

概要

同音異義語の仮名漢字変換での誤変換という問題に対し、単語間の意味的な関係を考慮した手法が提案されてきた。意味的情報の網羅性がそれらの手法を実現する上での問題であったが、近年、網羅性の高い意味的辞書として結合価パターンが作成された。

そこで、本研究では、単語連鎖確率を用いた仮名漢字変換の候補文に対して結合価パターンを用いた候補選択を行う手法を提案する。具体的には、単語連鎖確率を用いて出現確率の高い順に候補文を出力し、出力候補文に対して結合価パターンで意味的整合性を調査し、その有効性を検証する。

結果として、結合価パターンによって仮名漢字変換候補文を1位まで引き上げた割合は、動詞の同音異義語を含む試験文において14%、名詞の同音異義語を含む試験文において16%、実際の文において12%であった。よって、結合価パターンの仮名漢字変換に対する有効性が示された。

目次

1	はじめに	4
2	関連研究	5
2.1	単語連鎖確率を用いた仮名漢字変換	5
2.2	結合価パターンによる意味解析	6
2.2.1	結合価パターンについて	6
2.2.2	一般名詞意味属性体系	7
2.3	結合価パターンを用いた音声認識	8
3	仮名漢字変換アルゴリズム	9
3.1	アルゴリズムの概要	9
3.2	単語連鎖確率を用いた変換候補作成	10
3.2.1	単語辞書の作成	10
3.2.2	接続辞書の作成	10
3.3	文法的に不適切な文の削除	11
3.4	意味的に不適切な文の削除	12
3.4.1	削除条件	12
3.4.2	意味的に不適切な文の判定	12
3.4.3	任意格ルールベース	14
3.5	各手順の具体例	15
3.6	慣用表現の優先	17
3.6.1	原理	17
3.6.2	具体例	17
4	実装	19
4.1	仮名漢字変換システムの概要	19
4.2	変換候補出力部の実装	20
4.2.1	単語辞書	20
4.2.2	接続辞書	21
4.2.3	単語切り出し・出現確率計算	22
4.3	パターンマッチング部の実装	23
4.3.1	格変化	24
4.3.2	結合価パターンマッチング	24
4.3.3	任意格チェック	25

4.3.4	名詞句チェック	25
4.3.5	慣用表現チェック	25
5	評価実験	26
5.1	評価実験の目的と方法	26
5.2	評価基準	26
5.3	評価過程	27
5.3.1	正解の判定例	27
5.3.2	不正解の判定例	28
5.4	実験結果	30
5.4.1	実験1の結果	30
5.4.2	実験2の結果	32
5.4.3	実験3の結果	36
6	考察	38
6.1	誤り文が削除できないことについて	38
6.2	正解文が削除されることについて	39
7	おわりに	42

1 はじめに

現在，仮名漢字変換において高い変換精度を持つ手法として，単語連鎖確率を用いた手法 [1],[2] がある．しかし，単語間の局所的な関係しか考慮しておらず，同音異義語の仮名漢字変換での誤変換が問題となる．

これに対し，文の意味解析を用いた仮名漢字変換の手法を考える．過去の研究では，連語共起情報を意味素で記述しその関係を調べる方法 [3]，単文内で用言の格フレームを用い，意味的整合性から変換する方法 [4] などが提案されている．ただし，文献 [3]，[4] の手法では，意味素の体系を適切に設定することが必要であり，加えて，文献 [4] の手法では，一貫性のある格フレームを大規模に構築しなければならないことが問題となっている．しかし，近年，網羅性の高い意味的辞書として結合価パターン [5] が作成されたことにより，意味解析を用いる手法が現実味を帯びてきた．

過去の研究において，結合価パターンを用いた仮名漢字変換の手法 [6] が提案されている．この手法においては，実験対象に，名詞，または，動詞以外の漢字表記が全て決定している場合において，結合価パターンを用いて同音異義語の選択を行っている．結果として，結合価パターンを用いた同音異義語の選択において，特に同音異義語の動詞に対して効果があることがわかっている．

そこで，本研究では，結合価パターンを用い，入力文が全て平仮名文の場合に，結合価パターンで仮名漢字変換候補文の意味的整合性の判定を行うことで，正しい候補文を取得する手法を実現，その有効性を調査する．

本稿では，第 2 章で仮名漢字変換アルゴリズムについて述べ，第 3 章では評価実験の結果を述べる．そして，第 3 章の結果を踏まえ，第 4 章で結果の考察を行う．

2 関連研究

2.1 単語連鎖確率を用いた仮名漢字変換

文献 [2] より，単語連鎖確率を用いた仮名漢字変換について説明する．

まず，入力された平仮名ベタ書き文に対し，単語辞書を利用して単語候補を出力する．次に，出力された単語候補の中から出現確率が最大となる漢字仮名混じり文を選択し，仮名漢字変換結果とする．ここで，出現確率を求めるには，文の区切りごとに出現確率を計算していく．文献 [2] では単語ごとの区切りを使用している．

単語区切りの場合，意味的にまとまった単位であるため，仮名漢字変換精度は，別の切り出し手法と比較して高いとされる．

よって，本研究では，単語区切りによる連鎖確率計算を行う．具体的にはまず，入力文を単語列に分割する．単語列に分割した後，連鎖確率を求め，連鎖確率の積で出現確率を計算する．連鎖確率の計算には単語 bi-gram モデルを用いる．

例えば「わたしはたなかです」という文の場合，

/わたし/は/たなか/です/

/わたし/はた/なか/です/

/わたし/はたなか/です/

というように単語列ごとの分割を行う．そして，各単語区切りごとの単語の組み合わせを行う．組み合わせの一例を図 1 に示す．

入力文：わたしはたなかです

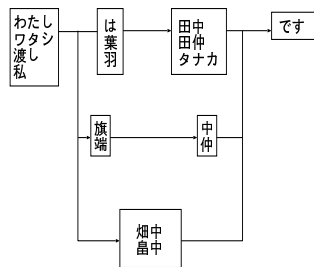


図 1: 組み合わせの一例

そして、単語列で隣接する単語間の連鎖確率を計算し、その組み合わせから出現確率を求める。「私/は/田中/です」という組み合わせの場合の出現確率は以下ようになる。

$$P\{\text{“私は田中です”}\} = P(\text{“私”} | \text{start}) \times P(\text{“は”} | \text{“私”}) \times P(\text{“田中”} | \text{“は”}) \times P(\text{“です”} | \text{“田中”}) \times P(\text{end} | \text{“です”})$$

ここで $P(\text{“私”} | \text{start})$ は、文の最初に「私」が出現する確率を表し、 $P(\text{end} | \text{“です”})$ は、文末に「です」が出現する確率を表す。

2.2 結合価パターンによる意味解析

2.2.1 結合価パターンについて

結合価パターンは、用言と格要素(名詞+格助詞)の意味的關係を記述したもので、用言と格要素間に意味的な制約を発生させて候補文の解析を行う。

本研究では、日本語語彙大系 [5] にある「構文意味辞書」の結合価パターンを基本にする。結合価パターンは約 16,000 件の日本語文型パターンを収録している。結合価パターンの例を示す。

パターン例 1: $N1(2483 \text{ 性質})$ が大きい

パターン例 2: $N1(4 \text{ 人})$ が若返る。

ここで、 $N1$ の後についた () 内の番号と単語は名詞の持つ意味属性番号と意味属性である。

2.2.2 一般名詞意味属性体系

本研究では、結合価パターンの名詞の意味属性を調べるために、一般名詞意味属性体系(図2)を用いる。一般名詞意味属性体系とは、単語の意味により体系的に分類、整理した語彙集である。日本語語彙大系に登録してある結合価パターンにある名詞は、一般名詞意味属性体系によって意味属性(図1中の#1~1964下の単語。#の数字は意味属性番号)を付けられている。

本研究で使用する一般名詞意味属性体系は、約40万語の名詞を、最大12段(図中の*1~12。名詞の属性の深さを表す)の木構造を構成する2,710の意味属性に分類している。また、一般名詞意味属性体系は、木構造を基本構成とし、上位の属性を持つ名詞は、木構造において自分より下位の名詞の意味属性を包含する性質がある。上位の意味属性が下位の意味属性を包含する事と名詞の持つ意味属性を利用し、結合価パターンは動詞の持つ意味を規定する事ができる。

