

概要

現在の音声認識では、言語モデルに単語 N -gram 言語モデルが用いられることが多い。しかし、この単語 N -gram 言語モデルは言語の隣接情報であり統計的性質を用いて正しい認識結果を推定するもので統語的な情報は含まれていない。そのため、統語的な情報を用いることにより認識精度は更に向上すると考えられる。そこで本研究では統語的な情報として結合価パターンを用いた。

本研究は、 N -gram と結合価パターンを用いることで、結合価パターンがどの程度認識性能を向上するかを検証した。

結合価パターンに適した (用言のパターンの中に 1 つ以上の格要素が存在する場合) 単文 100 文を実験データに用い、単語 bi-gram 言語モデルを用いた音声認識プログラムで得られた候補文に対して、結合価パターンを用いた。実験データ 100 文中 11 文に候補順位向上が見られた。また、第 1 候補文は 49 文から 54 文に増加した。しかしながら大部分の候補が結合価パターンを用いなくても向上が見られるものであった。結合価パターンを用いなければ候補順位向上が見られない文は 11 文中 2 文であった。そこで、パターン適用規則を見直し、用言のパターン全ての格要素が存在する場合により実験を行った。実験データは 100 文中 70 文を用いることができた。70 文中 23 文の候補順位向上が見られた。

結果より、単文認識においてパターン適用規則を「全ての格要素が存在する」とすることにより結合価パターンによって大きく認識精度向上する。

目次

1	まえがき	1
2	音声認識	2
2.1	音声分析	3
2.2	音響モデル	4
2.3	言語モデル	6
3	結合価パターンの候補文への適用	7
3.1	結合価パターン	7
3.2	一般名詞意味属性体系	8
3.3	結合価パターンの適用方法	9
4	実験	10
4.1	実験データ	10
4.2	音声認識プログラム	11
4.3	評価方法	13
4.4	実験手順	13
5	実験結果と考察	14
5.1	実験結果 (1)	14
5.2	出力結果例	15
5.2.1	実験データ例	15
5.2.2	候補順位向上例	16
6	考察	17
6.1	候補昇格文の調査	17
6.2	結合価パターン適用文の調査	17
6.3	結合価パターンの問題点	18
7	結合価パターン適用規則の見直し	19
7.1	実験用データ	19
7.2	候補順位向上例	21
7.3	考察	22

8	結論	23
9	まとめ	24

目 次

1	音声認識の基本的な流れ	2
2	HMM の例	4
3	HMM の結合	4
4	一般名詞意味属性体系	8
5	tree-trellis サーチのアルゴリズム	12
6	実験手順	13
7	実験結果	14
8	実験結果	20

表 目 次

1	結合価パターンの例	7
2	用言「受験する」の結合価パターン	9
3	入力単文の例	10
4	ATR-Bset 文の例	11
5	各候補までの累積正解率	14
6	用言「くれる」の結合価パターン	16
7	用言「遅れる」の結合価パターン	16
8	各候補までの累積正解率	20
9	用言「くれる」の結合価パターン	21
10	用言「遅れる」の結合価パターン	21

1 まえがき

音声認識は、音声分析、音素モデル、言語モデル、サーチ、適応、などの技術で構成されている。言語モデルは主に、 N -gram等の統計的な手法によるものと、ネットワーク文法、文脈自由型文法等の規則によって制約するものがある。規則によって制約する言語モデルは、規則の生成を人手により行うためその生成は困難であり、また規則は確定されおらず、非文を出力しやすい傾向がある。そのため、現在の音声認識はその精度から統計的な手法の一つである単語 N -gram が広く用いられている。しかし、単語 N -gram は言語の隣接統計情報であり、統語的な情報を含んでいない。そこで本研究では、日本語彙体系 [1] の「構文体系」に記載されている結合価パターンが格要素と用言に対する意味的制約効果を持つ点に着目し、統語的な情報として結合価パターンを用いた。 N -gram と結合価パターンを用いることで、結合がパターンがどの程度音声認識性能を向上するかを調査した。結合価パターンに適した (用言のパターンの中に 1 つ以上の格要素が存在する場合) 単文 100 文を実験データに用い、単語 bi-gram 言語モデルを用いた音声認識プログラムで得られた候補文に対して、結合価パターンを用いた。実験データ 100 文中 11 文に候補順位向上が見られた。また、第 1 候補文は 49 文から 54 文に増加した。しかしながら大部分の候補が結合価パターンを用いなくても向上が見られるものであった。結合価パターンを用いなければ候補順位向上が見られない文は 11 文中 2 文であった。そこで、パターン適用規則を見直し、用言のパターン全ての格要素が存在する場合により実験を行った。実験データは 100 文中 70 文を用いることができた。70 文中 23 文の候補順位向上が見られた。

以下第 2 章では音声認識について、第 3 章では結合価パターンによる候補文の削減について、第 4 章では実験について、第 5 章で実験結果について、第 6 章で考察について、第 7 章で結合価パターン適用方法の見直し、第 8 章で結論について、第 9 章でまとめを述べる

2 音声認識

音声認識とは下の図1のように、音声分析、音響モデル、言語モデル、探索過程などを持つ技術である。入力音声は音声分析により認識に有効な本質的な特徴が抽出される。あらかじめ大量の学習データから各音素の特徴を隠れマルコフモデル (HMM) により表現して音響モデルとして使用する。また、語彙、文法あるいは言語統計などにより、発声内容を規定するモデルが言語モデルである。言語モデルで規定された探索空間の中で、入力音声をもっともよく説明できるような音素モデルの列を捜し出すのが、探索過程の役割である。以上の構成要素により、与えられた文法の中で、もっとも入力された音声特徴時系列に合致する音素列を選び出し、認識結果として出力する。

