL ("でし ") POS DANTEI-DESI))

(SPELL ("です") POS DANTEI-DESU))
(SPELL ("でしょ") POS DANTEI-DESYO))
(SPELL ("な") POS DANTEI-NA))
(SPELL ("なら") POS DANTEI-NARA))

; 丁寧

(SPELL ("ませ") POS TEINEI-MASE))
(SPELL ("まし") POS TEINEI-MASI))
(SPELL ("ます") POS TEINEI-MASU))
(SPELL ("ますれ") POS TEINEI-MASURE))
(SPELL ("ましょ") POS TEINEI-MASYO))

; 否定

(SPELL ("ない") POS HITEI-NAI))
(SPELL ("なかる") POS HITEI-NAKARO))
(SPELL ("なかっ") POS HITEI-NAKATU))
(SPELL ("なけれ") POS HITEI-NAKERE))
(SPELL ("なく") POS HITEI-NAKU))
(SPELL ("ね") POS HITEI-NE))
(SPELL ("ぬ") POS HITEI-NU))
(SPELL ("ん") POS HITEI-NU))
(SPELL ("ず") POS HITEI-ZU))

; 推量

(SPELL ("べから") POS SUIRYOU-BEKARA))
(SPELL ("べき") POS SUIRYOU-BEKI))
(SPELL ("べく") POS SUIRYOU-BEKU))
(SPELL ("べし") POS SUIRYOU-BESI))
(SPELL ("まい") POS SUIRYOU-MAI))
(SPELL ("らしかっ") POS SUIRYOU-MASIKATU))
(SPELL ("らしい") POS SUIRYOU-RASII))
(SPELL ("らしけれ") POS SUIRYOU-RASIKERE))
(SPELL ("らしく") POS SUIRYOU-RASIKU))

; 過去

(SPELL ("だ") POS KAKO-DA))

(SPELL ("だら ") POS KAKO-DARA))

(SPELL ("だり ") POS KAKO-DARI))

(SPELL ("だろ ") POS KAKO-DARO))

(SPELL ("た ") POS KAKO-TA))

(SPELL ("たら ") POS KAKO-TARA))

(SPELL ("たり ") POS KAKO-TARI))

(SPELL ("たろ ") POS KAKO-TARO))

; 未来

(SPELL ("う ") POS MIRAI-U))

(SPELL ("よう ") POS MIRAI-YOU))

; 伝聞

(SPELL ("そう ") POS DENBUN-SOU))

; 様態

(SPELL ("そう ") POS YUTAI-SOU))

; 比況

(SPELL ("よう ") POS HIKYOU-YOU))

; 推定

(SPELL ("よう ") POS SUITEI-YOU))

; 希望

(SPELL ("た ") POS KIBOU-TA))

(SPELL ("たがら ") POS KIBOU-TAGARA))

(SPELL ("たがれ ") POS KIBOU-TAGARE))

(SPELL ("たがり ") POS KIBOU-TAGARI))

(SPELL ("たがろ ") POS KIBOU-TAGARO))

(SPELL ("たがる ") POS KIBOU-TAGARU))

(SPELL ("たがっ ") POS KIBOU-TAGATU))

(SPELL ("たい ") POS KIBOU-TAI))

(SPELL ("たかろ ") POS KIBOU-TAKARO))

(SPELL ("たかっ ") POS KIBOU-TAKATU))

(SPELL ("たけれ ") POS KIBOU-TAKERE))

(SPELL ("たく ") POS KIBOU-TAKU))

; 使役

(SPELL ("させ ") POS SIEKI-SASE))

(SPELL ("させれ ") POS SIEKI-SASERE))

(SPELL ("させる ") POS SIEKI-SASERU))

(SPELL ("せ ") POS SIEKI-SE))

(SPELL ("せれ ") POS SIEKI-SERE))

(SPELL ("せる ") POS SIEKI-SERU))

; 受身・その他

(SPELL ("られ ") POS UKEMI-RARE))

(SPELL ("られれ ") POS UKEMI-RARERE))

(SPELL ("られる ") POS UKEMI-RARERU))

(SPELL ("れ ") POS UKEMI-RE))

(SPELL ("れれ ") POS UKEMI-RERE))

(SPELL ("れる ") POS UKEMI-RERU))

::: 用言語尾

; 一段動詞語尾

(SPELL ("れ ") POS D1-RE))

(SPELL ("る ") POS D1-RU))

; 五段動詞語尾

(SPELL ("か ") POS D5K-A))

(SPELL ("き ") POS D5K-I))

(SPELL ("く ") POS D5K-U))

(SPELL ("け ") POS D5K-E))

(SPELL ("こ ") POS D5K-O))

(SPELL ("い ") POS D5K-ONBIN))

(SPELL ("さ ") POS D5S-A))

(SPELL ("し ") POS D5S-I))

(SPELL ("す ") POS D5S-U))

(SPELL ("せ ") POS D5S-E))

(SPELL ("そ ") POS D5S-0))

(SPELL ("た ") POS D5T-A))

(SPELL ("ち ") POS D5T-I))

(SPELL ("つ ") POS D5T-U))

(SPELL ("て ") POS D5T-E))

(SPELL ("と ") POS D5T-O))

(SPELL ("っ ") POS D5T-ONBIN))

(SPELL ("な ") POS D5N-A))

(SPELL ("に ") POS D5N-I))

(SPELL ("ぬ ") POS D5N-U))

(SPELL ("ね ") POS D5N-E))

(SPELL ("の ") POS D5N-O))

(SPELL ("ん ") POS D5N-ONBIN))

(SPELL ("ま ") POS D5M-A))

(SPELL ("み ") POS D5M-I))

(SPELL ("む ") POS D5M-U))

(SPELL ("め ") POS D5M-E))

(SPELL ("も ") POS D5M-O))

(SPELL ("ん ") POS D5M-ONBIN))

(SPELL ("ら ") POS D5R-A))

(SPELL ("り ") POS D5R-I))

(SPELL ("る ") POS D5R-U))

(SPELL ("れ ") POS D5R-E))

(SPELL ("ろ ") POS D5R-O))

(SPELL ("っ ") POS D5R-ONBIN))

(SPELL ("わ ") POS D5W-A))

(SPELL ("い ") POS D5W-I))

(SPELL ("う ") POS D5W-U))

(SPELL ("え ") POS D5W-E))

(SPELL ("お ") POS D5W-O))

(SPELL ("っ ") POS D5W-ONBIN))

(SPELL ("が") POS D5G-A))
(SPELL ("ぎ") POS D5G-I))
(SPELL ("ぐ") POS D5G-U))
(SPELL ("げ") POS D5G-E))
(SPELL ("ご") POS D5G-O))
(SPELL ("い") POS D5G-ONBIN))

(SPELL ("ば") POS D5B-A))
(SPELL ("び") POS D5B-I))
(SPELL ("ぶ") POS D5B-U))
(SPELL ("べ") POS D5B-E))
(SPELL ("ぼ") POS D5B-O))
(SPELL ("ん") POS D5B-ONBIN))

; 形容詞語尾

(SPELL ("い") POS K-I))
(SPELL ("から") POS K-KARA))
(SPELL ("かろ") POS K-KARO))
(SPELL ("かる") POS K-KARU))
(SPELL ("かっ") POS K-KATU))
(SPELL ("けれ") POS K-KERE))
(SPELL ("く") POS K-KU))

目錄

CONTENTS

1. Introduction

2. The History of the Project

3. The Organization of the Project

4. The Methodology

5. The Results

6. The Discussion

7. The Conclusion

8. The Acknowledgements

9. The References

10. The Appendix

11. The Glossary

12. The Index

13. The Bibliography

14. The List of Figures

15. The List of Tables

16. The List of Abbreviations

付録 B 品詞接続規則

形態素解析に用いた品詞間接続規則を以下に示す。この文法規則は [西村 78] 等を参考にして作成した。

(A B C D)

という形は、品詞 A の直後に品詞 B, C, D が接続し得ることを表す。

::: 助詞

; 格助詞

(kaku-ga kakari-dokoroka touten)

(kaku-wo huku-nado huku-bakari huku-hodo huku-kurai huku-dake huku-made
huku-sura kakari-sae kakari-dokoroka kakari-kiri kakari-sika
kakari-koso kakari-mo touten)

(kaku-ni huku-ya huku-nado huku-bakari huku-hodo huku-kurai huku-dake
huku-made huku-sura kakari-sae kakari-dokoroka kakari-kiri
kakari-sika kakari-koso kakari-mo kakari-ha touten)