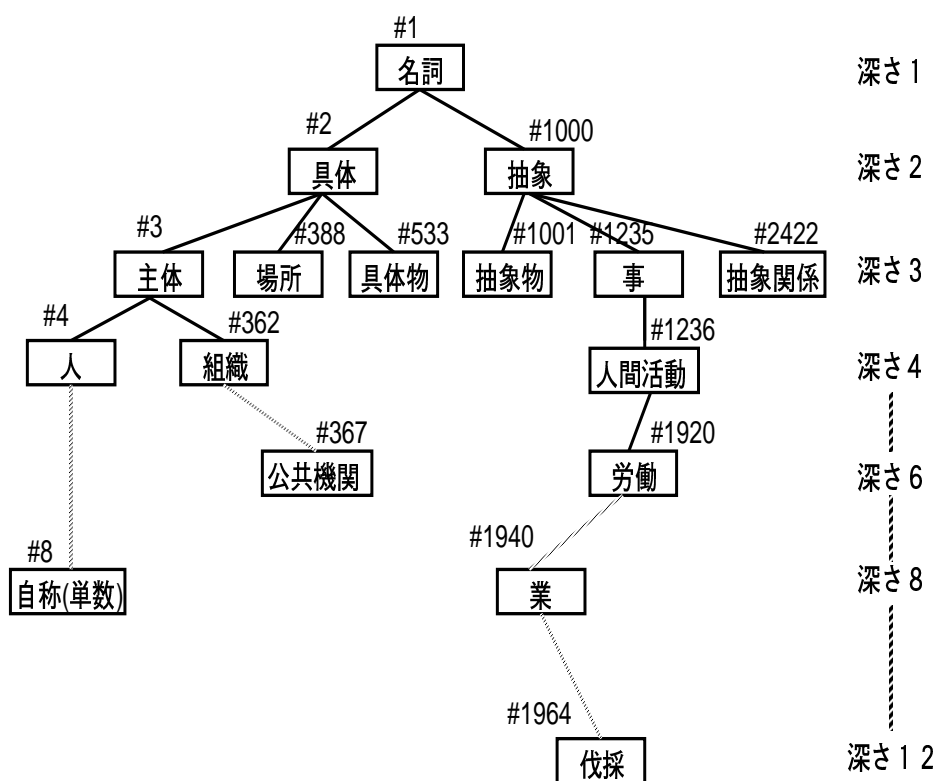


図2: 一般名詞意味属性体系の一部

2.3 結合価パターンを用いた音声認識

過去の研究において，音声入力文の候補文に対し，結合価パターンによる意味解析を用いて候補選択を行う手法 [7] が提案されている．この手法による結合価パターンの使用手法としては，結合価パターンで指定されている格要素全てが適合しない限り意味的に正しいとみなさないとしている．以下に具体例を示す．

音声入力文：私は彼の高校を受験する．

この音声入力文に対し，言語モデルに単語 bi-gram を用いた音声認識プログラムにおいて，以下の 2 文が出力された．

第 1 位候補文 私からの高校と受験する．

第 2 位候補文 私は彼の高校を受験する．

ここで「受験する」のパターンは以下の通りである．

パターン 1： $N1(4 \text{ 人})$ [は/が] $N2(362 \text{ 組織})$ を受験する

パターン 2： $N1(4 \text{ 人})$ [は/が] $N2(1426 \text{ 試験})$ を受験する

パターン 3： $N1(4 \text{ 人})$ [は/が] 受験する

上記のパターン内で，第 1 位候補文を満たすパターンは存在しないので，第 1 位候補文は非文と判断し削除する．

第 2 位候補文については，パターン 1 において「 $N1(4 \text{ 人})$ は」の部分が「私は」と「 $N2(362 \text{ 組織})$ を」の部分が「高校を」と合致するので結合価パターンの条件を全て満たしている．したがって，第 2 位候補文を第 1 位候補文とする．

以上の結合価パターンの候補文選択方式により，単文を音声入力することで出力された，出現確率の高い上位 16 候補文に対して実験を行ったところ，候補文の第 1 位正解率が 17% 向上し，結合価パターンは認識精度向上に有効であることが示されている．しかし，実際の文には省略が生じるために結合価パターンの指定する全格要素が適合する場合は限定的なものである．したがって，本研究では，文の格要素の省略にも対応した手法を提案する．

3 仮名漢字変換アルゴリズム

3.1 アルゴリズムの概要

結合価パターンを使用する際は，前もって入力文に対して形態素解析を行い，意味属性や単語区切りの情報を取得しておく必要がある．しかし，本研究で用いるような平仮名文に対して形態素解析をすることは困難である．

また，平仮名文を形態素解析することができたとしても，結合価パターンで同音異義語の候補の解析ができるのは，用言と名詞の部分に限られる．したがって，結合価パターンのみでは平仮名文の全体に対して仮名漢字変換を行うことができない．

よって，本研究では，あらかじめ候補文を出力し，その候補文に対し，結合価パターンで解析を行うことにした．そこで用いるのが，単語連鎖確率を用いた仮名漢字変換である．単語連鎖確率を用いた仮名漢字変換は，確からしい順に変換候補を大量に出力することができる．変換候補に対し，結合価パターンを用いることで，意味的な適切性を判定することができる．

よって，両者の利点を活かし，統合したアルゴリズムを作成する．アルゴリズムの概要を以下に示す．

(手順 1：候補作成)

(手順 2：文法的に不適切な文の削除)

(手順 3：意味的に不適切な文の削除)

(手順 4：慣用表現の優先)

(手順 5：補欠選択)

ここで，補欠選択とは，(手順 2) および (手順 3) で全ての候補が削除された場合に，出現確率の最も高い候補文を選択することである．以下に手順の詳細を示す．

3.2 単語連鎖確率を用いた変換候補作成

(手順1)では，文献[2]の単語連鎖確率を用いた仮名漢字変換を行う．出現確率の高い文から順に，上位32文の変換候補を作成する．

「わたしはたなかです」と入力した場合の一例

- 1：私は田中です
- 2：私は田仲です
- 3：私はタナカです
- 4：わたしは田中です
- 5：わたしは田仲です
- ...
- 32：渡しは棚か出す

変換候補作成時には，元となる単語を検索するために単語辞書が必要となる．また，出現確率を計算するために，単語と単語が繋がる確率を記した接続辞書が必要である．よって，日本語語彙大系から約35万語の単語辞書を，EDRコーパス約20万文から接続辞書をそれぞれ作成した．

3.2.1 単語辞書の作成

単語辞書は，日本語語彙体系に登録されている，漢字用語辞書，漢字一文字辞書，カタカナ語辞書，ひらがな語辞書，混ぜ書き語辞書を統合した辞書を作成した．辞書は「あ～ん」のインデックス別に構成し，読みと漢字の字面を記載している．

3.2.2 接続辞書の作成

本研究では接続辞書の約800万文字の漢字仮名混じり文を含んだEDRコーパスを使用している．具体的な単語連鎖確率の計算としては，EDRコーパスは文が形態素区切りで分割されているので，その形態素ごとに漢字仮名の出現頻度を数え，最後にこれらの連鎖確率を計算した．

3.3 文法的に不適切な文の削除

(手順2)では、出力した32文の候補文に対し、ALT-JAWSを用いた形態素解析を行う。ALT-JAWSは品詞の接続関係に誤りのある部分をエラーとして出力することができる。したがって、ALT-JAWSを用いて形態素解析を行うことで、品詞間の接続関係に誤りのある候補文を削除することができる。そして、文法的に不適切な文を発見し、結合価パターンに必要な文法的な文の区切り、名詞意味属性等を導き出す。文を形態素解析した例を以下に示す。

(正常に出力した場合)

入力候補文：彼は論文を日本語から英語に直す。

1. 彼 (1710,[23,48])/は (7530)
2. 論文 (1100,[1046,1111])/を (7430)
3. 日本語 (1a00,[1064])/から (7410)
4. 英語 (1a00,[1064])/に (7430)
5. 直す (2136)/。 ([P]0110)

この場合では、正しい文が入力されているので、エラーは出力されない。

(エラーを出力した場合)

入力候補文：彼は論文を日本語絡英語に直す。

1. 彼 (1710,[23,48])/は (7530)
2. 論文 (1100,[1046,1111])/を (7430)
3. 日本語 (1a00,[1064])/絡 ([*]1100,[1])/英語 (1a00,[1064])/に (7430)
4. 直す (2136)/。 ([P]0110)