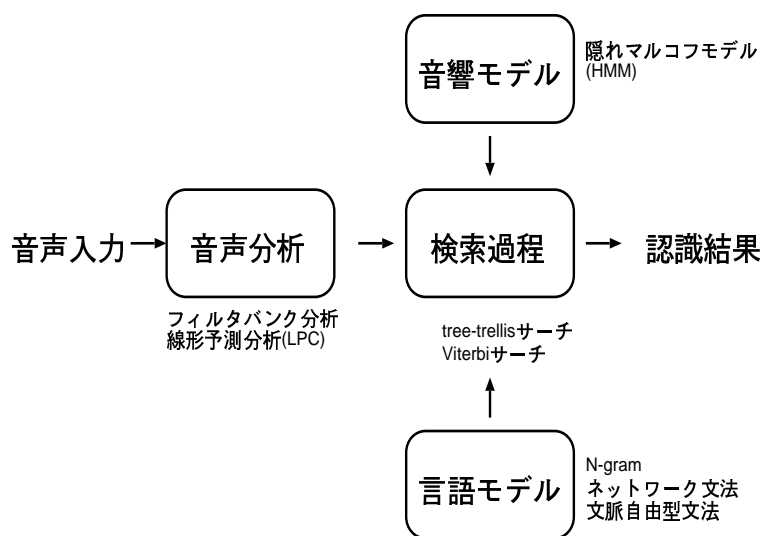


図 1: 音声認識の基本的な流れ

2.1 音声分析

音声信号は信号の性質が常に変化している非定常信号であるものの、100分の1秒程度の短時間区間では定常的な信号とみなすことができるので、定常過程 (stationary process) に対するスペクトル推定 (spectral estimation) の方法を利用することができる。切り出し区間の両端に急激な変化が起こらないように、ハミング窓 (Hamming window)、ハニング窓 (Hanning window) などの窓関数を波形にかけてスペクトル分析を行う。この1分析フレーム (analysis frame) の長さを15~30ms程度にし、5~20ms程度ずつシフトしながら音声分析を行う。

以前は音声の特徴を抽出する手段として、FFT (高速フーリエ変換) により計算したスペクトルパラメータが用いられていた。現在では、音声データに対して特徴パラメータ抽出を行い、スペクトルパラメータに変換したものをを用いる。特徴パラメータ抽出を行う方法としてフィルタバンク分析 (filter bank analysis) や線形予測法 (LPC, linear predictive coding) がある。特徴パラメータには、ケプストラムやメルケプストラムなどが用いられる。ケプストラムは波形の短時間振幅スペクトルの対数を逆フーリエ変換したものと定義され、スペクトル包絡と微細構造を近似的に分離できる特徴がある。

特徴パラメータには、FFTのスペクトルを対数変換したものに逆フーリエ変換したメル周波数ケプストラム係数 (MFCC, Mel Frequency Cepstrum Coefficient) やデルタケプストラム (MFCCの1次差分, Δ MFCC)、対数パワー (LogPow) やデルタ対数パワー (LogPow の1次差分, Δ LogPow) などがある。

2.2 音響モデル

音素 HMM

従来、音声認識には音声パターンの時間軸上の変動を吸収できる動的計画法を用いた非線形伸縮パターンマッチング法 (DP マッチング) が用いられていた。しかし、以下のような利点のために、隠れマルコフモデル (HMM) が近年主流になりつつある。

- スペクトル時系列の統計的変動をモデルのパラメータに反映できる図 2 のように、確率モデルの状態の遷移が時間遷移を表し、その遷移は確率 a_{ij} によって表されているので、通常範囲の時間変動を吸収することができる。また、出力シンボルが周波数のパラメータを表し、そのシンボル k の出現は出力確率 $b_{ij}(k)$ で表すため、ある程度の周波数方向の変動も吸収できる。
- 比較的簡単なモデルのパラメータ推定法が存在する。
- 確率モデルでは二つのモデルを結合する際、図 3 のようにそのモデル間の結合も確率で表すことができるため、滑らかに結合した新たなモデルを得ることができる。
- 認識時の計算量が比較的少ない。

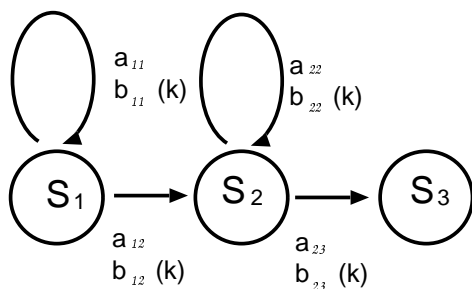


図 2: HMM の例

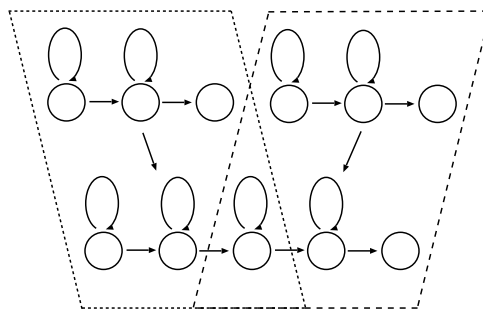


図 3: HMM の結合

一方、統計的計算が基本となるため、次に挙げるような欠点も存在する。

- モデルのパラメータを決定するための学習処理が、やや複雑で計算量が多い。
- 不特定話者で音素単位の認識を行うためには、学習に音素のバランスのとれた大量の音声が必要とする。