(kaku-de no-no huku-ya huku-nado huku-bakari huku-hodo huku-kurai
huku-dake huku-made huku-sura kakari-sae kakari-dokoroka
kakari-kiri kakari-sika kakari-koso kakari-mo kakari-ha touten)

(kaku-to no-no huku-ya huku-nado huku-bakari huku-hodo huku-kurai
huku-dake huku-made huku-sura kakari-sae kakari-dokoroka
kakari-kiri kakari-sika kakari-koso kakari-mo kakari-ha touten)

(kaku-he kaku-to no-no huku-ya huku-nado huku-bakari huku-hodo
huku-kurai huku-dake huku-made huku-sura kakari-sae
kakari-dokoroka kakari-kiri kakari-sika kakari-koso kakari-mo
kakari-ha touten)

(kaku-kara kaku-ga no-no huku-ya huku-nado huku-bakari huku-hodo

huku-kurai huku-dake huku-made huku-sura kakari-sae
kakari-dokoroka kakari-kiri kakari-sika kakari-koso kakari-mo
kakari-ha touten)

(kaku-yori no-no huku-ya huku-nado huku-kurai huku-dake huku-made
huku-sura kakari-sae kakari-dokoroka kakari-kiri kakari-sika
kakari-koso kakari-mo kakari-ha touten)

; 「の」助詞

(no-no kaku-ga kaku-wo kaku-ni kaku-de kaku-to kaku-he kaku-kara
kaku-yori huku-ya huku-nado huku-bakari huku-hodo huku-kurai
huku-dake huku-made huku-sura kakari-sae kakari-dokoroka
kakari-kiri kakari-sika kakari-koso kakari-mo kakari-ha
kakari-ka hikyou-you touten)

; 副助詞

(huku-ya kaku-ga kaku-wo kaku-ni kaku-de kaku-to kaku-he kaku-kara
kaku-yori no-no hukumade huku-kurai huku-bakari huku-dake
huku-hodo huku-sura kakari-sae kakari-ka kakari-dokoroka
kakari-mo kakari-ha kakari-sika kakari-koso kakari-kiri touten)

(huku-nado kaku-ga kaku-wo kaku-ni kaku-de kaku-to kaku-he kaku-kara
kaku-yori no-no huku-made huku-kurai huku-bakari huku-sura
kakari-sae kakari-ka kakari-dokoroka kakari-mo kakari-ha
kakari-sika kakari-koso touten)

(huku-made kaku-ga kaku-wo kaku-ni kaku-de kaku-to kaku-he kaku-kara
kaku-yori no-no huku-ya huku-nado huku-kurai huku-hodo
huku-sura kakari-sae kakari-ka kakari-dokoroka kakari-mo
kakari-ha kakari-sika kakari-koso kakari-kiri touten)

(huku-kurai kaku-ga kaku-wo kaku-ni kaku-de kaku-to kaku-he kaku-kara
kaku-yori no-no huku-ya huku-nado huku-made huku-bakari
huku-h-do huku-sura kakari-sae kakari-ka kakari-dokoroka
kakari-mo kakari-ha kakari-sika kakari-koso kakari-kiri touten)

(huku-bakari kaku-ga kaku-wo kaku-ni kaku-de kaku-to kaku-he kaku-kara
kaku-yori no-no huku-nado huku-made huku-bakari huku-hodo
huku-sura kakari-sae kakari-ka kakari-dokoroka kakari-mo
kakari-ha kakari-sika touten)

(huku-dake kaku-ga kaku-wo kaku-ni kaku-de kaku-to kaku-he kaku-kara

no-no huku-kurai huku-sura kakari-sae kakari-ka kakari-dokoroka
kakari-mo kakari-ha kakarisika kakari-kiri touten)

(huku-hodo kaku-ga kaku-wo kaku-ni kaku-de kaku-to kaku-he kaku-kara
kaku-yori no-no huku-ya huku-nado touten)

(huku-sura kakari-dokoroka kakari-mo touten)

(huku-nagara kaku-ni no-no kakari-mo touten)

(huku-ten m)

(huku-koro kaku-ni kaku-kara kaku-yori huku-made touten)

; 係助詞

(kakari-kiri kakari-sika kakari-dokoroka syuu-ka touten)

(kakari-tote kakari-mo touten)

(kakari-sae kakari-mo kakari-dokoroka touten)

(kakari-koso kaku-ga kakari-ha touten)

(kakari-mo kaku-ga kakari-dokoroka dantei-arū dantei-aro dantei-ari
dantei-atu dantei-are k-nai k-nakara k-nakaro k-naku k-nakatu
k-nakaru k-nakere k-nasa touten)

(kakari-ha dantei-arū dantei-aro dantei-ari dantei-atu dantei-are
k-nai k-nakara k-nakaro k-naku k-nakatu k-nakaru k-nakere
k-nasa touten)

(kakari-sika kakari-dokoroka touten)

; 接続助詞

(setuzoku-te kakari-ha kakari-mo touten kuten)

(setuzoku-to touten)

(setuzoku-1 touten)

(setuzoku-2 touten)

::: 助動詞

; 断定

(dantei-da denbun-sou setuzoku-1 setuzoku-tote setuzoku-to kuten)

(dantei-daro mirai-u)

(dantei-datu kako-ta kako-tara kako-tari kako-taro)

(dantei-na suitei-you kaku-yori huku-toka huku-nado huku-bakari
huku-kurai huku-dake huku-hodo huku-made shyuu-monoka)

setuzoku-mononara setuzoku-monono setuzoku-2 touten)
 (dantei-nara setuzoku-ba touten)
 (dantei-de dantei-aru dantei-ara dantei-aro dantei-atu dantei-are
 k-nai k-nakara k-naku k-nakatu k-nararu k-nakere k-nakare
 touten)
 (dantei-desu syuu-ka syuu-monoka kuten)
 (dantei-desyo mirai-u)
 (dantei-desi kako-ta kako-tara kako-tari kako-taro setuzoku-te setuzoku-1
 setuzoku-2 kuten)
 (dantei-aru suiryuu-mai suiryuu-beki suiryuu-bekara buiryuu-beku denbun-sou
 hikyou-you suitei-you dantei-daro dantei-desyo suiryuu-rasii
 suiryuu-rasiku suiryuu-rasikatu suiryuu-rasikere syuu-ka
 syuu-monoka syuu-na setuzoku-1 setuzoku-2 setuzoku-tote
 setuzoku-to setuzoku-mononara setuzoku-monono m kuten)
 (dantei-ara hitei-nu hitei-zu hitei-ne)
 (dantei-aro mirai-u)
 (dantei-ari youtai-sou teinei-masu teinei-mase teinei-masyo teinei-masi
 kibou-tai kibou-takaro kibou-takatu kibou-takere kibou-ta
 huku-nagara touten)
 (dantei-atu kako-ta kako-tara kako-tari kako-taro)
 (dantei-are setuzoku-ba)

; 丁寧

(teinei-masu suiryuu-mai kaku-ni kaku-yori huku-toka huku-nado huku-ya
 huku-bakari huku-kurai huku-dake huku-hodo huku-made syuu-ka
 syuu-monoka setuzoku-1 setuzoku-2 setuzoku-tote setuzoku-to
 setuzoku-mononara setuzoku-monono m kuten)
 (teinei-mase hitei-nu)
 (teinei-masyo mirai-u)
 (teinei-masi kako-ta kako-tara kako-tari kako-taro setuzoku-te)
 (teinei-masure setuzoku-ba)

; 否定

(hitei-nai denbun-sou hikyou-you suitei-you dantei-daro dantei-desyo
 suiryuu-rasii suiryuu-rasiku suiryuu-rasikatu suiryuu-rasikere
 youtai-nai kaku-yori huku-nagara huku-toka huku-nado

huku-barari huku-kurai huku-dake huku-hodo huku-made syuu-ka
syuu-monoka setuzoku-1 setuzoku-2 setuzoku-tote setuzoku-to
setuzoku-mononara setuzoku-monono m kuten)

(hitei-nakaro mirai-u)

(hitei-naku k-nai k-nakara k-nakaro k-naku k-nakatu k-nakaru k-nakare
k-nakere setuzoku-te setuzoku-tomo touten)

(hitei-nakatu kako-ta kako-tara kako-tari kako-taro)

(hitei-nakere setuzoku-ba)