この場合では、「彼は論文を日本語から英語に直す」という文の誤り文であるが、格助詞「から」の部分が用言等で使用される「絡」になっている。「絡」の後には活用形がかからなくてはならないが、この場合、直後に名詞の「英語」が来ているため、接続関係がおかしくなる。したがって、その部分は文法的に誤りがあると見なさ

れ，エラーとして [*] が出力されている．よってこの候補文は削除される．

3.4 意味的に不適切な文の削除

3.4.1 削除条件

結合価パターンを使用する上で，候補を採用する条件を以下のように設定する．

候補を採用する条件： 候補文の全ての格要素と用言の間の意味的結束性が正しい

一部でも格要素の結束性が確認できない候補文は候補から削除する．ところで，実際の文には結合価パターンの制約を受けない任意格が存在するため，結合価パターンだけでは，全ての格要素の結束性が確認できない．さらに，本研究の現状では，格要素を構成する名詞が名詞句である場合，名詞句の主名詞にしか一般名詞意味属性による制約が働かない．そこで，本研究では，任意格を判定する為に，任意格のルールベースを用いる．名詞句の問題については，単語辞書，および，接続辞書を用いて，確認をとる．

任意格のルールベースは，EDR コーパスから収集した例文の単文 100 文より任意格要素を集めて作成する．任意格要素を判定するルールは，名詞部分を一般名詞意味属性で制約し，かつ，使用可能な格助詞を指定する．

3.4.2 意味的に不適切な文の判定

以下の段階を経て意味的に不適切な文を候補から削除する．

(手順 3-1．態による格助詞の変換)

結合価パターンは能動態で定義される．受身や使役の文を結合価パターンに対応させるためには，単純に受身、使役文を結合価パターンに登録する方法や既存の結合価パターン自体を変形して入力文に対応させる方法等がある．本手法では，候補文が能動態でない場合，文献 [8] の格助詞変換規則に従い，パターンマッチングの為の格助詞を補足する．

(手順 3-2 . 結合価パターンのマッチング)

まず候補文の用言で結合価パターンを [5] より検索する . 次に結合価パターンの規定する格助詞と一致する格要素を候補文から全て抽出し , 格要素への意味属性制約を検査する . ここで , 格助詞が一致するにも関わらず , 意味属性制約が満たされないならば , その結合価パターンは適用しない . なお , 結合価パターンの規定する格助詞が候補文で必ずしも使われる必要はない . マッチする結合価パターンが見つからないならば , 候補文を削除する .

(手順 3-3 . 任意格の検査)

(手順 3-2) の段階で検査されなかった格要素は , 任意格ルールベースで検査する . 不適切な格要素が 1 つでもある場合 , 候補文を削除する .

(手順 3-4 . 名詞句の検査)

(手順 3-2) および (手順 3-3) の段階で検査された格要素が名詞句ならば , 名詞句を構成する単語間の接続関係を接続辞書により確認する . 接続辞書に接続のない名詞は , 未知の名詞句として , その候補文を削除する . 例えば「私の本」という名詞句の場合なら「私」と「の」の接続関係「の」と「本」の接続関係を調べ , 接続例が見つければ正しい名詞句とする . 接続例が見つからない場合 , その名詞句を含む候補文は削除する .

3.4.3 任意格ルールベース

任意格ルールベースに登録したルールを表 2 に示す .

表 1: 任意格ルールベース表

名詞意味属性	格助詞
(2585 数量)	も
(2610 場)	から
(2670 時間)	から
(3 人)	に
(1155 制度)	に
(1236 人間活動)	に
(2610 場)	に
(2670 時間)	に
(362 組織)	で
(388 場所)	で
(1236 人間活動)	で
(2054 事象)	で
(2422 抽象的關係)	で
(2507 状態)	で
(2610 場)	で
(2670 時間)	で

ただし , 特殊なルールとして「(533 具体物) と (533 具体物) を」という組み合わせがある .

この表は , EDR コーパスから抽出した例文 100 文中 , 任意格を含む 25 文より , 意味属性の汎化を行い作成した .

3.5 各手順の具体例

(手順 3-1)： 態による格助詞の変換の具体例

格変化規則の具体例と，格変化対応表を以下に示す。

例： 契約が交わされている。

「契約を交わす」と認識する。(「が」格 「を」格に変化)

表 2: 受身、使役の格変化対応表

受身、使役	が格	に格	を格	と格	から格	によって格
平叙文	を格	が格	を格	と格	が格	が格

ただし、「に」格についてはかかる名詞が主体となる場合は格変化せず「を」格についてはかかる名詞が主体となる場合のみ「が」格に格変化するものとする特別規則を設けている．具体的な適応範囲は(4人)以下と(534生物)以下の意味属性の名詞である．

また「は」格については「が」格と同じ扱いとする．

(手順 3-2)： 結合価パターンのマッチングの具体例

結合価パターンのマッチングにおける具体例を以下に示す．

正しいと判定する場合

例 1-1

入力文： $N1$ が $N2$ に行く。

パターン： $N1$ が $N2$ に行く。

(パターンの $N1$ および $N2$ の意味属性制約が満たされている場合)

例 1-2

入力文： $N1$ が $N2$ に行く。

パターン： $N1$ が $N2$ に $N3$ で行く。

(パターンの $N1$ および $N2$ の意味属性制約が満たされている場合)

「例 1-1」のように、入力文とパターンの格要素が全て一致する場合、および「例 1-2」のように、パターンの指定する格要素は余るものの、入力文の格要素が全てパターンと一致する場合は、正しい文であると判断する。

誤りであると判定する場合

例 2-1

入力文： $N1$ が $N2$ に行く。

パターン： $N1$ が $N2$ に行く。

(パターンの $N1$ もしくは $N2$ の意味属性制約が満たされていない場合)

例 2-2

入力文： $N1$ に行く。

パターン： $N1$ が逝く。

(用言自体が間違っている場合)

[例 2-1] のように、どれか 1 つでも意味属性に誤りがある場合、および、[例 2-2] のように、用言自体が異なる場合は誤りとする。

任意格ルールベースを用いた判断が必要な場合

例 3-1

入力文： $N2$ を $N3$ で買う。

パターン： $N1$ が $N2$ を買う。

(パターンの $N2$ の意味属性制約が満たされている場合)

例 3-2

入力文： $N1$ で買う。

パターン： $N2$ が $N3$ を買う。

[例 3-1] のように、結合価パターンで一部解析できない格要素がある場合、および、[例 3-2] のように用言パターンは存在するものの、格要素で一致するものが無い場合は、任意格ルールベースを使用する対象とする。

(手順 3-3)： 任意格ルールベースの作成例

1： まず、任意格のある文を用意する。

例文： やがて姉もサウジアラビアへ働きに行った。

対応パターン： $N1(4人)$ が $N2(388場所\ 2610場)$ へ行く

対応パターンで適合するのは「姉も(が)」と「サウジアラビアへ」の部分だけである。したがって、任意格部分は「働きに」の部分ということになる。

2： 任意格部分を抽出する

任意格： 「働き」 + 「に格」

3： 意味属性を汎化する

「働き」の意味属性は(2051 活動)である。これを、できる限り意味が合致するところまで汎化する。

汎化： (2051 活動)「意味属性の深さの段8」 (1236 人間活動)「意味属性の深さの段4」

よって(1236 人間活動) + 「に格」をルールベースに追加する。

3.6 慣用表現の優先

3.6.1 原理

結合価パターンには一般表現と慣用表現があり、一般表現は意味属性で、慣用表現は字面で、それぞれ格要素に制約を発生させる。

日常的な文では、慣用表現の読みと同じになるような漢字の使い方はしないのが普通であるので、慣用表現のパターンが当てはまる候補文は、優先的に選択することが妥当である。

(例)

一般表現： $N1(3主体)$ が $N2(4人)$ を絞る

慣用表現： $N1(4人)$ が 知恵 を絞る

3.6.2 具体例

例えば、上の例で言うと「ちえをしぼる」という入力文の出力に、以下のような候補が出力されたとする。

候補文1位： 千絵を絞る

候補文 2 位：千絵を搾る

候補文 3 位：知恵を絞る

候補文 4 位：知恵を搾る

これらの文の場合，全ての候補文が結合価パターンで正しいと判断されている．したがって，候補文 1 位の「千絵を絞る」が取得されることになるが，候補文 1 位や 2 位は，日常的な文としての使用頻度は少ない．よって，日常的に使われる，候補文 3 位の「知恵を絞る」を取得する．