音声認識に用いる HMM は次の六つの集合 $M = (S, Y, A, B, \pi, F)$ で定義する。

S : 状態の有限集合; $S = \{s_i\}$

A : 状態遷移確率の集合; $A = \{a_{ij}\}$

a_{ij} は状態 s_i から s_j への遷移確率。

$$\sum_j a_{ij} = 1$$

B : 出力確率の集合; $B = \{b_{ij}(k)\}$

$b_{ij}(k)$ は状態 s_i から状態 s_j へ遷移する際にシンボル k を出力する確率。

$$\sum_k b_{ij}(k) = 1 \quad (\text{離散出力分布型 HMM の場合})$$

$$\int_{-\infty}^{\infty} b_{ij}(k) dk = 1 \quad (\text{連続出力分布型 HMM の場合})$$

Π : 初期状態確率の集合; $\Pi = \{\pi_i\}$

π_i は初期状態が s_i である確率。

$$\sum_j \pi_j = 1$$

F : 最終状態の集合

ここで、出力シンボルの集合 Y は有限集合の場合と無限集合の場合があり、B の表現方法も異なる。有限集合の場合は出力確率分布は離散的であり、無限集合の場合では出力確率分布はガウス分布のような連続分布である。離散出力分布型 HMM を用いる場合、音声特徴パラメータをベクトル量子化により離散シンボルに変換する。しかし、ベクトル量子化により量子化誤差が生じるため、認識結果に影響を与える可能性がある。そのため、量子化誤差のない連続出力分布型 HMM を用いて音声認識を行うことが多い。

2.3 言語モデル

単語 N -gram 言語モデル

認識における単語系列 W の出現確率の推定値 $P(W)$ を与える。 $P(W)$ を単語系列 W の言語モデル尤度と呼ぶ。候補の単語系列 W_1, W_2 があるとき、 $P(W_1) > P(W_2)$ ならば W_1 の方が W_2 よりも正解である可能性が高い。 W_1^R がある単語系列であると仮定する。

$$W_1^R = \omega_1 \omega_2 \cdots \omega_R$$

単語系列の出現確率を各単語の条件付き出現確率の積と考え、次のようにして $P(W_1^R)$ を求める。

$$P(W) = P(\omega_1)P(\omega_2|\omega_1)P(\omega_3|\omega_1\omega_2)\cdots P(\omega_R|W_1^{R-1})$$

しかし、すべての単語とすべての単語系列長に対して、すべての条件付き確率を推定することは事実上不可能である。そこで、以下のように W_1^{i-1} を W_{i-N+1}^{i-1} で近似することで、これを現実的に作成可能なモデルにしている。

$$P(\omega_i|W_1^{i-1}) \simeq P(\omega_i|W_{i-N+1}^{i-1})$$

これは先行する $N - 1$ 単語のみについての条件付き出現確率を求めるものである。実際には、条件付き確率 $P(\omega_i|W_{i-N+1}^{i-1})$ の計算に大量の学習用テキストコーパスと莫大な量の計算が必要となるため、 $N=2$ の bi-gram、あるいは少数の候補についての場合でも $N=3$ の tri-gram が実現上の限界となる。後述する本実験で用いられる音声認識プログラムには bi-gram モデルを用いている。bi-gram モデルは簡潔さ・有効性などの点から言語モデルとして広く用いられている。bi-gram モデルは着目している単語 ω_i の直前の単語 ω_{i-1} だけを考慮するもので、極めて局所的な単語の共起条件を表しているといえる。

$$P(\omega_i|\omega_{i-1})$$

3 結合価パターンの候補文への適用

結合価パターンは、名詞の持つ意味属性により用言と格要素(名詞+格助詞)の意味的共起関係体系化したものである。結合価パターンの使用により名詞と動詞の意味的な誤りが調査可能である。

3.1 結合価パターン

本研究では、日本語語彙体系 [1] の「構文体系」のパターンを使用する。「構文体系」は日本語の用言 6000 語の文型パターンを記したものである。表 1 に結合価パターンの例を示す(括弧内は意味属性)。結合価パターンでは、格要素として一般名詞意味属性が、用言として動詞が用いられている。

表 1: 結合価パターンの例

$N1(3 \text{ 主体})$ が $N2(1236 \text{ 人間活動 } 2064 \text{ 変動})$ に 動く
$N1(962 \text{ 機械})$ が $N2(2335 \text{ 電気})$ で 動く
$N1(3 \text{ 主体})$ が $N2(986 \text{ 乗り物})$ を $N3(388 \text{ 場所 } 2610 \text{ 場})$ に 呼ぶ
$N1(533 \text{ 具体物})$ は $N2(709 \text{ 成分})$ を 呼ぶ

格要素は「(3 主体) が」、や「(1236 人間活動 2064 変動) に」のように、一般名詞意味属性によって分類、規制された名詞((3 主体)、(1236 人間活動 2064 変動)、等)と助詞を組み合わせられて構成されている。そして格要素により単一の用言は意味的に分類、規制され、複数の結合価パターンが存在する。

3.2 一般名詞意味属性体系

一般名詞意味属性 (図 4) とは、対象と単語の関係を対象の捉え方に注目して分類したものである。一般名詞意味属性では約 40 万語の単語を約 2700 種類の意味属性に分類されている。日本語のシソーラスとしての規模、語彙数、意味分解能は世界最大級のものである。