(hitei-nu denbun-sou hikyou-you suitei-you dantei-dro dantei-desho
suiryou-rasii suiryou-rasiku suiryou-rasikatu suiryou-rasikere
kaku-yori huku-nagara huku-toka huku-nado huku-bakari
huku-kurai huku-dake huku-hodo huku-made syuu-ka syuu-monoka
setuzoku-1 setuzoku-2 setuzoku-tote setuzoku-to
setuzoku-mononara setuzoku-monono m kuten)

(hitei-zu touten)

(hitei-ne setuzoku-ba)

; 推量

(suiryou-beki hikyou-you suitei-you dantei-da dantei-daro dantei-datu
dantei-na dantei-nara dantei-de dantei-desu dantei-desyo
dantei-desi suiryou-rasii suiryou-rasiku suiryou-rasikatu
suiryou-rasikere syuu-ka m)

(suiryou-bekara hitei-nu hitei-zu hitei-ne)

(suiryou-beku touten)

(suiryou-besi kuten)

(suiryou-mai huku-toka huku-nado syuu-ka setuzoku-1 setuzoku-tote
setuzoku-to kuten)

(suiryou-rasii denbun-sou syuu-ka syuu-monoka setuzoku-1 setuzoku-tote
setuzoku-monono setuzoku-2 m kuten)

(suiryou-rasiku k-nai k-nakara k-nakaro k-naku k-nakatu k-nakaru k-nakere
kakari-ha kakari-mo kakari-koso kakari-sae setuzoku-te touten)

(suiryou-rasikatu kako-ta kako-tara kako-tari kako-taro touten)

(suiryou-rasikere setuzoku-ba)

; 過去

(kako-ta denbun-sou hikyou-you suitei-you dantei-daro dantei-desho

suiryou-rasii suiryou-rasiku suiryou-rasikatu suiryou-rasikere
kaku-yori huku-toka huku-nado huku-bakari huku-kurai huku-dake
huku-hodo huku-made syuu-ka syuu-monoka setuzoku-2
setuzoku-tote setuzoku-monono m kuten)

(kako-tara setuzoku-ba touten)

(kako-tari touten touten)

(kako-taro mirai-u)

(kako-da denbun-sou hikyou-you suitei-you dantei-daro dantei-desho
suiryou-rasii suiryou-rasiku suiryou-rasikatu suiryou-rasikere
kaku-yori huku-toka huku-nado huku-bakari huku-kurai huku-dake
huku-hodo huku-made syuu-ka syuu-monoka setuzoku-2
setuzoku-tote setuzoku-monono m kuten)

(kako-dara setuzoku-ba touten)

(kako-dari touten touten)

(kako-daro mirai-u)

; 未来

(mirai-u huku-toka huku-nado syuu-ka setuzoku-tomo setuzoku-1
setuzoku-tote setuzoku-to setuzoku-mononara kuten)

(mirai-you huku-toka huku-nado syuu-ka setuzoku-tomo setuzoku-1
setuzoku-tote setuzoku-to setuzoku-mononara kuten)

; 伝聞

(denbun-sou dantei-da dante-de dantei-desu)

; 様態

(youtai-sou dantei-da dantei-daro dantei-datu dantei-na dantei-nara
dantei-de dantei-desu dantei-desyo dantei-desi)

; 比況

(hikyou-you dantei-da dantei-daro dantei-datu dantei-na dantei-nara
dantei-de dantei-desu dantei-desyo dantei-desi kaku-ni)

; 推定

(suitei-you dantei-da dantei-datu dantei-de dantei-desu dantei-desi)

; 希望

(kibou-tai denbun-sou suitei-you dantei-daro dantei-sesyo suiryō-rasii
suiryō-rasiku suiryō-rasikatu suiryō-rasikere huku-nagara
huku-toka huku-nado huku-bakari huku-kurai huku-dake huku-hodo
huku-made syuu-ka syuu-monoka setuzoku-1 setuzoku-2
setuzoku-tote setuzoku-to setuzoku-monono m kuten)

(kibou-takaro mirai-u)

(kibou-taku k-nai k-nakara k-nakaro k-naku k-nakatu k-nakaru k-nakere
setuzoku-te setuzoku-tomo touten)

(kibou-takatu kako-ta kako-tara kako-tari kako-taro)

(kibou-takere setuzoku-ba)

(kibou-ta youtai-sou)

(kibou-tagaru suiryō-mai denbun-sou hikyou-you suitei-you dantei-daro
dantei-desyo suiryō-rasii suiryō-rasiku suiryō-rasikatu
suiryō-rasikere kaku-ni kaku-yori huku-toka huku-nado
huku-bakari huku-kurai huku-dake huku-hodo huku-made syuu-ka
syuu-monoka syuu-na setuzoku-1 setuzoku-2 setuzoku-tote
setuzoku-to setuzoku-mononara setuzoku-monono m kuten)

(kibou-tagara hitei-nai hitei-nakaro hitei-naku hitei-nakatu hitei-nakere
hitei-nu hitei-zu hitei-ne)

(kibou-tagaro mirai-u)

(kibou-tagari youtai-sou teinei-masu teinei-mase teinei-masyo teinei-masi
teinei-masure huku-tutu huku-nagara kakari-ha kakari-mo
kakari-koso kakari-sae touten)

(kibou-tagatu kako-ta kako-tara kako-tari kako-taro setuzoku-te)

(kibou-tagare setuzoku-ba)

; 使役

(sieki-seru suiryō-mai suiryō-beki suiryō-bekara suiryō-beku
suiryō-besi denbun-sou hikyou-you suitei-you dantei-daro
dantei-desyo suiryō-rasii suiryō-rasiku suiryō-rasikatu
suiryō-rasikere kaku-ni kaku-yori huku-ya huku-bakari
huku-kurai huku-dake huku-hodo huku-made syuu-ka syuu-monoka
syuu-na setuzoku-1 setuzoku-2 setuzoku-tote setuzoku-to
setuzoku-mononara setuzoku-monono m kuten)

(sieki-se youtai-sou teinei-masu teinei-mase teinei-masyo teinei-masi

teinei-masure kibou-tai kibou-takaro kibou-taku kibou-takatu
kibou-takere kibou-ta kibou-tagaru kibou-tagara kibou-tagaro
kibou-tagari kibou-tagatu kibou-tagare kako-ta kako-tara
kako-tari kako-taro mirai-you hitei-nai hitei-nakaro hitei-naku
hitei-nakatu hitei-nakere hitei-nu hitei-zu hitei-ne huku-tutu
huku-nagara kakari-ha kakari-mo kakari-koso kakari-sae
setuzoku-te touten)

(sieki-sere setuzoku-ba)

(sieki-saseru suiryō-mai suiryō-beki suiryō-bekara suiryō-beku
suiryō-besi denbun-sou hikyō-you suitei-you dantei-daro
dantei-desyō suiryō-rasii suiryō-rasiku suiryō-rasikatu
suiryō-rasikere kaku-ni kaku-yori huku-ya huku-bakari
huku-kurai huku-dake huku-hodo huku-made syū-ka syū-monoka
syū-na setuzoku-1 setuzoku-2 setuzoku-tote setuzoku-to
setuzoku-mononara setuzoku-monono m kuten)

(sieki-sase youtai-sou teinei-masu teinei-mase teinei-masyō teinei-masi
teinei-masure kibou-tai kibou-takaro kibou-taku kibou-takatu
kibou-takere kibou-ta kibou-tagaru kibou-tagara kibou-tagaro
kibou-tagari kibou-tagatu kibou-tagare kako-ta kako-tara
kako-tari kako-taro mirai-you hitei-nai hitei-nakaro hitei-naku
hitei-nakatu hitei-nakere hitei-nu hitei-zu hitei-ne huku-tutu
huku-nagara kakari-ha kakari-mo kakari-koso kakari-sae
setuzoku-te touten)

(sieki-sasere setuzoku-ba)

; 受身

(ukemi-reru suiryō-mai suiryō-beki suiryō-bekara suiryō-beku
suiryō-besi denbun-sou hikyō-you suitei-you dantei-daro
dantei-desyō suiryō-rasii suiryō-rasiku suiryō-rasikatu
suiryō-rasikere kaku-ni kaku-yori huku-ya huku-bakari
huku-kurai huku-dake huku-hodo huku-made syū-ka syū-monoka
syū-na setuzoku-1 setuzoku-2 setuzoku-tote setuzoku-to
setuzoku-mononara setuzoku-monono m kuten)