4 実装

4.1 仮名漢字変換システムの概要

本システムは，第3章で示した仮名漢字変換アルゴリズムを元に作成している．仮名漢字変換システムの概要について，図3に示す．

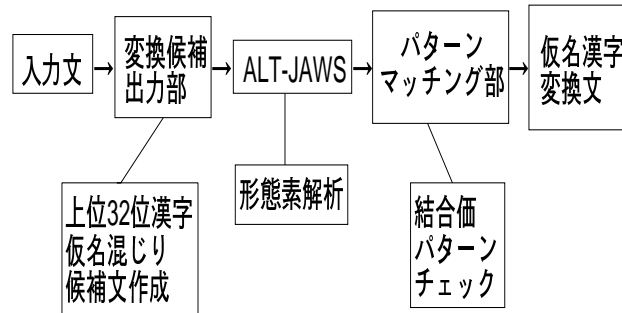


図 3: 仮名漢字変換システムの概要

各部について説明する．

「変換候補出力部」は，入力した平仮名文に対し，単語連鎖確率によって，出現確率の高い上位 32 位の候補文を出力する部分であり，仮名漢字変換アルゴリズムの(手順 1)の部分にあたる．

「ALT-JAWS」は出力候補文に対し，形態素解析をする部分であり，仮名漢字変換アルゴリズムの(手順 2)の部分にあたる．ALT-JAWSとは，NTTとの共同研究の下で使用している形態素解析ツールである．本研究では，この部分の実装には関わっていない．

「パターンマッチング部」は，形態素解析を行った候補文に対し，態による格変化，結合値パターンでのチェック，任意格ルールベースのチェック，名詞句部分のチェック，慣用表現の優先を行う部分であり，仮名漢字変換アルゴリズムの(手順 3~4)の部分にあたる．

なお，仮名漢字変換アルゴリズムの(手順 5)の部分については，上位 32 候補全文が削除された段階で行われる．

4.2 変換候補出力部の実装

変換候補出力部の概要を図4に示す。

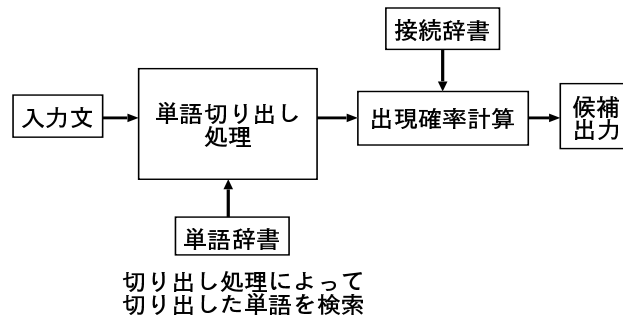


図 4: 変換候補出力部の概要

まず、文を単語単位に分割するため、単語切り出し処理を行う。切り出した部分が単語であるか、また、変換候補がどれだけあるかを調べるため、ここで、単語辞書を使用する。

次に、切り出した部分それぞれが連鎖する確率を求める。連鎖する確率は、あらかじめ集計し、接続辞書に登録しておく。この接続辞書を用いて、各連鎖確率を合計した値が候補文の出現確率となる。

次節より、各部分の実装の詳細について示す。

4.2.1 単語辞書

単語辞書作成には、日本語語彙体系の単語辞書を統合したものをを用いる。この時、重複がないようにソートユニークをかけておく。

「あ～ん」の順番にソート後、「あ～ん」のディレクトリを作成し、一文字目のインデックスとして、単語検索時に使用する。ただし、辞書の内容は膨大であるため、辞書を上から検索しては時間がかかる。したがって、単語の二文字目、三文字目も辞書の中にインデックス化する。

このとき、ディレクトリ内部には単語を収録したファイルと、インデックス用のファイルの2つを作成する。単語を収録したファイルには、単語の読みと字面の組み合わせを記載する。インデックス用のファイルには、単語のあるバイト位置を記載し、バイト位置を読み込む事により、メモリに辞書を読み込まずに指定の単語を取り出す。

4.2.2 接続辞書

接続辞書作成には、EDR コーパスを用いる。EDR コーパスの一部を示す。

例文：腕に入れ墨をした店員は、くやしがった。

- 1(腕/ウデ/名詞)2(に/ニ/助詞)
- 3(入れ墨/イレスミ/名詞)4(を/ヲ/助詞)
- 5(し/シ/動詞)6(た/タ/助動詞)
- 7(店員/テンイン/名詞)8(は/ハ/助詞)
- 9(、/、/記号)10(くやしが/クヤシガ/動詞)
- 11(っ/ツ/語尾)12(た/タ/助動詞)13(。/。)

このように、EDR コーパスは形態素単位で文に区切りを入れている。これを利用して、接続辞書を作成する。上記の例を用いると、

「腕」「に」「入れ墨」「を」「し」「た」「店員」「は」「、」「くやしが」「っ」「た」「。」の出現回数それぞれ一回が集計できる。

また、連鎖事例として「start/腕」「腕/に」「に/入れ墨」「入れ墨/を」「を/し」「し/た」「た/店員」「店員/は」「は/、」「、/くやしが」「くやしが/っ」「っ/た」「た/。」。「。/end」の事例をそれぞれ一回ずつ集計できる。ここで「start/腕」は「腕」で文が始まる場合、「。/end」は「。」で文が終わる場合を指す。

この集計を全文に行う。最後に、連鎖事例ごとにソートしてファイルにまとめ、各一文字目をインデックス化する。

集計例：CFD0A4C0 /腕/に/ 20 154

「CFD0A4C0」部分：

「CFD0」は「腕」の行の文字コードを指す。「A4C0」は「に」の行の文字コードを指す。この文字コードのインデックスを利用して、接続辞書の検索に使用する。

「/腕/に/」部分：

連鎖事例の字面部分である。これを用い、単語連鎖の字面の一致を調査する。

「20 154」部分：

「20」は「腕」の次に「に」が連鎖した事例回数で、「154」は「腕」の出現回数である。これにより「腕」の次に「に」が連鎖する確率を計算できる。この場合「腕」の次に「に」が連鎖する確率は、 0.13 ($P(\text{に}|\text{腕}) = \langle \text{「腕/に/」の出現回数} \rangle / \langle \text{「腕」の出現回数} \rangle = 20/154$) である。

4.2.3 単語切り出し・出現確率計算

次に、単語切り出し部分について説明する。

1：単語辞書からあらかじめ単語候補を抜き出す。

以下に例を示して説明する。

例文：わたしはたなかです。

「わ」のインデックス：輪，和，綿，私 …

「た」のインデックス：他，太，他紙 …

「し」のインデックス：氏，市 …

「は」のインデックス：齒，派，旗，畑中 …

これら抜き出した単語を検索に用いる。

2：文を分割する。

例文：わたしはたなかです。(10文字)

わたしはたなかです。(10:0)

わたしはたなかです/。(9:1)

わたしはたなかで/す。(8:2)

わたしはたなかで/す/。(8:1:1)

再帰を用いて区切りを行う。この分割の際、前から順に単語を検索する。この例の場合では、8文字の単語は存在しないので、これらの候補は全て消去され、3文字区切りとなった時「私」という単語が現れて初めて後ろの単語が検索される。このとき、単語同士のつながりを親子関係として、ポインタでつないでおく。

ただし，このアルゴリズムでは，区切りを保存しないため，区切り回数が多くなってしまふ．本研究では，実行速度については考慮していないが，改善の余地のある部分である．

3：全区切りについて単語が出揃った場合，出現確率を計算する．

出現確率の計算を，接続辞書から連鎖組み合わせを検索して行う．

このとき，出現確率の計算時に，各単語の連鎖確率の積を計算しているが，積の場合では数値が非常に大きくなってしまふので，対数の和で出現確率を求める．また，連鎖が無い組み合わせの場合は，接続辞書で最も連鎖しにくい組み合わせの確率， -12.5 を加算する．