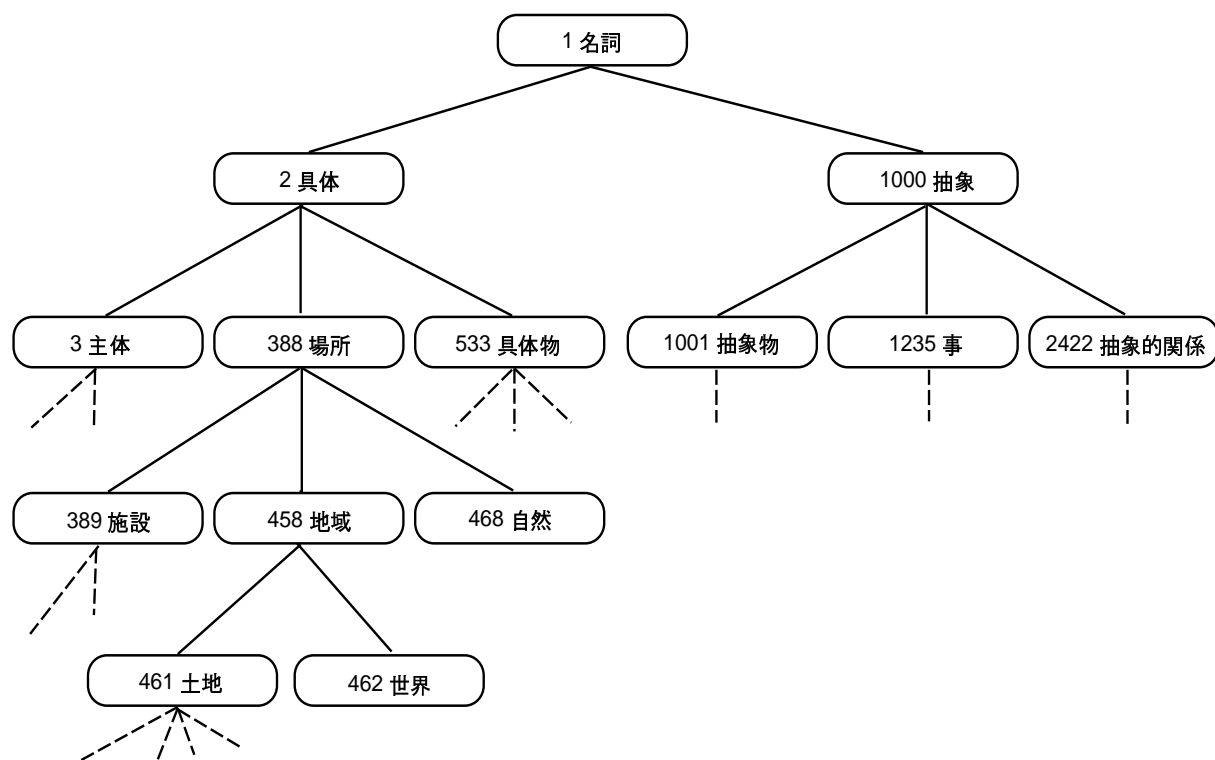


図 4: 一般名詞意味属性体系

図 8 は意味属性体系の構造の一部を記したものである。この意味属性体系には、下位の属性は上位の属性に付随するといった特性がある。例えば、「地域」という意味属性には「土地」、「世界」などの属性が含まれているが、「地域」の上位属性の「場所」には「地域」だけでなく、「施設」、「自然」といった属性に含まれる単語すべてが内包されている。

3.3 結合価パターンの適用方法

結合価パターンの有効性を評価する方法を、以下の文を例に説明する。

<音声入力文> 私は彼の高校を受験する。 (1)

音声認識プログラムに例文(1)を入力した場合、第1候補文、第2候補文として以下の2文が出力されたと仮定する。

第1候補文：私からの高校と受験する。

第2候補文：私は彼の高校を受験する。

この2つの候補文で用いられている用言「受験する」の結合価パターンを以下に示す。
(括弧内は名詞意味属性)

表 2: 用言「受験する」の結合価パターン

①	N1(4人) [は/が] N2(362組織) を 受験する
②	N1(4人) [は/が] N2(1426試験) を 受験する
③	N1(4人) [は/が] 受験する

本実験では、用言に対するパターンに1つ以上の格要素が存在するときにパターンを満たしていると判断する。上記のパターン内で、第1候補文を満たすパターンは存在しないので、第1候補文は非文と判断し削除する。第2候補文を満たすパターン①が存在するため、第2候補文を入力文の正解候補文第1位とする。

4 実験

4.1 実験データ

実験データとして、単文を用いる。音声入力話者1名(男性)により行う。入力単文は結合価パターンに適したものを採用する。以下に入力単文の一部を示す。(全文は付録参照)

表 3: 入力単文の例

風は東から吹いている。
彼は仕事に身が入る。
私は子供に名前を付けた。
彼は展覧会に絵を出す。
彼は約束を破った。
彼女は悲鳴を上げる。
我々はその話を前に聞いた。
•
•
•

4.2 音声認識プログラム

音声認識プログラムとして当研究室にある tree-trellis 型の音声認識プログラムを用いる。音声認識プログラムは言語モデルに単語 bi-gram(2-gram) を用い、語彙数は 8192 単語とする。音素 hmm の学習は、HTK[2] を用いる。ATR-Aset5240 単語の男性話者 10 名による発話データを用いて不特定話者モデルを作成し、実験データと同一話者による ATR-Bset503 文の同一発話文を用いて連結学習し音素モデルを作成する。言語モデルに使用する単語 bi-gram(2-gram) データは、毎日新聞 1993 年度 1 年文(約 108 万文)と実験データの入力単文を用いて作成する。

以下に音素モデル作成に用いた ATR-Bset の文の一部を示す。(全文は付録参照)