(ukemi-re youtai-sou teinei-masu teinei-mase teinei-masyō teinei-masi
teinei-masure kibou-tai kibou-takaro kibou-taku kibou-takatu
kibou-takere kibou-ta kibou-tagaru kibou-tagara kibou-tagaro

kibou-tagari kibou-tagatu kibou-tagare kako-ta kako-tara
kako-tari kako-taro mirai-you hitei-nai hitei-nakaro hitei-naku
hitei-nakatu hitei-nakere hitei-nu hitei-zu hitei-ne huku-tutu
huku-nagara kakari-ha kakari-mo kakari-koso kakari-sae
setuzoku-te touten)

(ukemi-rere setuzoku-ba)

(ukemi-rareru suiryō-mai suiryō-beki suiryō-bekara suiryō-beku
suiryō-besi denbun-sou hikyou-you suitei-you dantei-daro
dantei-desyo suiryō-rasii suiryō-rasiku suiryō-rasikatu
suiryō-rasikere kaku-ni kaku-yori huku-ya huku-bakari
huku-kurai huku-dake huku-hodo huku-made syuu-ka syuu-monoka
syuu-na setuzoku-1 setuzoku-2 setuzoku-tote setuzoku-to
setuzoku-mononara setuzoku-monono m kuten)

(ukemi-rare youtai-sou teinei-masu teinei-mase teinei-masyo teinei-masi
teinei-masure kibou-tai kibou-takaro kibou-taku kibou-takatu
kibou-takere kibou-ta kibou-tagaru kibou-tagara kibou-tagaro
kibou-tagari kibou-tagatu kibou-tagare kako-ta kako-tara
kako-tari kako-taro mirai-you hitei-nai hitei-nakaro hitei-naku
hitei-nakatu hitei-nakere hitei-nu hitei-zu hitei-ne huku-tutu
huku-nagara kakari-ha kakari-mo kakari-koso kakari-sae
setuzoku-te touten)

(ukemi-rarere setuzoku-ba)

; サ変語尾

(sahen-suru suiryō-mai suiryō-beki suiryō-beku suiryō-bekara
suiryō-beku suiryō-besi denbun-sou hikyou-you suitei-you
dantei-daro dantei-desyo suiryō-rasii suiryō-rasiku
suiryō-rasikatu suiryō-rasikere kaku-ni kaku-yori huku-ya
huku-toka huku-nado huku-bakari huku-kurai huku-dake huku-hodo
huku-made syuu-ka syuu-monoka syuu-na setuzoku-1 setuzoku-2
setuzoku-tote setuzoku-to setuzoku-mononara setuzoku-monono
touten kuten)

(sahen-si youtai-sou teinei-masu teinei-mase teinei-masyo teinei-masi
teinei-masure kibou-tai kibou-takaro kibou-taku kibou-takatu
kibou-takere kibou-ta kibou-tagaru kibou-tagara kibou-tagaro
kibou-tagari kibou-tagatu kibou-tagare kako-ta kako-tara

kako-tari kako-taro mirai-you hitei-nai hitei-nakaro hitei-naku
 hitei-nakatu hitei-nakere huku-tutu huku-nagara kakari-ha
 kakari-mo kakari-koso kakari-sae setuzoku-te syuu-ro touten)
 (sahen-se hitei-nu hitei-zu hitei-ne syuu-yo)
 (sahen-sa sieki-seru sieki-se sieki-sere ukemi-reru ukemi-re ukemi-rere)
 (sahen-sure setuzoku-ba)

;;; 動詞

; 一段

(d1 d1-ru d1-re youtai-sou teinei-masu teinei-mase teinei-masyo
 teinei-masi teinei-masure kibou-tai kibou-takaro kibou-taku
 kibou-takatu kibou-takere kibou-ta kibou-tagaru kibou-tagara
 kibou-tagaro kibou-tagari kibou-tagatu kibou-tagare kako-ta
 kako-tara kako-tari kako-taro mirai-you hitei-nai hitei-nakaro
 hitei-naku hitei-nakatu hitei-nakere hitei-nu hitei-zu hitei-ne
 sieki-saseru sieki-sase sieki-sasere ukemi-rareru ukemi-rare
 ukemi-rarere huku-tutu huku-nagara kakari-ha kakari-mo
 kakari-koso kakari-sae setuzoku-te syuu-yo syuu-ro touten)
 (d1-ru suiryoku-mai suiryoku-beki suiryoku-bekara suiryoku-beku
 suiryoku-besi denbun-sou hikyoku-you suitei-you dantei-daro
 dantei-desyo suiryoku-rasii suiryoku-rasiku suiryoku-rasikatu
 suiryoku-rasikere kaku-ni kaku-yori huku-ya huku-toka huku-nado
 huku-bakari huku-kurai huku-dake huku-hodo huku-made syuu-ka
 syuu-monoka syuu-na setuzoku-1 setuzoku-2 setuzoku-tote
 setuzoku-to setuzoku-mononara setuzoku-monono m touten kuten)
 (d1-re setuzoku-ba)

; 力変

(dk-kuru suiryoku-mai suiryoku-beki suiryoku-beku suiryoku-bekara
 suiryoku-beku suiryoku-besi denbun-sou hikyoku-you suitei-you
 dantei-daro dantei-desyo suiryoku-rasii suiryoku-rasiku
 suiryoku-rasikatu suiryoku-rasikere kaku-ni kaku-yori huku-ya
 huku-toka huku-nado huku-bakari huku-kurai huku-dake huku-hodo
 huku-made syuu-ka syuu-monoka syuu-na setuzoku-1 setuzoku-2
 setuzoku-tote setuzoku-to setuzoku-mononara setuzoku-monono

touten kuten)

(dk-ki youtai-sou teinei-masu teinei-mase teinei-masyo teinei-masi
teinei-masure kibou-tai kibou-takaro kibou-taku kibou-takatu
kibou-takere kibou-ta kibou-tagaru kibou-tagara kibou-tagaro
kibou-tagari kibou-tagatu kibou-tagare kako-ta kako-tara
kako-tari kako-taro huku-tutu huku-nagara kakari-ha kakari-mo
kakari-koso kakari-sae setuzoku-te touten)

(dk-ko mirai-you hitei-nai hitei-nakaro hitei-naku hitei-nakatu
hitei-nakere hitei-nu hitei-zu hitei-ne sieki-saseru sieki-sase
sieki-sasere ukemi-rareru ukemi-rare ukemi-rarere syuu-i)

(dk-kure setuzoku-ba)

; サ変

(ds ds-suru ds-si ds-se ds-sa ds-sure)

(ds-suru suiryō-mai suiryō-beki suiryō-beku suiryō-bekara
suiryō-beku suiryō-besi denbun-sou hikyō-you suitei-you
dantei-daro dantei-desyō suiryō-rasii suiryō-rasiku
suiryō-rasikatu suiryō-rasikere kaku-ni kaku-yori huku-ya
huku-toka huku-nado huku-bakari huku-kurai huku-dake huku-hodo
huku-made syuu-ka syuu-monoka syuu-na setuzoku-1 setuzoku-2
setuzoku-tote setuzoku-to setuzoku-mononara setuzoku-monono
touten kuten)

(ds-si youtai-sou teinei-masu teinei-mase teinei-masyo teinei-masi
teinei-masure kibou-tai kibou-takaro kibou-taku kibou-takatu
kibou-takere kibou-ta kibou-tagaru kibou-tagara kibou-tagaro
kibou-tagari kibou-tagatu kibou-tagare kako-ta kako-tara
kako-tari kako-taro mirai-you hitei-nai hitei-nakaro hitei-naku
hitei-nakatu hitei-nakere huku-tutu huku-nagara kakari-ha
kakari-mo kakari-koso kakari-sae setuzoku-te syuu-ro touten)

(ds-se hitei-nu hitei-zu hitei-ne syuu-yo)

(ds-sa sieki-seru sieki-se sieki-sere ukemi-reru ukemi-re ukemi-rere)

(ds-sure setuzoku-ba)

; 五段

(d5k d5k-a d5k-i d5k-u d5k-e d5k-o d5k-onbin)