この出現確率の計算を，各単語区切りの単語の組み合わせの数だけ行う．ただし，この時候補数が非常に多くなってしまふため，出力候補数は上位 32 候補に絞り込む．また，候補数を絞りこんだとしても，出力を区切りごとに行うために，やはり出力候補数は多くなってしまふ．そこで，出現確率が出た時点で，最も出現確率の高い数値を逐次記録しておき， -10 の範囲で出力のしきい値を作成しておく．そして，しきい値以下の出現確率を持つ候補文は出力しないこととする．

出力例 以下に，入力文を入れた際の出力例について示す．

まず，文を平仮名で入力する．

入力文：わたしがれつのさいごになる。

出力例としては以下のようなものがある．

出力例 1： -67.6457 /私/がれ/募/最期/荷菜/る/。

出力例 2： -51.9714 /渡士/が/列/の/最後/に/な/る/。

出力例 3： -33.5527 /私/が/列/の/最後/に/な/る/。

一番前についた数値が出現確率である．実際のところ，出現確率の高いものほど正しい文が出力されやすい．

4.3 パターンマッチング部の実装

パターンマッチング部の概要を図 5 に示す．

ここでは (手順 3~4) のアルゴリズムの手順に従って処理を行う．

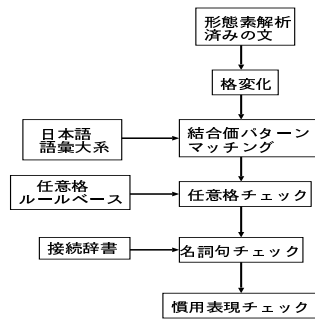


図 5: パターンマッチング部の概要

4.3.1 格変化

格変化を行うためには、候補文が受身形、もしくは使役形であることを確認しなくてはならない。本プログラムは、単文対応のプログラムであるので、用言が一番後ろに来ると仮定している。したがって、その用言の付属語の状態を判別することで、受身形、使役形の確認を取る。

形態素解析を行った時点で、用言の付属語の状態は品詞コードとして明記されている。よって受身形ならば (711X)(712X)、使役形ならば (713X)(714X)(715X) の品詞コードを持つ付属語の場合に格変化規則を適用する。

4.3.2 結合価パターンマッチング

結合価パターンマッチングを行うためには、まず、形態素解析済みの候補文から用言と格要素の情報を取得する。取得した用言より結合価パターンを検索し、候補文の格要素と一致するかを一つずつ調査する。適合が認められない文は削除する。

出力例について以下に示す。入力形態素解析文として「彼はステレオを付ける。」という文を入力する。

入力形態素解析文 彼はステレオを付ける。

1. 彼 (1710,[23,48])/は (7530)
2. ステレオ (1100,[923,972])/を (7430)
3. 付ける (2416,[1487,2162,2239,2273])/。 ([P]0110)

出力は以下のようになる。

入力文の字面：彼はステレオを付ける。

適合要素 1：N1 彼

適合要素 2：N2 ステレオ

適合パターン N1が N2を 付ける

適合パターンの意味属性対応 N1(4人) N2(968 電気機器 903 灯火)

この場合，N1，および，N2の意味属性が適合している．

4.3.3 任意格チェック

任意格のチェックについては，結合価パターンマッチングプログラムにある，意味属性の親子関係を調べるプログラムに，任意格部分の意味属性をあらかじめ登録することで作成した．順位の集計は人手で行っている．

4.3.4 名詞句チェック

名詞句チェックについては，接続辞書において，連鎖の事例があるかないかで行う．この事例の調査については，接続辞書を検索するプログラムを流用することで行ったが，順位の集計は人手で行っている．

4.3.5 慣用表現チェック

慣用表現チェックは，結合価パターンマッチングの段階で同時に行うことができる．慣用表現の場合は，字面での一致を調査するため，処理を別としている．したがって，慣用表現のパターンが適合した場合には，そのまま候補順位を 1 位へと引き上げる．

5 評価実験

5.1 評価実験の目的と方法

本研究では，以下の3つの実験を行う．

(実験1) 本手法の動作の確認，および，任意格のデータベース作成を目的とする．EDR コーパスの単文集を対象とする．

(実験2) 同音異義語に対する結合価パターンの効果の調査を目的とする．IPAL 動詞辞書，および，IPAL 名詞辞書に登録された，多義のある基本動詞，および，基本名詞から作成した単文集を対象とする．

(実験3) 実際の文に対する結合価パターンの効果の調査を目的とする．毎日新聞95年度記事 [9] の単文集を対象とする．

ただし，本稿では未知語が含まれる文を対象外とする．よって，単語辞書に含まれている単語で構成された例文100文を用いる．

5.2 評価基準

評価は「正解文との完全一致を正解とする基準」，および「人手で正解文と見比べて判定する基準」の2つを設ける．人手の判定を設けた理由は，仮名漢字の表記の正しさが絶対的には定まらないこと，および，語義選択が元々不可能である場合があることに対処するためである．例えば「復習に時間を掛ける」と「復讐に時間を掛ける」はいずれも意味的に正しいので正解とする．

5.3 評価過程

5.3.1 正解の判定例

(手順1)で求めた32候補文の中から，本手法によって正解文を引き出した例を下に示す．

例として，ここでは「確率を尋ねる。」の平仮名文を入力する．

入力文：かくりつをたずねる。

正解文：確率を尋ねる。

(手順1)部分

ここで，32位までの候補文が出力される．

候補文1位：確率を訪ねる。

候補文2位：確立を訪ねる。

候補文3位：確率を尋ねる。

...

候補文32位：格率をたずねる。

この場合，完全一致，および，人手での正解文は候補文3位である．

(手順2)部分

各候補文を形態素解析する．

確率を尋ねる。

1. 確率 (1100,[2596])/を (7430)
2. 訪ねる (2416,[1699])/。 ([P]0110)

確立を訪ねる。

1. 確立 (1230,[2074])/を (7430)
2. 訪ねる (2416,[1699])/。 ([P]0110)

確率を尋ねる。

1. 確率 (1100,[2596])/を (7430)
2. 尋ねる (2416,[1424,1427,1511])/。 ([P]0110)

(手順 3-4) 部分

エラーがないことが確認でき，これらの候補文は能動態なので，結合価パターンのマッチングを行う．

「訪ねる」の結合価パターン

パターン：N1(4人)がN2(388場所 2610場)を訪ねる

「尋ねる」の結合価パターン

パターン：N1(3主体)がN2(*)をN3(3主体)に尋ねる

候補文 1 位の場合「確率」の意味属性は (2596 計算値) なので「訪ねる」のパターン「N2(388 場所 2610 場)を」の部分の意味属性とは親子関係にはない．よって削除される．

候補文 2 位の場合「確立」の意味属性は (2074 確立) なので「訪ねる」のパターン「N2(388 場所 2610 場)を」の部分の意味属性とは親子関係にはない．よって削除される．

候補文 3 位の場合「確率」の意味属性は (2596 計算値) なので「尋ねる」のパターン「N2(*)を訪ねる」の部分の意味属性と親子関係にある．したがって，結合価パターンに合致したことになる．全ての格要素の結束性を確認できたので，この候補文を正しい文として取得する．よって，正解文が選ばれた．

5.3.2 不正解の判定例

(手順 1) で求めた 32 候補文の中から，本手法で不正解となった例を下に示す．

例として，ここでは「当時の人々は貝を取る。」の平仮名文を入力する．

入力文：とうじのひとびとはかいをとる。

正解文：当時の人々は貝を取る。

(手順 1) 部分

ここで，32 位までの候補文が出力される．

候補文 1 位：当時の人々は海を取る。

候補文 2 位：当時の人々は貝を取る。

...