表 4: ATR-Bset 文の例

A01	あらゆる 現実を すべて 自分の ほうへ ねじ曲げたのだ。
A02	一週間ばかり ニューヨークを 取材した。
A03	テレビゲームや パソコンで ゲームをして 遊ぶ。
A04	物価の 変動を 考慮して 給付水準を 決める 必要が ある。
A05	救急車が 十分に 動けず、 救助作業が 遅れている。
	・
	・
	・

tree-trellis 型音声認識プログラム

tree-trellis 型音声認識プログラムとは連続単語認識アルゴリズムとして tree-trellis サーチを用いたものである。tree-trellis サーチは、各時刻・各状態において最尤の単語列を知ることができ、N 位までの累積尤度の単語列 (N-best リスト) を出力することができる。各時刻・各状態において最尤の単語列を知ることができるので、音素モデルではポーズを認識しながら言語モデルではポーズをスキップすることによってポーズがある音声でも誤認識を起こしにくい。言語モデルには Left-Right 型のデータが使用できる。tree-trellis サーチのアルゴリズムを図 5 に示す。

tree-trellis サーチの trellis 計算は、前時刻の同一状態を意味するグリッドの尤度と前時刻の 1 つ前の状態のグリッドの尤度の 2 状態を加えて現時刻のグリッドの尤度を計算す

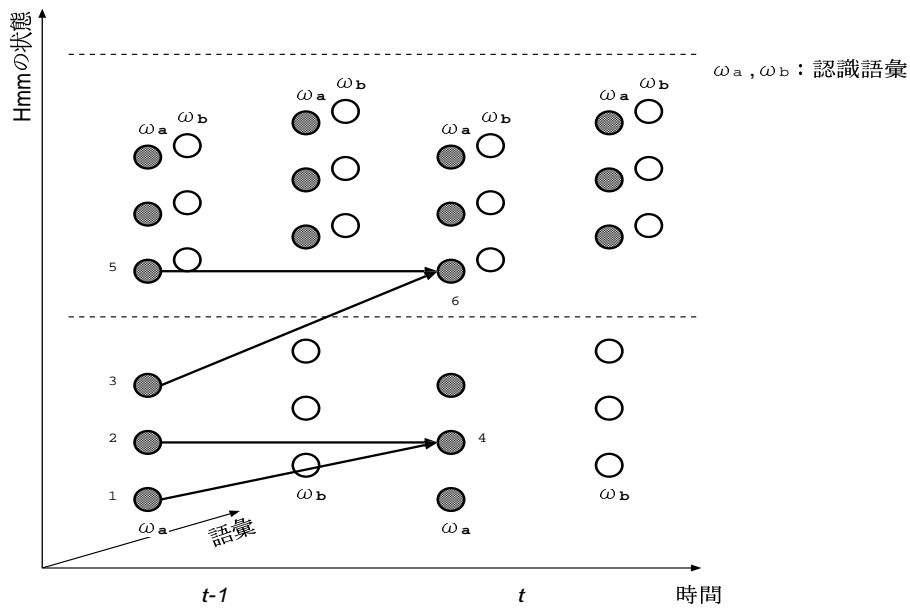


図 5: tree-trellis サーチのアルゴリズム (4 状態 3 ループ)

る。例えば、 4 は 1 からの遷移と 2 からの遷移の累積尤度の総和である。しかし、 6 のようなグリッドは、前時間の同一のグリッドの累積尤度 (5) とレベルが 1 つ下の単語の最終状態を意味するグリッド (3) の累積尤度の総和とする。この計算を全てのグリッドに対して行う。

4.3 評価方法

本研究は、結合価パターン適用前と適用後の累計正解率を比べ評価する

4.4 実験手順

実験の手順について、おおまかな手順を図6に示す。

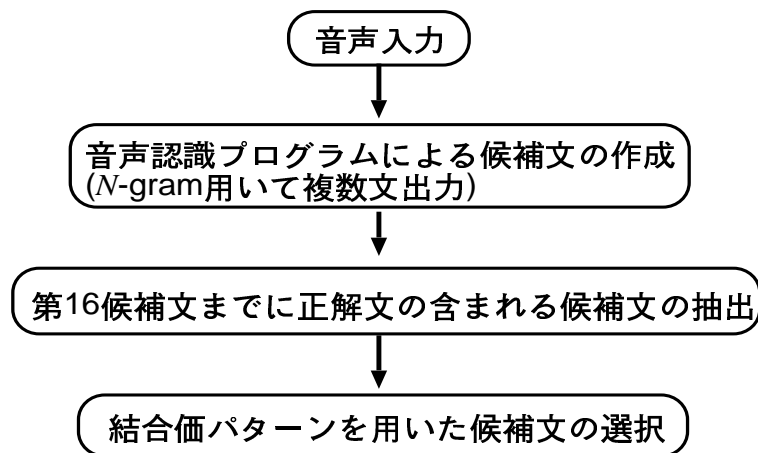


図 6: 実験手順

入力文に対して音声認識プログラムを用いて候補文を生成する。候補文は N -gram を用いて複数文出力され、尤度順に第 16 位までとする。第 16 位までに正解文の出力された候補文を実験データとして用いる。得られた候補文に対して結合価パターンを用いて有効性を調べる。

5 実験結果と考察

5.1 実験結果 (1)