(d5k-a hitei-nai hitei-nakaro hitei-naku hitei-nakatu hitei-nakere

- hitei-nu hitei-ne hitei-zu sieki-seru sieki-se sieki-sere
ukemi-reru ukemi-re ukemi-rere)
- (d5k-i youtai-sou teinei-masu teinei-mase teinei-masyo seinei-masi
teinei-masure kibou-tai kibou-takaro kibou-taku kibou-takatu
kibou-takere kibou-ta kibou-tagaru kibou-tagara kibou-tagaro
kibou-tagari kibou-tagatu kibou-tagare huku-tutu huku-nagara
kakari-ha kakari-mo kakari-koso kakari-sae touten)
- (d5k-u suiryō-mai suiryō-beki suiryō-bakara suiryō-beku
suiryō-besi suiryō-rasii suiryō-rasiku suiryō-rasikatu
suiryō-rasikere denbun-sou hikyou-you suitei-you dantei-daro
dantei-desyo kaku-ni kaku-yori huku-ya huku-toka huku-nado
huku-bakari huku-kurai huku-dake huku-hodo huku-made syuu-ka
syuu-monoka syuu-na setuzoku-1 setuzoku-2 setuzoku-tote
setuzoku-to setuzoku-mononara setuzoku-monono touten kuten)
- (d5k-e setuzoku-ba syuu-yo kuten)
- (d5k-o mirai-u)
- (d5k-onbin kako-ta kako-tara kako-tari kako-taro setuzoku-te)
- (d5s d5s-a d5s-i d5s-u d5s-e d5s-o)
- (d5s-a hitei-nai hitei-nakaro hitei-naku hitei-nakatu hitei-nakere
hitei-nu hitei-ne hitei-zu sieki-seru sieki-se sieki-sere
ukemi-reru ukemi-re ukemi-rere)
- (d5s-i youtai-sou teinei-masu teinei-mase teinei-masyo seinei-masi
teinei-masure kibou-tai kibou-takaro kibou-taku kibou-takatu
kibou-takere kibou-ta kibou-tagaru kibou-tagara kibou-tagaro
kibou-tagari kibou-tagatu kibou-tagare kako-ta kako-tara
kako-tari kako-taro huku-tutu huku-nagara kakari-ha kakari-mo
kakari-koso kakari-sae setuzoku-te touten)
- (d5s-u suiryō-mai suiryō-beki suiryō-bakara suiryō-beku
suiryō-besi suiryō-rasii suiryō-rasiku suiryō-rasikatu
suiryō-rasikere denbun-sou hikyou-you suitei-you dantei-daro
dantei-desyo kaku-ni kaku-yori huku-ya huku-toka huku-nado
huku-bakari huku-kurai huku-dake huku-hodo huku-made syuu-ka
syuu-monoka syuu-na setuzoku-1 setuzoku-2 setuzoku-tote
setuzoku-to setuzoku-mononara setuzoku-monono touten kuten)
- (d5s-e setuzoku-ba syuu-yo kuten)
- (d5s-o mirai-u)

(d5t d5t-a d5t-i d5t-u d5t-e d5t-o d5t-onbin)

(d5t-a hitei-nai hitei-nakaro hitei-naku hitei-nakatu hitei-nakere
hitei-nu hitei-ne hitei-zu sieki-seru sieki-se sieki-sere
ukemi-reru ukemi-re ukemi-rere)

(d5t-i youtai-sou teinei-masu teinei-mase teinei-masyo seinei-masi
teinei-masure kibou-tai kibou-takaro kibou-taku kibou-takatu
kibou-takere kibou-ta kibou-tagaru kibou-tagara kibou-tagaro
kibou-tagari kibou-tagatu kibou-tagare huku-tutu huku-nagara
kakari-ha kakari-mo kakari-koso kakari-sae touten)

(d5t-u suiryou-mai suiryou-beki suiryou-bakara suiryou-beku
suiryou-besi suiryou-rasii suiryou-rasiku suiryou-rasikatu
suiryou-rasikere denbun-sou hikyou-you suitei-you dantei-daro
dantei-desyo kaku-ni kaku-yori huku-ya huku-toka huku-nado
huku-bakari huku-kurai huku-dake huku-hodo huku-made syuu-ka
syuu-monoka syuu-na setuzoku-1 setuzoku-2 setuzoku-tote
setuzoku-to setuzoku-mononara setuzoku-monono touten kuten)

(d5t-e setuzoku-ba syuu-yo kuten)

(d5t-o mirai-u)

(d5t-onbin kako-ta kako-tara kako-tari kako-taro setuzoku-te)

(d5n d5n-a d5n-i d5n-u d5n-e d5n-o d5n-onbin)

(d5n-a hitei-nai hitei-nakaro hitei-naku hitei-nakatu hitei-nakere
hitei-nu hitei-ne hitei-zu sieki-seru sieki-se sieki-sere
ukemi-reru ukemi-re ukemi-rere)

(d5n-i youtai-sou teinei-masu teinei-mase teinei-masyo seinei-masi
teinei-masure kibou-tai kibou-takaro kibou-taku kibou-takatu
kibou-takere kibou-ta kibou-tagaru kibou-tagara kibou-tagaro
kibou-tagari kibou-tagatu kibou-tagare huku-tutu huku-nagara
kakari-ha kakari-mo kakari-koso kakari-sae touten)

(d5n-u suiryou-mai suiryou-beki suiryou-bakara suiryou-beku
suiryou-besi suiryou-rasii suiryou-rasiku suiryou-rasikatu
suiryou-rasikere denbun-sou hikyou-you suitei-you dantei-daro
dantei-desyo kaku-ni kaku-yori huku-ya huku-toka huku-nado
huku-bakari huku-kurai huku-dake huku-hodo huku-made syuu-ka
syuu-monoka syuu-na setuzoku-1 setuzoku-2 setuzoku-tote
setuzoku-to setuzoku-mononara setuzoku-monono touten kuten)

(d5n-e setuzoku-ba syuu-yo kuten)

(d5n-o mirai-u)

(d5n-onbin kako-da kako-dara kako-dari kako-daro setuzoku-de)

(d5m d5m-a d5m-i d5m-u d5m-e d5m-o d5m-onbin)

(d5m-a hitei-nai hitei-nakaro hitei-naku hitei-nakatu hitei-nakere
hitei-nu hitei-ne hitei-zu sieki-seru sieki-se sieki-sere
ukemi-reru ukemi-re ukemi-rere)

(d5m-i youtai-sou teinei-masu teinei-mase teinei-masyo seinei-masi
teinei-masure kibou-tai kibou-takaro kibou-taku kibou-takatu
kibou-takere kibou-ta kibou-tagaru kibou-tagara kibou-tagaro
kibou-tagari kibou-tagatu kibou-tagare huku-tutu huku-nagara
kakari-ha kakari-mo kakari-koso kakari-sae touten)

(d5m-u suiryō-mai suiryō-beki suiryō-bakara suiryō-beku
suiryō-besi suiryō-rasii suiryō-rasiku suiryō-rasikatu
suiryō-rasikere denbun-sou hikyō-you suitei-you dantei-daro
dantei-desyō kaku-ni kaku-yori huku-ya huku-toka huku-nado
huku-bakari huku-kurai huku-dake huku-hodo huku-made syū-ka
syū-monoka syū-na setuzoku-1 setuzoku-2 setuzoku-tote
setuzoku-to setuzoku-mononara setuzoku-monono touten kuten)

(d5m-e setuzoku-ba syū-yo kuten)

(d5m-o mirai-u)

(d5m-onbin kako-da kako-dara kako-dari kako-daro setuzoku-de)

(d5r d5r-a d5r-i d5r-u d5r-e d5r-o d5r-onbin)

(d5r-a hitei-nai hitei-nakaro hitei-naku hitei-nakatu hitei-nakere
hitei-nu hitei-ne hitei-zu sieki-seru sieki-se sieki-sere
ukemi-reru ukemi-re ukemi-rere)

(d5r-i youtai-sou teinei-masu teinei-mase teinei-masyo seinei-masi
teinei-masure kibou-tai kibou-takaro kibou-taku kibou-takatu
kibou-takere kibou-ta kibou-tagaru kibou-tagara kibou-tagaro
kibou-tagari kibou-tagatu kibou-tagare huku-tutu huku-nagara
kakari-ha kakari-mo kakari-koso kakari-sae touten)