候補文 32 位：当時の人々は階を取る。

この場合，完全一致，および，人手での正解文は候補文 2 位である。

(手順 2) 部分

各候補文を形態素解析に通す。

当時の人々は海を取る。

1. 当時 (1500,[2714,2700,2699])/の (7410)
2. 人々 (1100,[5,2606])/は (7530)
3. 海 (1100,[507])/を (7430)
4. 取る (2386)(2386, 捕る)(2386, 採る)(2386, 盗る)/。 ([P]0110)

当時の人々は貝を取る。

1. 当時 (1500,[2714,2700,2699])/の (7410)
2. 人々 (1100,[5,2606])/は (7530)
3. 貝 (1100,[545,842,631])/を (7430)
4. 取る (2386)(2386, 捕る)(2386, 採る)(2386, 盗る)/。 ([P]0110)

(手順 3-4) 部分

エラーがないことが確認でき，これらの候補文は能動態なので，結合価パターンのマッチングを行う。マッチングの際「は」格を「が」格として認識する。

「取る」の結合価パターン

パターン：N1(*) が N2(“学位” 594 手 *) を取る。

候補文第 1 位の場合，対応パターンの N1，N2，部分は何を入れても良いとされている。格助詞が一致するので，結合価パターンはこの候補文を正しい文として取得する。名詞句の接続確認も問題がなく，下位に慣用表現の候補文もなかったため，この文が 1 位となる。しかし「海」は「取る」ことができないので，この候補文は誤りである。

5.4 実験結果

実験結果はアルゴリズムの各段階での効果を比べるため，手順ごとに N-best で正解率を集計する．実験 1 から実験 3 の結果を表 3～表 6 に示す．各手順の上段は完全一致，下段は人手による判定の結果である．また，表 3～表 6 の結果をグラフにしたものを図 6～図 13 に示す．

5.4.1 実験 1 の結果

実験 1 集計結果を表 3, 図 6,7 に示す．

表 3: 実験 1 の結果 (クローズドテスト)

手順		候補数	1 位	~ 2 位	~ 4 位	~ 8 位	~ 16 位	~ 32 位
手順 1 のみ	完全一致		83 %	92 %	95 %	98 %	98 %	98 %
	人手判定		96 %	97 %	98 %	99 %	100 %	100 %
手順 2 まで	完全一致		78 %	87 %	91 %	93 %	93 %	93 %
	人手判定		95 %	96 %	97 %	98 %	99 %	99 %
手順 4 まで	完全一致		66 %	75 %	79 %	80 %	80 %	80 %
	人手判定		83 %	83 %	85 %	85 %	85 %	85 %
全手順	完全一致		79 %	89 %	93 %	94 %	94 %	94 %
	人手判定		96 %	97 %	99 %	99 %	99 %	99 %

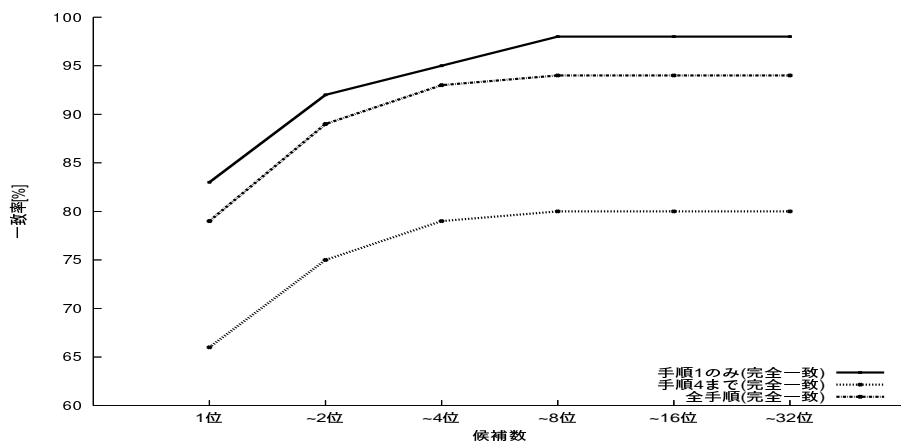


図 6: 実験 1 の結果 (クローズドテスト・完全一致)

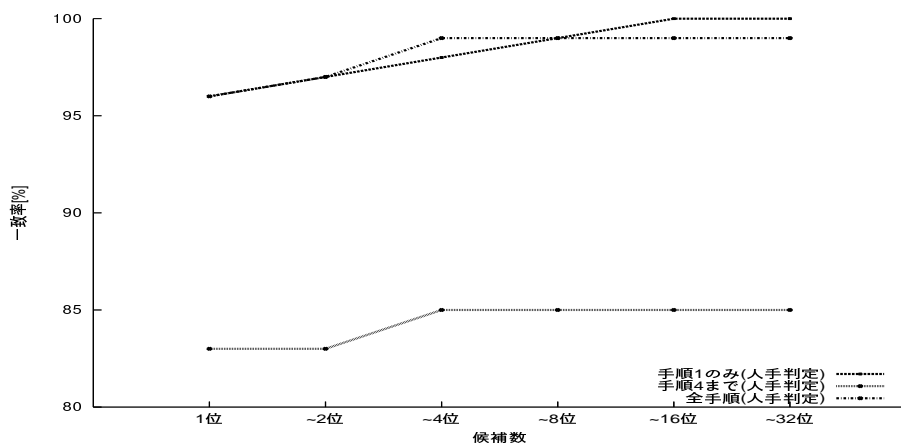


図 7: 実験 1 の結果 (クローズドテスト・人手判定)

「(手順 1)のみ」の結果より，(手順 1)の動作が正常に行われたことがわかる．また，図 4, 図 5 のグラフからも分かる通り，結合価パターンによって正解文が削除された割合が多い．しかし，クローズドテストでは，ほとんどの正解文が 1 位にあるため，結合価パターンによる順位引き上げは換算されていない．それを考慮に入れて，実験 2 に移る．

5.4.2 実験2の結果

実験2の同音異義語の動詞を含む文での結果を表4, 図8,9に示す.

表4: 実験2の結果 (同音異義語の動詞を含む文)

手順		候補数	1位	~2位	~4位	~8位	~16位	~32位
手順1のみ	完全一致		34%	46%	55%	61%	72%	73%
	人手判定		55%	68%	74%	81%	88%	88%
手順2まで	完全一致		35%	46%	56%	62%	70%	71%
	人手判定		54%	67%	74%	79%	84%	84%
手順4まで	完全一致		35%	42%	50%	51%	54%	54%
	人手判定		58%	63%	65%	66%	67%	67%
全手順	完全一致		44%	52%	62%	63%	68%	68%
	人手判定		67%	74%	78%	80%	83%	83%

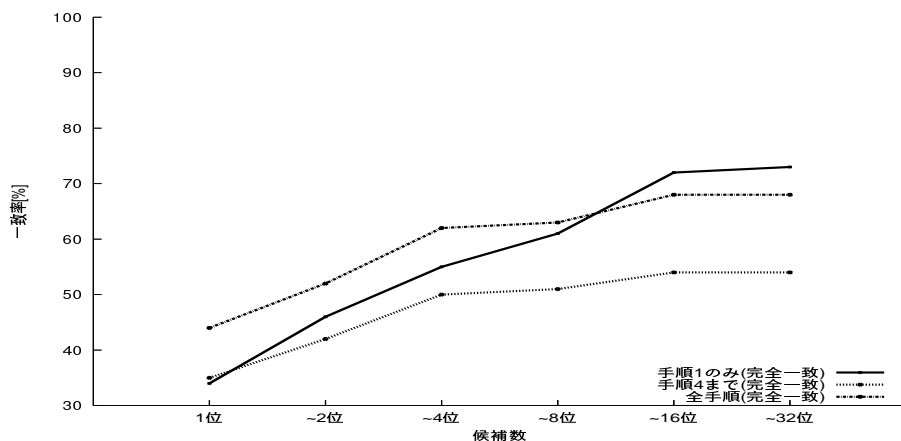


図 8: 実験 2 の結果 (同音異義語の動詞を含む文・完全一致)

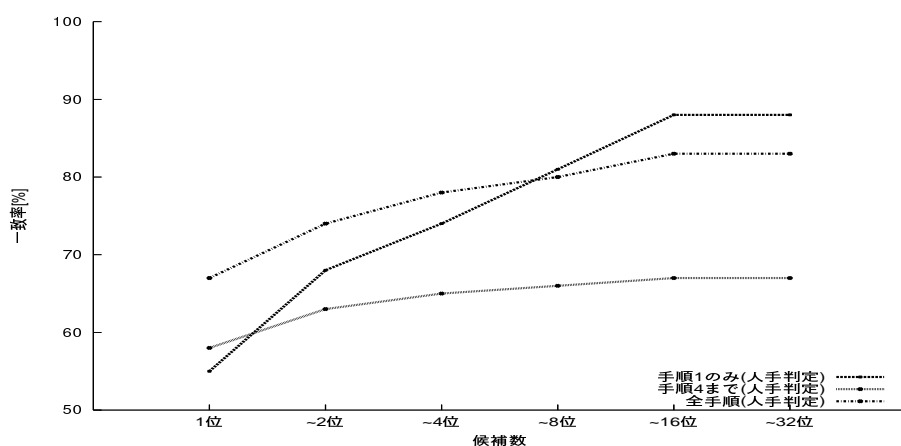


図 9: 実験 2 の結果 (同音異義語の動詞を含む文・人手判定)

「同音異義語の動詞を含む文」での結果において、「(手順1)のみ」と「(手順4)まで」の結果の比較では、正解率に多少の向上が見られる。「全手順」を行った場合は、格段に正解率が上昇している。また、「(手順1)のみ」から「(手順4)まで」の間で、結合価パターンによって下位の候補を一位に引き上げた割合は14%であり、正解率に多少の向上が見られることから、結合価パターンは、正解文を削除する以上に、正解文を引き上げることができることがわかる。

実験2の同音異義語の名詞を含む文での結果を表5, 図10,11に示す.