実験データとして単文 100 文を用いた。

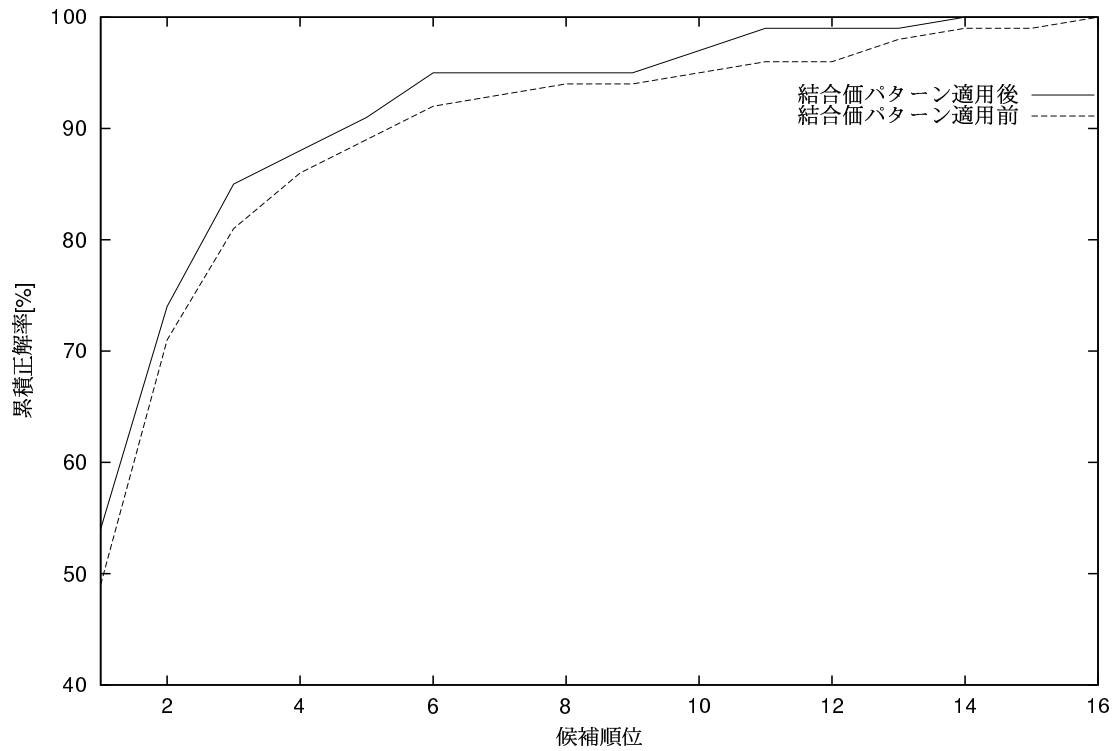


図 7: 実験結果

表 5: 各候補までの累積正解率

候補	結合価パターン無	結合価パターン有
第 1 候補	49% (49/100)	54% (54/100)
第 2 候補	71% (71/100)	74% (74/100)
第 4 候補	86% (86/100)	88% (88/100)
第 8 候補	94% (94/100)	95% (95/100)

第 1 候補では結合価パターン適用前は 49 文、適用後は 54 文であった。

5.2 出力結果例

5.2.1 実験データ例

入力文 (2) の音声認識プログラムの出力結果 (実験データ) を以下に示す。

< 音声入力文 > 私はそのニュースに驚いた (2)

```
1 ( 358.38794 ) 私はそのニュースに驚いた
0 134 5 42 1466 6 3416 9
2 ( 409.09436 ) う私はそのニュースに驚いた
1867 134 5 42 1466 6 3416 9
3 ( 422.99939 ) 私はそのニュースには驚いた
0 134 5 42 1466 6 5 3416 9
4 ( 424.80951 ) 私はそのニュースに驚いた
0 134 5 42 1466 6 3416 0 9
5 ( 425.14111 ) 私はその店に驚いた
0 134 5 42 1377 6 3416 9
6 ( 451.33582 ) 私はそのニュースに誇る一帯
0 134 5 42 1466 6 7192 6895
7 ( 461.32666 ) 私はそのニュースに驚いた区
0 134 5 42 1466 6 3416 11 4019
8 ( 461.65387 ) 私はそのニュースに行きたい
0 134 5 42 1466 6 1225 86
9 ( 461.68341 ) 私はそのニュースに届いた
0 134 5 42 1466 6 4263 9
10 ( 462.76263 ) 私はそのニュースにと驚いた
0 134 5 42 1466 6 8 3416 9
11 ( 468.14755 ) 私はソニーはせにほど来た
0 134 5 6271 5 2095 1993 272 634 9
12 ( 470.87003 ) 私はソニーはずにほど来た
0 134 5 6271 5 1312 1993 272 634 9
13 ( 471.79947 ) 私はそのニュースに音を行った
0 134 5 42 1466 6 1517 4 238 9
14 ( 471.99435 ) 私はそのニュースに心を行った
0 134 5 42 1466 6 647 4 238 9
15 ( 475.51596 ) う私はそのニュースに驚いた
1867 134 5 42 1466 6 3416 0 9
16 ( 482.80804 ) 私はそのニュースに横を行った
0 134 5 42 1466 6 2604 4 238 9
```

各偶数行の数字は単語のインデックスであり「134 5」が「私 は」のように対応している。この文では第1候補に正解文が存在する。この文で用いられる用言「驚く」に「N1(3主体 535動物)が N2(*)に 驚く」のパターンがあるので、この第1候補にある正解文は結合価パターンに適している。

5.2.2 候補順位向上例

出力結果の中で結合価パターンにより候補が向上した文を以下に記載する。

<音声入力文> 母が私に新しい本をくれた (3)

音声認識プログラムに文(3)を入力した場合、以下の4候補が出力された。正解文は第4候補である。

第1候補文 歯が私に新しいのんをくれた

第2候補文 歯が私のに新しいほど遅れた

第3候補文 母が私に新しい方をくれた

第4候補文 母が私に新しい本をくれた

第1、3、4候補は用言「くれる」のパターン①が存在、第2候補は用言「遅れる」パターンが存在しない。第2候補文は削除され正解文である第4候補文が第3候補に昇格した。