(d5r-u suiryō-mai suiryō-beki suiryō-bakara suiryō-beku
suiryō-besi suiryō-rasii suiryō-rasiku suiryō-rasikatu
suiryō-rasikere denbun-sou hikyō-you suitei-you dantei-daro
dantei-desyō kaku-ni kaku-yori huku-ya huku-toka huku-nado
huku-bakari huku-kurai huku-dake huku-hodo huku-made syū-ka
syū-monoka syū-na setuzoku-1 setuzoku-2 setuzoku-tote

setuzoku-to setuzoku-mononara setuzoku-monono touten kuten)
 (d5r-e setuzoku-ba syuu-yo kuten)
 (d5r-o mirai-u)
 (d5r-onbin kako-ta kako-tara kako-tari kako-taro setuzoku-te)
 (d5w d5w-a d5w-i d5w-u d5w-e d5w-o d5w-onbin)
 (d5w-a hitei-nai hitei-nakaro hitei-naku hitei-nakatu hitei-nakere
 hitei-nu hitei-ne hitei-zu sieki-seru sieki-se sieki-sere
 ukemi-reru ukemi-re ukemi-rere)
 (d5w-i youtai-sou teinei-masu teinei-mase teinei-masyo seinei-masi
 teinei-masure kibou-tai kibou-takaro kibou-taku kibou-takatu
 kibou-takere kibou-ta kibou-tagaru kibou-tagara kibou-tagaro
 kibou-tagari kibou-tagatu kibou-tagare huku-tutu huku-nagara
 kakari-ha kakari-mo kakari-koso kakari-sae touten)
 (d5w-u suiryou-mai suiryou-beki suiryou-bakara suiryou-beku
 suiryou-besi suiryou-rasii suiryou-rasiku suiryou-rasikatu
 suiryou-rasikere denbun-sou hikyuu-you suitei-you dantei-daro
 dantei-desyo kaku-ni kaku-yori huku-ya huku-toka huku-nado
 huku-bakari huku-kurai huku-dake huku-hodo huku-made syuu-ka
 syuu-monoka syuu-na setuzoku-1 setuzoku-2 setuzoku-tote
 setuzoku-to setuzoku-mononara setuzoku-monono touten kuten)
 (d5w-e setuzoku-ba syuu-yo kuten)
 (d5w-o mirai-u)
 (d5w-onbin kako-ta kako-tara kako-tari kako-taro setuzoku-te)
 (d5g d5g-a d5g-i d5g-u d5g-e d5g-o d5g-onbin)
 (d5g-a hitei-nai hitei-nakaro hitei-naku hitei-nakatu hitei-nakere
 hitei-nu hitei-ne hitei-zu sieki-seru sieki-se sieki-sere
 ukemi-reru ukemi-re ukemi-rere)
 (d5g-i youtai-sou teinei-masu teinei-mase teinei-masyo seinei-masi
 teinei-masure kibou-tai kibou-takaro kibou-taku kibou-takatu
 kibou-takere kibou-ta kibou-tagaru kibou-tagara kibou-tagaro
 kibou-tagari kibou-tagatu kibou-tagare huku-tutu huku-nagara
 kakari-ha kakari-mo kakari-koso kakari-sae touten)
 (d5g-u suiryou-mai suiryou-beki suiryou-bakara suiryou-beku
 suiryou-besi suiryou-rasii suiryou-rasiku suiryou-rasikatu
 suiryou-rasikere denbun-sou hikyuu-you suitei-you dantei-daro
 dantei-desyo kaku-ni kaku-yori huku-ya huku-toka huku-nado

huku-bakari huku-kurai huku-dake huku-hodo huku-made syuu-ka
 syuu-monoka syuu-na setuzoku-1 setuzoku-2 setuzoku-tote
 setuzoku-to setuzoku-mononara setuzoku-monono touten kuten)
 (d5g-e setuzoku-ba syuu-yo kuten)
 (d5g-o mirai-u)
 (d5g-onbin kako-da kako-dara kako-dari kako-daro setuzoku-de)
 (d5b d5b-a d5b-i d5b-u d5b-e d5b-o d5b-onbin)
 (d5b-a hitei-nai hitei-nakaro hitei-naku hitei-nakatu hitei-nakere
 hitei-nu hitei-ne hitei-zu sieki-seru sieki-se sieki-sere
 ukemi-reru ukemi-re ukemi-rere)
 (d5b-i youtai-sou teinei-masu teinei-mase teinei-masyo seinei-masi
 teinei-masure kibou-tai kibou-takaro kibou-taku kibou-takatu
 kibou-takere kibou-ta kibou-tagaru kibou-tagara kibou-tagaro
 kibou-tagari kibou-tagatu kibou-tagare huku-tutu huku-nagara
 kakari-ha kakari-mo kakari-koso kakari-sae touten)
 (d5b-u suiryou-mai suiryou-beki suiryou-bakara suiryou-beku
 suiryou-besi suiryou-rasii suiryou-rasiku suiryou-rasikatu
 suiryou-rasikere denbun-sou hikyou-you suitei-you dantei-daro
 dantei-desyo kaku-ni kaku-yori huku-ya huku-toka huku-nado
 huku-bakari huku-kurai huku-dake huku-hodo huku-made syuu-ka
 syuu-monoka syuu-na setuzoku-1 setuzoku-2 setuzoku-tote
 setuzoku-to setuzoku-mononara setuzoku-monono touten kuten)
 (d5b-e setuzoku-ba syuu-yo kuten)
 (d5b-o mirai-u)
 (d5b-onbin kako-da kako-dara kako-dari kako-daro setuzoku-de)

;;; 形容詞

(k k-karo k-katu k-ku k-i k-kere k-karu k-kara youtai-sou)
 (k-karu suiryou-beki suiryou-bakara suiryou-beku suiryou-besi)
 (k-i denbun-sou suitei-you dantei-daro dantei-desyo suiryou-rasii
 suiryou-rasiku suiryou-rasikatu suiryou-rasikere kaku-yori
 huku-nagara huku-ya huku-toka huku-nado huku-barari huku-kurai
 huku-dake huku-hodo huku-made syuu-ka syuu-monoka setuzoku-1
 setuzoku-2 setuzoku-tote setuzoku-to setuzoku-monono touten
 kuten)
 (k-katu kako-ta kako-tara kako-tari kako-taro)

(k-karo mirai-u)

(k-kara hitei-nu hitei-ne hitei-zu)

(k-ku k-nai k-nakara k-nakaro k-naku k-nakatu k-nakaru k-nakere
k-nasa kakari-ha kakari-mo kakari-koso kakari-sae kakari-sika
setuzoku-te setuzoku-tomo touten)

(k-kere setuzoku-ba)

(k-nai denbun-sou suitei-you dantei-daro dantei-desyo suiryō-rasii
suiryō-rasiku suiryō-rasikatu suiryō-rasikere kaku-yori
huku-nagara huku-toka huku-nado huku-bakari huku-kurai
huku-dake huku-hodo huku-made syuu-ka syuu-monoka setuzoku-1
setuzoku-2 setuzoku-tote setuzoku-to setuzoku-mononara
setuzoku-monono m kuten)

(k-nakara hitei-zu hitei-ne hitei-nu)

(k-nakaro mirai-u)

(k-naku k-nai k-nakara k-nakaro k-naku k-nakatu k-nakaru k-nakare
k-nakere setuzoku-te setuzoku-tomo touten)

(k-nakatu kako-ta kako-tara kako-tari kako-taro)

(k-nakaru suiryō-beki suiryō-bekara suiryō-beku)

(k-nakere setuzoku-ba)

(k-nasa youtai-sou)

(k-onaji hikyō-you suitei-you dantei-da dantei-daro dantei-datu
dantei-desu dantei-desyo dantei-desi suiryō-rasii
suiryō-rasiku suiryō-rasikatu suiryō-rasikere kaku-yori
huku-nagara syuu-ka syuu-monoka m)

(k-onajikara hitei-zu hitei-nu hitei-ne)

;;; 名詞

(m dantei-da dantei-daro dantei-datu dantei-na dantei-nara
dantei-de dantei-desu dantei-desyo dantei-desi kaku-ga kaku-wo
kaku-ni kaku-de kaku-to kaku-he kaku-kara kaku-yori no-no
huku-ten huku-ya huku-nado huku-made huku-kurai huku-bakari
huku-dake huku-hodo huku-sura huku-koro kakari-kiri
kakari-tote kakari-sae kakari-koso kakari-ha kakari-mo
kakari-sika kakari-dokoroka syuu-ka kantou-ne kantou-yo
sahen-suru sahen-si sahen-se sahen-sa sahen-sure m s
touten kuten)

;;; 副詞

(h touten)

;;; 連体詞

(r touten)