表5: 実験2の結果(同音異義語の名詞を含む文)

手順		候補数					
		1位	~2位	~4位	~8位	~16位	~32位
手順1のみ	完全一致	36%	53%	63%	70%	77%	82%
	人手判定	58%	72%	83%	89%	94%	98%
手順2まで	完全一致	37%	55%	64%	72%	78%	82%
	人手判定	61%	74%	84%	90%	95%	98%
手順4まで	完全一致	37%	50%	60%	66%	66%	67%
	人手判定	67%	73%	80%	82%	82%	83%
全手順	完全一致	42%	59%	69%	76%	79%	81%
	人手判定	73%	82%	89%	93%	95%	97%

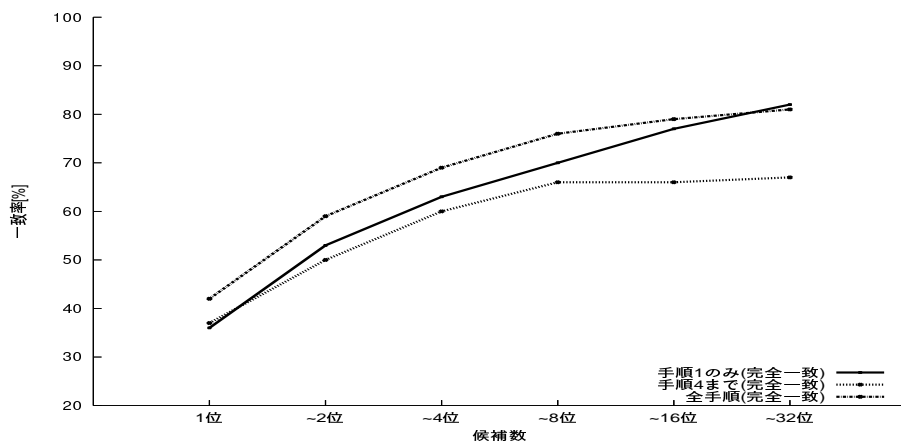


図 10: 実験 2 の結果 (同音異義語の名詞を含む文・完全一致)

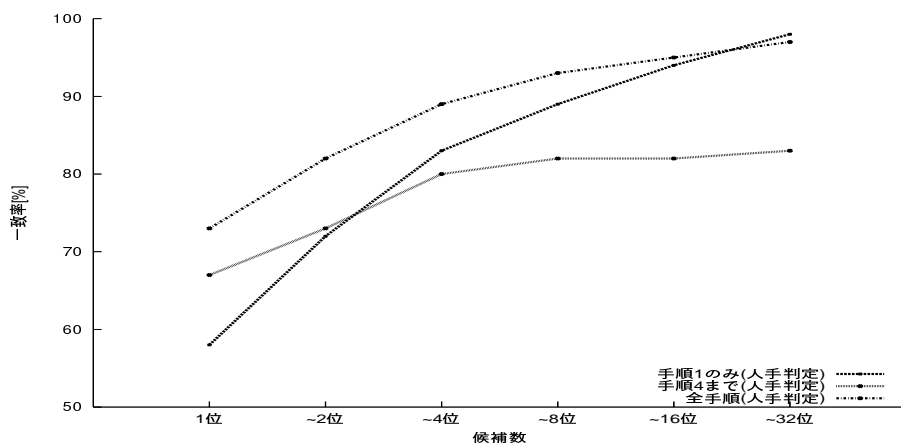


図 11: 実験 2 の結果 (同音異義語の名詞を含む文・人手判定)

「同音異義語の名詞を含む文」での結果では，手順を踏まえるに従って正解率が上昇している．また「(手順 1)のみ」から「(手順 4)まで」の間で，結合価パターンによって下位の候補を一位に引き上げた割合は 16%であった．以上より，結合価パターンは同音異義語に対し，効果があることがわかる．

5.4.3 実験3の結果

実験3の結果を表5, 図12,13に示す.

表6: 実験3の結果(毎日新聞記事)

手順		候補数	1位	~2位	~4位	~8位	~16位	~32位
手順1のみ	完全一致		52%	66%	76%	80%	82%	87%
	人手判定		64%	70%	79%	85%	86%	91%
手順2まで	完全一致		52%	68%	77%	79%	83%	86%
	人手判定		64%	72%	80%	86%	87%	90%
手順4まで	完全一致		40%	52%	57%	57%	57%	57%
	人手判定		55%	58%	60%	60%	60%	60%
全手順	完全一致		58%	71%	77%	78%	78%	81%
	人手判定		73%	78%	81%	83%	83%	86%

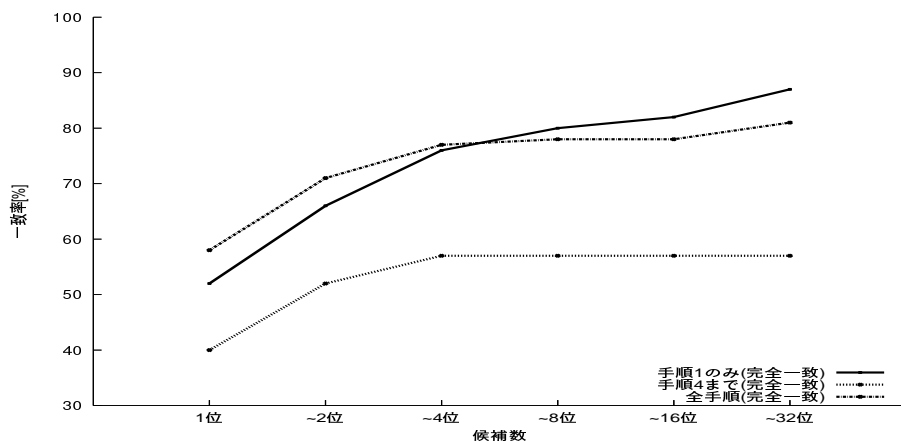


図 12: 実験 3 の結果 (毎日新聞記事・完全一致)

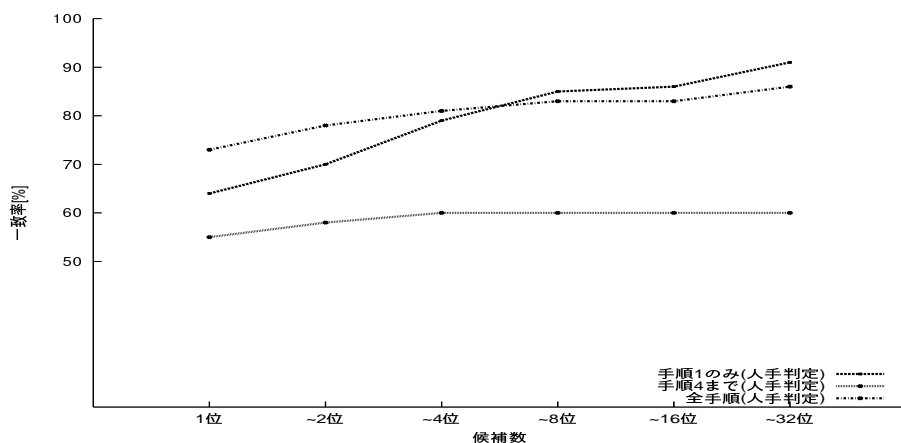


図 13: 実験 3 の結果 (毎日新聞記事・人手判定)

「全手順」を行った場合，正解率が向上している．よって，本手法は実際の文に対しても有効であることがわかる．また，「(手順 1)のみ」から「(手順 4)まで」の間で，結合価パターンによって下位の候補を一位に引き上げた割合は 12%であった．しかし，「(手順 4)まで」の結果を見ると，「(手順 1)のみ」の結果よりも正解率が低かった．次章では，この理由について考察する．