以下に用言「くれる」、「遅れる」のパターンを示す。(括弧内は意味属性)

表 6: 用言「くれる」の結合価パターン

①N1(3主体) [が/から] N2(388場所 533具体物 1001抽象物) を N3(3主体) [に/へ] くれる
②N1(4人) が N2(4人) [に/へ] N3("ゲンコツ") を くれる

表 7: 用言「遅れる」の結合価パターン

①N1(4人) が N2(374企業 367公共機関 986乗り物 1236人間活動) に 遅れる
②N1(986乗り物 1235事) が 遅れる
③N1(1000抽象) が N2(2585数量 2670時間) 遅れる
④N1(362組織) は N2(1001抽象物 1236人間活動) が 遅れる
⑤N1(3主体 1001抽象物 1236人間活動) が N2(2302進歩 1218流行 2518形勢 2697期間(歴史)) に 遅れる
⑥N1(984時計 2152流布 2302進歩 2394成長) が 遅れる
⑦N1(362組織) は N2(1001抽象物 1236人間活動) が N3(3主体) [から/より] 遅れる
⑧N1("文化" 1002抽象物(精神) 1155制度 458地域) が 遅れる
⑨N1(984時計) が N2(2585数量 2670時間) 遅れる

6 考察

6.1 候補昇格文の調査

候補順位が昇格した文は 100 文中 11 文であった。候補順位が昇格した 11 文を調査した。

11 文中 10 文は以下の出力例文 (4) の第 1 候補のように文末に名詞が来ることで削除されたものであった。

< 音声入力文 >	彼女は英語を始めた	(4)
第 1 候補文	彼女は英国は <u>自宅</u>	
第 2 候補文	彼女は A 子を始めた	
第 3 候補文	彼女は絵を始めた	
第 4 候補文	彼女は猫を始めた	
第 5 候補文	彼女は A 子を始めた	
第 6 候補文	彼女は円を始めた	
第 7 候補文	彼女は英国家事を得た	
第 8 候補文	彼女は英語を始めた	

このように、文末に用言ではなく名詞が来てしまう場合があった。これは認識プログラムの言語モデルに単語 bi-gram を用いているので、名詞で終る文章が出力された。このような候補は結合価パターンを用いなくても削除できてしまう。今回は単文を用いたのでこのような結果が多数見られたのだと考えられる。

6.2 結合価パターン適用文の調査

結合価パターン適用した文の中には以下のような意味が不明な語を含む文が存在した。

出力候補文 歯が私に新しいのんをくれた

この文の「のん」は意味的に不明な語句である。しかし、結合価パターンはこのような意味的に不明な未知語が存在していてもパターンが適用してしまっている。

6.3 結合価パターンの問題点

今回の実験データでは結合価パターンに適した単文を用いた。これは、結合価パターンでカバーしきれていないパターンが存在する場合があります、入力文や候補文が統語的にあっているにもかかわらず削除してしまう場合があるのを防ぐためである。

7 結合価パターン適用規則の見直し

次に、結合価パターンの適用規則の見直しを行った。条件は以下の通り。

- ・ 結合価パターンで全ての格要素が存在するものを結合価パターンに適しているとする。(ただし(*)は省略されてもよい)

7.1 実験用データ

実験用のデータには本実験データの中から、規則見直しの条件に適しているものを用いた(実験データ 100 文中 70 文(付録の結合価パターン適用規則見直し実験データ参照))。

実験結果

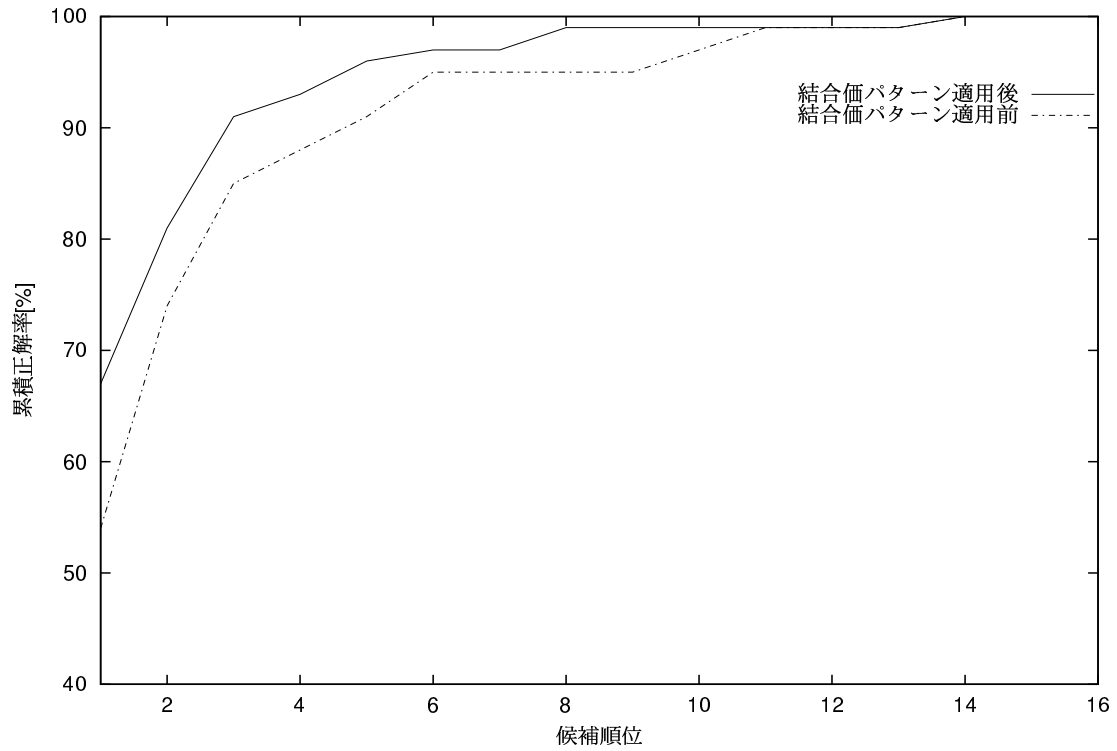


図 8: 実験結果

表 8: 各候補までの累積正解率

候補	結合価パターン無	結合価パターン有
第 1 候補	50% (35/70)	67% (47/70)
第 2 候補	71% (50/70)	81% (57/70)
第 4 候補	84% (59/70)	93% (65/70)
第 8 候補	94% (66/70)	99% (69/70)