;;; 接続詞

(s touten)

;;; 読点

(touten d1 d5k d5s d5t d5n d5m d5r d5w d5g d5b
dk-kuru dk-ki dk-ko dk-kure
ds-suru ds-si ds-se ds-sa ds-sure
k-nai k-nakara k-nakaro k-naku k-nakatu k-nakaru k-nakere
k-nasa k-onazi k-onazikara
m k h r s)

謝辞

本研究は、著者が筑波大学大学院博士課程工学研究科電子・情報工学専攻において5年間にわたり行ったもので、その間終始多大なる御指導をいただいた筑波大学電子・情報工学系 板橋秀一教授に謹んで感謝の意を表します。

東京工科大学の西野博二教授には、学群学生時代に自然言語処理研究を御指導いただきました。ここに深厚なる謝意を表します。

ニューラルネットの数理・設計に関して御指導をいただきました筑波大学電子・情報工学系 平井有三助教授に謹んで感謝いたします。

本論文をまとめるにあたり、有益な御指導・御助言を頂きました筑波大学電子・情報工学系 池辺八洲彦教授、並びに 同 西原清一 教授に感謝いたします。

自然言語処理に関して御指導・御助言を頂きました通商産業省工業技術院電子技術総合研究所 自然言語研究室の横山晶一室長に感謝いたします。

研究を行うにあたり、激励を頂いた筑波大学 知能情報研究室の諸氏に感謝します。

参考文献

- [Aho72] Aho, A. V. and Ulman, J. D. : The Theory of Parsing, Translation, and Compiling, Vol.1 and 2, Prentice-Hall (1972).
- [麻生 88] 麻生英樹 : ニューラルネットワーク情報処理, 産業図書 (1988).
- [Charniak83] Charniak, E. : Passing markers: A theory of contextual influence in language comprehension, *Cognitive Science*, Vol.7, pp.171-190 (1983).
- [Farlman83] Farlman, S. E., Hinton, G. E., and Sejnowski, T. J. : Massively parallel architectures for AI: NETL, Thistle, and Boltzman Machines, *Proceedings of the National Conference on Artificial Intelligence AAAI-83*, pp.109-113 (1983).
- [福島 91] 福島俊一 : 形態素抽出マシン MEX-II の試作と評価, 情報処理学会第 42 回全国大会論文集 1C-8 (1991).
- [郡司 86] 郡司隆男 : 日本語句構造文法, 人工知能学会誌, Vol.1, No.2, pp.45-52 (1989).
- [橋本 46] 橋本進吉 : 「国語学概論」, 岩波書店 (1946).
- [Higuchi91a] Higuchi, T., Furuya, T., Handa, K., Takahashi, N., Nishiyama, H., and Kokubu, A. : *IXM2 : A Parallel Associative Processor*, 18th Int'l, Symp. on Computer Architecture (1991).
- [Higuchi91b] Higuchi, T., Kitano, H., Furuya, T., Handa, K., Takahashi, N., and Kokubu, A. : *IXM2 : A Parallel Associative Processor for Knowledge Processing*, AAAI (1991).
- [Hillis85] Hillis, W. D. : *The Connection Machine*, MIT Press, Cambridge (1985).
- [Hirst87] Hirst, G. : *Semantic interpretation and the resolution of ambiguity*, Cambridge University Press, Cambridge (1987).

- [Hopfield84] Hopfield, J. J. : Neurons with Graded Response Have Collective Computational Properties like those of Two-state Neurons, *Proceedings of the National Academy of Science USA* 81, pp.3088-3092 (1984).
- [Hopfield85] Hopfield, J. J. and Tank, D. W. : Neural Computation of Decisions in Optimization Problems, *Biological Cybernetics*, Vol.52, pp.141-152 (1985).
- [木村 90] 木村和広, 鈴岡節, 伊藤悦雄, 天野真家 : 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論文集 9-3, pp.301-304 (1990).
- [金田一 74] 金田一京助, 金田一春彦, 見坊豪紀, 柴田武, 山田忠雄編 : 新明解国語辞典第二版 (磁気テープ版), 三省堂 (1974).
- [Kirkpatrick83] Kirkpatrick, S., Gellat, C. D., and Vecchi, M. P. : Optimization by simulated annealing, *Science* 220, pp.671-680 (1983).
- [澗瀉 90] 澗瀉健一, 荒木健治, 宮永喜一, 栃内香次 : 表層より得られる単語共起関係とその評価, 情報処理学会研究報告, 90-NL-80 (1990).
- [Marcus80] Marcus, M. P. : Algorithm Schemata and Data Structures in Syntactic Processing, The MIT Press (1980).
- [松川 89] 松川智義, 長尾真, 中村順一 : 共起関係に注目した DM 分解と確率的推定による単語のクラスタリング, 情報処理学会研究報告, 89-NL-72 (1989).
- [松本 89] 松本一則, 黒岩真吾, 鈴木雅実, 榊博史 : 共起関係データの蓄積と利用のための基礎実験, 情報処理学会研究報告, 89-NL-73 (1989).
- [森 89] 森辰則, 中川裕志 : Connectionist Model による構文解析モデル, 情報処理学会論文誌, Vol.30, No.4, pp.447-456 (1989).
- [村川 82] 村川堅太郎, 江上波夫, 山本達郎, 林健太郎 : 詳説 世界史 (再訂版), 山川出版社 (1982).
- [村瀬 89] 村瀬功, 中川聖一 : ボルツマンマシンによる文節ラティスの係り受け解析, 情報処理学会第 38 回全国大会論文集 5E-8, pp.378-379 (1989).

- [長尾 78] 長尾真, 辻井潤一, 山上明, 建部周二: 国語辞書の記憶と日本語文の自動分割, 情報処理, Vol.19, No.6, pp.514-521 (1978).
- [長尾 83] 長尾真: 言語工学, 昭晃堂 (1983).
- [西村 64] 西村恕彦: 機械翻訳研究会資料 (1964).
- [西村 78] 西村恕彦, 水谷静夫, 尾上佳介, 田中幸子: 日本語基本文法 — 単文篇 —, 電子技術総合研究所研究報告 第 783 号 (1978).
- [荻野 89] 荻野孝野, 山本由起雄, 清野正樹, 縄田路子, 内田裕士: 共起関係に注目した動詞分類について, 情報処理学会研究報告, 89-NL-71 (1989).
- [尾関 86] 尾関和彦: 最適文節列を選択するための多段決定アルゴリズム, 電子通信学会技術研究報告, SP86-32, pp.41-48 (1986).
- [奥村 90] 奥村明俊, 山端潔, 村木一至: ニューラルネットワークによる日本語係り受け構造の学習, 第 4 回人工知能学会全国大会論文集 11-5, pp.353-356 (1990).
- [Prager86] Prager, R. W., Harrison, T. D., and Fallside, F.: Boltzmann Machines for Speech Recognition, *Computer Speech and Language*, Vol.1, pp.3-27 (1986).
- [Quillian67] Quillian, M. R.: Word Concepts: A Theory and Simulation of Some Basic Semantic Capabilities, *Behavioral Science*, Vol.12, pp.410-430 (1967).
- [Rumelhart86] Rumelhart, D. E. et al.: *Parallel Distributed Processing*, The MIT Press, Cambridge (1986).
- [杉村 88] 杉村領一, 赤坂宏二, 久保幸弘, 松本裕治, 佐野洋: 論理型形態素解析 LAX, Proc. of the Logic Programming Conf., ICOT, pp.213-222 (1988).
- [白井 86] 白井克彦, 浜田朋之, 鈴木径夫: 係り受け関係の共起性に基づいた構文解析, 情報処理学会第 32 回全国大会論文集 4S-1, pp.1617-1618 (1986).
- [鈴木 86] 鈴木径夫, 白井克彦: 係り受け解析における共起関係利用の効果, 情報処理学会第 33 回全国大会論文集 3K-5, pp.1807-1808 (1986).