6 考察

実験3の人手判定において「(手順4)まで」により、32個の候補文を全て削除した件数の割合は36%であった。したがって「(手順4)まで」は64%の判定結果を出力したことになり、候補を正しく選択することの適合率($\langle \text{正解数} \rangle / \langle \text{出力数} \rangle$)は86%である。これは「(手順1)のみ」の適合率64%よりも高い。

実際に「(手順1)のみ」から「(手順4)まで」で、結合価パターンが下位の仮名漢字候補文を1位に引き上げた割合が12%であったことから、結合価パターンは、確かに誤り文を削除することに成功している。したがって、結合価パターンは、同音異義語に対し有効であることがわかる。

しかし、一方で再現率は「1~32位」までをみると「(手順4)まで」は60%であるが「(手順1)のみ」は91%である。この差より、結合価パターンによって正解である文が削除されている件数も多いことがわかる。

そこで「(手順4)」の適合率と再現率の向上に向け、誤り事例の分析を行う。

6.1 誤り文が削除できないことについて

「(手順4)まで」における「1位」と「1~32位」の正解率の差より、誤り文が削除できなかった量がわかる。原因、および、該当候補文の数を表11に示す。

原因	文数
(1-1) 結合価パターンが原因	2
(1-2) 結合価パターン以外が原因	3

以下に具体例を載せる。

(原因 1-1) の例

入力文：しきちないにちいさなばばがある。

正解文：敷地内に小さな馬場がある。

誤り候補文：敷地内に小さな祖母がある。

次のパターンで判断したため誤った。

パターン：N1(*)がN2(388 場所 533 具体物 1000 抽象)にある

この場合「祖母(人間)」に対して「ある」は誤りである。

(原因 1-2) の例

入力文： てんしんのことばがよみがえる。

正解文： 転進の言葉がよみがえる。

誤り候補文： 点心の言葉がよみがえる。

パターン： $N1(*)$ が蘇る

誤り候補文において「点心」と「言葉」は、人手での評価では繋がらないと判断し、誤りとした。

以上より、(原因 1-1) の場合、結合価パターンの制約条件の変更が必要となる。しかし、「『彼が～の立場にある』という場合では正しい」というように、一般に制約条件の変更は容易ではない。(原因 1-2) は、人手での判定では誤りとしたが、「の」型名詞句は、状況や考え方によっては繋がることもある。したがって、これら 2 つの原因の解決は困難である。

6.2 正解文が削除されることについて

(手順 2) における「～32 位」と(手順 4) における「～32 位」の正解率の差より、正解するはずの候補文が、意味的削除で削除されてしまった量がわかる。その原因を、実験 3 の結果より調べ、表 12 にまとめる。

表 8: 正解文が削除された原因

原因	件数
(2-1) 結合価パターンの名詞意味属性の制約が不足	10
(2-2) 任意格データベースに登録された名詞意味属性の制約が不足	7
(2-3) 結合価パターンの不足	5
(2-4) 接続辞書のデータ不足	4
(2-5) 格変化規則が不適切	3

以下に具体例を載せる。

(原因 2-1) の例

正解文： 国籍 を拒否された。

パターン： $N1(3$ 主体) が $N2(3$ 主体 1236 人間活動) を拒否する

「国籍」の意味属性は「(1203 籍)」であり、(3 主体 1236 人間活動)の意味属性とは親子関係にないため、パターンが適合せず、正解文が削除された。

(原因 2-2) の例

正解文：これまで、同公社内部に保管されていた。

パターン：N1(3 主体)がN2(533 具体物 1001 抽象物)を保管する

正解文には、任意格の「に格」が含まれている。しかし、任意格のデータベース中に「内部(2623 内部)に」に対応する意味属性が無いので、正解文が削除された。

(原因 2-3) の例

正解文：機長と副操縦士のやりとりが生々しい。

パターン：なし

「生々しい」は、単語辞書には登録されているが、結合価パターン辞書には「生々しい」が存在しない。よって、正解文が削除された。

(原因 2-4) の例

正解文：派遣予定もない。

「派遣予定」の名詞の接続において「派遣」と「予定」の接続が接続辞書に載っていないため、正解文が削除された。

(原因 2-5) の例

正解文：同氏自身の具体的な見解は示さなかった。

パターン：N1(3 主体)がN2(*)をN3(1109 文書 1062 言語)に示す

「見解は」の対応する部分は「N2(*)を」なので「は格」を「を格」に変化させなくてはならない。しかし、格変化規則では「は格」を「が格」に変換することが優先されるため、結合価パターンが誤って適合したことが原因で、正解文が削除された。「が格」を優先した理由は「は格」を「が格」、および「を格」の両方と対応させた場合、誤った判断をすることが動作確認の際に多く見られたためである。例えば「私は作る」の誤り文で「渡しは作る」という文がある。この場合「を格」を用いると「渡し(橋)を作る」と解釈されてしまう。そのため、本稿では使用を控えている。

以上より，(原因 2-1)～(原因 2-4)ではデータベースの拡充，(原因 2-5)では格変化規則の改良がそれぞれ必要となる．この結果より，データベース類を拡充するだけで，原因件数を大幅に減らせることがわかる．よって，精度向上のためには，第一にデータベースの拡充が必要である．また，格変化規則も改良の余地があるため，精度向上に貢献できる．

7 おわりに

本稿では、単語連鎖確率を用いた仮名漢字変換の候補文に対して、結合価パターンを用いた候補選択を行うという統合的な仮名漢字変換アルゴリズムを提案した。

IPALの基本動詞/基本名詞に関する同音異義語を含む仮名文、および、毎日新聞95年度記事の仮名文それぞれ100文を対象とした仮名漢字変換の評価実験において、結合価パターンによって下位の仮名漢字変換候補文を1位まで引き上げた割合は、それぞれ14%、16%、および12%であった。このことから結合価パターンの同音異義語、および、実際の文に対する仮名漢字変換への有効性を確認することができた。

また、本研究のアルゴリズムを全手順について行った場合の1位正解率については、それぞれ12%、15%、および9%の向上が見られた。よって、本アルゴリズムの有効性が確認できた。今後の課題は、データベースの拡充、および、格変化規則の改良である。

謝辞

最後に，本研究において御指導頂きました鳥取大学工学部知能情報工学科計算機C研究室の池原教授，村上助教授，徳久助手に厚く御礼申し上げます．

また「確率的言語モデルによる仮名漢字変換」について詳細な解説をして下さいました森 信介氏に，御礼申し上げます．

なお，本研究で使用した ALT-JAWS は NTT との共同研究の下で使用させて頂きました．

参考文献

- [1] 村上: 漢字かなの TRIGRAM をもちいたかな漢字変換方法, 情報処理学会第 43 回全国大会, 7H-3, Vol.3, pp.287-288(1991).
- [2] 森, ほか 3 名: 確率的モデルによる仮名漢字変換, 情報処理学会論文誌, Vol.40, No.7, pp.2946-2953(1999).
- [3] 本間, 山階, 小橋: 連語解析を用いたべた書きかな漢字変換, 情報処理学会論文誌, Vol.27, No.11, pp.1062-1067(1986).
- [4] 大島, ほか 3 名: 格文法によるかな漢字変換の多義解消, 情報処理学会論文誌, Vol.27, No.7, pp.679-687(1986).
- [5] 池原, ほか 7 名: 日本語語彙大系, 岩波書店 (1997).
- [6] 元永: 意味的共起関係を用いた同音異義語のかな漢字変換, 鳥取大学大学院修士論文 (2000).
- [7] 村上, ほか 3 名: 結合価パターンを用いた音声認識, 情報科学技術フォーラム (FIT-2003), 第 2 分冊 F-024, pp.255-256(2003).
- [8] 吉田, 池原, 村上: 入力文に対する結合価パターン対の選択方法について, 言語処理学会第 8 回年次大会発表論文集, pp.299-302(2001).
- [9] 日外アソシエーツ株式会社: 毎日新聞 95 年度版 CD-ROM(1995).
- [10] 吉田, ほか 3 名: 結合価パターンを用いた仮名漢字変換候補の選択, 言語処理学会第 10 回年次大会, 発表予定 (2004).