7.2 候補順位向上例

出力結果の中で結合価パターンにより候補が向上した文を以下に記載する。(全文は追加実験付録参照)

<音声入力文> 母が私に新しい本をくれた (1)

音声認識プログラムに文(1)を入力した場合、以下の4候補が出力された。正解文は第4候補である。

- 第1候補文 歯が私に新しいのんをくれた
- 第2候補文 歯が私のに新しいほど遅れた
- 第3候補文 母が私に新しい方をくれた
- 第4候補文 母が私に新しい本をくれた

第3、4候補は用言「くれる」のパターン①が存在、第1候補は存在しない。第2候補は用言「遅れる」パターンが存在しない。第1、2候補文は削除され正解文である第4候補文が第2候補に昇格した。(格要素を全て一致させることにより、第3候補までしか上昇しなかったものが第2候補まで昇格した)

以下に用言「くれる」、「遅れる」のパターンを示す。(括弧内は意味属性)

表 9: 用言「くれる」の結合価パターン

①N1(3主体) [が/から] N2(388 場所 533 具体物 1001 抽象物) を N3(3主体) [に/へ] くれる
②N1(4人) が N2(4人) [に/へ] N3(“ゲンコツ”) を くれる

表 10: 用言「遅れる」の結合価パターン

①N1(4人) が N2(374 企業 367 公共機関 986 乗り物 1236 人間活動) に 遅れる
②N1(986 乗り物 1235 事) が 遅れる
③N1(1000 抽象) が N2(2585 数量 2670 時間) 遅れる
④N1(362 組織) は N2(1001 抽象物 1236 人間活動) が 遅れる
⑤N1(3主体 1001 抽象物 1236 人間活動) が N2(2302 進歩 1218 流行 2518 形勢 2697 期間(歴史)) に 遅れる
⑥N1(984 時計 2152 流布 2302 進歩 2394 成長) が 遅れる
⑦N1(362 組織) は N2(1001 抽象物 1236 人間活動) が N3(3主体) [から/より] 遅れる
⑧N1(“文化” 1002 抽象物(精神) 1155 制度 458 地域) が 遅れる
⑨N1(984 時計) が N2(2585 数量 2670 時間) 遅れる

7.3 考察

候補順位が昇格した文は 70 文中 23 文であった。候補順位が昇格した 23 文を調査した。結合価パターンのみで削除できたものは 23 文中 18 文、名詞が文末に来ることで候補順位が向上した文は 8 文、両方で削除され候補が向上したものが 3 文あった。第 1 候補文は 35 文から 47 文に増加した。

8 結論

本実験では単文認識の精度向上に結合価パターンがどの程度有効であるかを調査した。結果として格要素が1つ以上存在すればパターンに適しているとする規則(①)では100文中11文において候補順位が向上した。全ての格要素が存在する場合のみパターンに適しているとする規則(②)では70文中23文において候補順位が向上した。また、第1候補文は①では49文から54文に、②では35文から47文に増加し、有効性が見られた。しかし、出力結果を調査すると、①では文末が名詞であることで候補削除される文がほとんどであり、用言や格要素が不適切で結合価パターンに適せずに削除できたものは11文中2文であった、②では23文中8文であった。また、結合価パターンに適した候補文の中には未知語でパターンがあたるものが存在した。未知語を考慮して結合価パターンを適用することにより、効果を向上させることが可能であると考えられる。今回の実験データは結合価パターンに適した単文を用いた。しかし、結合価パターンは全てのパターンを網羅していない。用言に対するパターンがない場合は正解文であっても削除してしまう場合があると考えられる。

9 まとめ

本研究では、統語的な情報である結合価パターンが認識精度向上にどの程度役立つかを検証した。実験の結果、パターン適用規則「1つ以上の格要素が存在する」で単文の認識精度向上は見られるが、結合価パターンを用いずとも別フィルターをもちいることでも大きな精度向上がみられる。そこで、パターン適用規則を見直し「全ての格要素が存在する」では結合価パターンによる認識精度向上が見られた。今後の課題は、今回実験データとして単文100文を用いたが複文重文の場合での結合価パターンの有効性の調査、意味的に不明な語句が出力文の中に出て来たときの未知語処理や結合価パターンの被覆問題の考慮が必要である。

謝辞

本研究において1年間指導して頂いた鳥取大学知能情報工学科計算機C研究室の池原悟教授, 村上 仁一助教授と徳久 雅人助手にお礼を申し上げます。ALT-JAWSはNTTとの共同研究の下で使用させて頂きました。そして、1年間卒業研究をするにあたりお世話になった研究室の方々にお礼申し上げます。

参考文献

- [1] 池原、宮崎、白井、他：日本語語彙体系、岩波書店 (1997)
- [2] Hidden Markov Model Toolkit(HTK)
<http://htk.eng.cam.ac.uk>
- [3] 毎日新聞 (1993.1 ~ 1993.12)：毎日新聞社