- [高橋 88] 高橋直人, 板橋秀一: 単語共起頻度を利用した形態素解析, 情報処理学会研究報告, 88-NL-69 (1988).
- [高橋 89a] 高橋直人, 板橋秀一: 単語共起頻度を用いた日本語解析, 情報処理学会第 38 回全国大会論文集 5E-3, pp.368-369 (1989).
- [高橋 89b] 高橋直人: 単語共起頻度を利用した日本語解析, 筑波大学大学院博士課程工学研究科修士論文 (1989).
- [高橋 90a] 高橋直人, 板橋秀一: 相互結合型ニューラルネットワークによる日本語の係り受け解析, 情報処理学会第 40 回全国大会論文集 4F-7, pp.464-465 (1990).
- [高橋 90b] Takahashi, N. and Itahashi, S.: Japanese Sentence Analysis Utilizing Mutually Connected Neural Network, *Proceedings of PRICAI'90*, pp.257-262 (1990).
- [高橋 90c] 高橋直人, 板橋秀一: ニューラルネットによる日本語形態素・係り受け解析, 情報処理学会研究報告, 90-NL-80 (1990).
- [高橋 91a] 高橋直人, 板橋秀一, 平井有三: 日本語形態素解析用ニューラルネットワークについて, 情報処理学会第 42 回全国大会論文集 1C-2 (1991).
- [高橋 91b] 高橋直人, 板橋秀一: 文章中の後続単語を連想するネットワークモデルについて, 情報処理学会第 43 回全国大会論文集 2G-9 (1991).
- [高橋 91c] 高橋直人, 板橋秀一: 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 (1991).
- [Takefuji89] Takefuji, Y. and Lee, K. C.: A Near-Optimum Parallel Planarization Algorithm, *Science*, Vol.245, pp.1221-1223 (1989).
- [田村 87] 田村淳, 安西祐一郎: 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).
- [田中 89] 田中穂積: 自然言語解析の基礎, 産業図書 (1989).
- [田中・吉田 88] 田中康仁, 吉田将: 自然言語の知識獲得 (その 1), 情報処理学会研究報告, 88-NL-69 (1988).

[田中・吉田 89] 田中康仁, 吉田将 : 語と語の関係について, 情報処理学会研究報告, 89-NL-73 (1989).

[上村 90] 上村龍太郎 : Recognition and Generation of Natural Language with Recurrent Neural Network, 情報処理学会研究報告, 90-NL-77 (1990).

[Waltz85] Waltz, 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).

[吉村 82] 吉村賢治, 日高達, 吉田将 : 日本語文の形態素解析における最長一致法と文節数最少法について, 情報処理学会自然言語研究会資料 30-7 (1982).

[湯浅 90] 湯浅太一 : コネクションマシンのソフトウェア, *bit*, Vol.22, No.11, 共立出版 (1990).

発表論文一覧

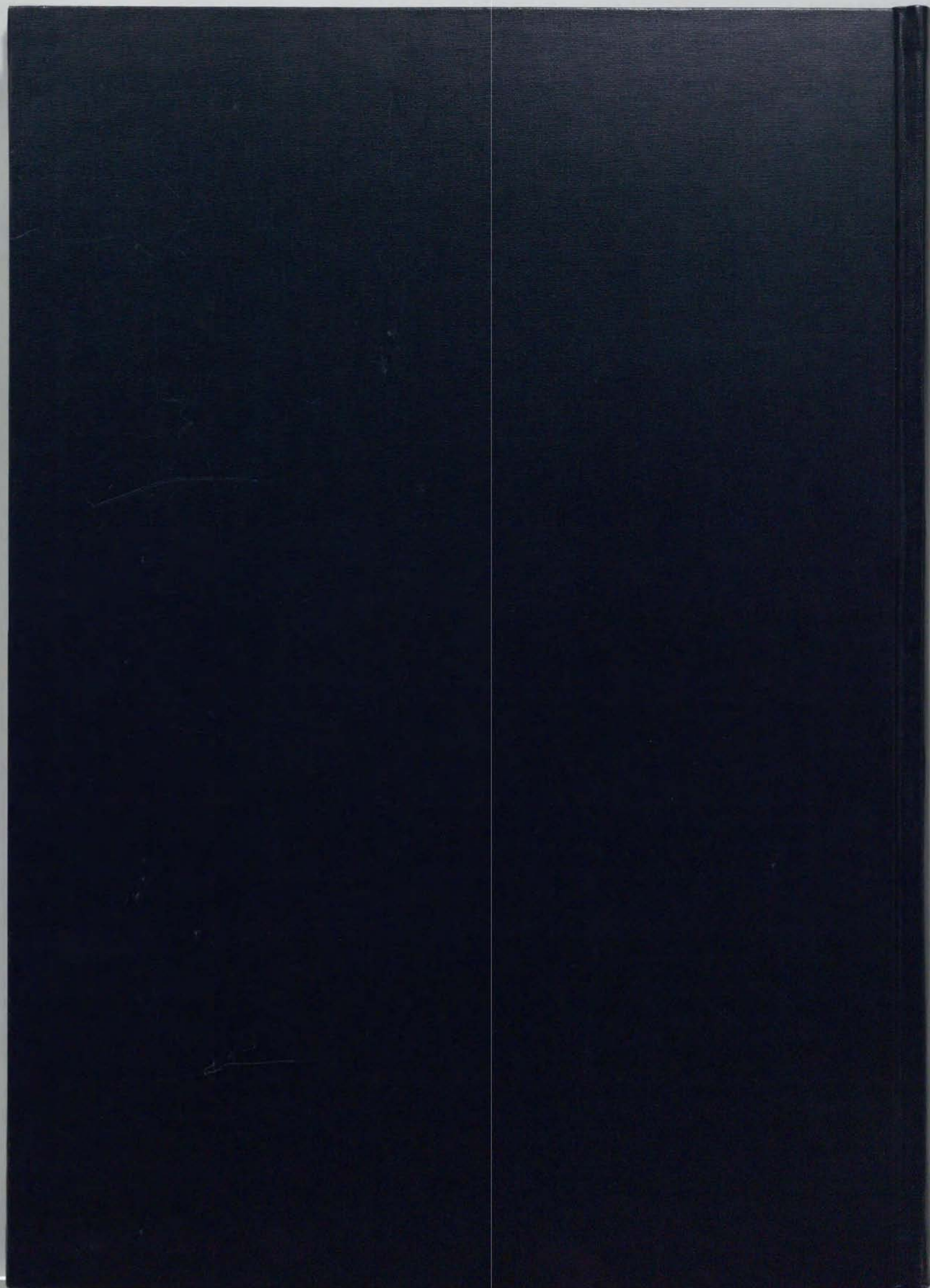
審査付公表済

- 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 (1991).
- Takahashi and Itahashi : *Japanese Sentence Analysis Utilizing Mutually Connected Neural Network*, PRICAI'90 (1990)
- Higuchi, T., Furuya, T., Handa, K., Takahashi, N., Nishiyama, H., and Kokubu, A. : *IXM2 : A Parallel Associative Processor*, 18th Int'l, Symp. on Computer Architecture (1991).
- Higuchi, T., Kitano, H., Furuya, T., Handa, K., Takahashi, N., and Kokubu, A. : *IXM2 : A Parallel Associative Processor for Knowledge Processing*, AAI (1991).

その他口頭発表

- 三浦, 高橋, 板橋 : 結合価文法における名詞の Distribution と意味素性, 情報処理学会第 36 回全国大会 4U-1 (1988)
- 高橋, 板橋 : 単語共起頻度を利用した形態素解析, 情報処理学会研究報告, 88-NL-69 (1988).
- 高橋, 板橋 : 単語共起頻度を用いた日本語解析, 情報処理学会第 38 回全国大会 5E-3 (1989).
- 高橋, 板橋 : 相互結合型ニューラルネットワークによる日本語の係り受け解析, 情報処理学会第 40 回全国大会 4F-7 (1990).
- 樋口, 高橋, 半田, 古谷, 国分 : 意味記憶システム IX - 意味ネットマシン IXM2 による word disambiguation の実験 -, 情報処理学会第 40 回全国大会 4F-8 (1990).

- 高橋, 板橋 : ニューラルネットワークによる日本語形態素・係り受け解析, 情報処理学会研究報告, 90-NL-80 (1990).
- 高橋, 板橋, 平井 : 日本語形態素解析用ニューラルネットワークについて, 情報処理学会第42回全国大会 1C-2 (1991).
- 高橋, 板橋 : 文章中の後続単語を連想するネットワークモデルについて, 情報処理学会第43回全国大会 2G-9 (1991).



inches 1 2 3 4 5 6 7 8
cm 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

Kodak Color Control Patches

© Kodak, 2007 TM: Kodak



Kodak Gray Scale



© Kodak, 2007 TM: Kodak

A 1 2 3 4 5 6 M 8 9 10 11 12 13 14 15 B 17 18 19

