# 概要

機械翻訳においては複数の訳語を持つ語の訳語選択が問題となるが、結合価パターンは、訳語選択の精度向上に有効である。しかし、実際の文章では、語順変更、省略等により結合価パターンをそのまま適用できない場合がある。また、省略によって結合価パターン候補が多数できるため、パターン候補の絞り込み方法が重要な問題となる。

本研究は、語順変更、省略に対応した精度の高い結合価パターン対選択方式の考案を目的とする。まず、入力文の形態素解析を行い、文を格要素ごとに分割する事で、語順変更可能な照合を行う。さらに照合後に残った候補パターンを絞るため、格要素への点数付けのシステムを設ける事で省略に対応した結合価パターン照合を行う。

結果としては、94 %の入力文を一意に絞る事ができ、パターン候補を絞るのに有効である事が分かった。しかし、複数の性質を持つ格助詞については、別の法則を用いる等、改良の余地がある事も分かった。

# 目 次

**1 はじめに** **4**

**2 結合価パターンによる意味的制約** **5**
2.1 結合価パターン . . . . . 5
2.2 一般名詞意味属性体系 . . . . . 6
2.3 単文の意味的整合性の判定 . . . . . 7

**3 結合価パターンの照合** **8**
3.1 格助詞の点数付け . . . . . 8
3.2 格助詞の配点について . . . . . 9
3.3 点数計算の例 . . . . . 10

**4 受身、使役文への対処方法** **11**

**5 評価実験の方法** **12**
5.1 実験の手順 . . . . . 12
5.2 用いる文のルール . . . . . 13
5.3 人手による評価基準 . . . . . 14

**6 実験の結果** **16**
6.1 格変化、点数計算の有効性 . . . . . 16
6.2 格助詞、意味属性の深さの有効性 . . . . . 18
6.3 格変化、点数計算使用データの詳細 . . . . . 19
6.4 パターン無しの内訳 . . . . . 23
6.5 登録されていなかった動詞 . . . . . 24
6.6 受身、使役文への対応精度 . . . . . 25
6.7 実験に使用したデータ . . . . . 26

**7 考察** **27**
7.1 誤ったパターンを導き出した場合 . . . . . 27
7.1.1 「は」「も」格が別の性質を持つ場合 . . . . . 28
7.1.2 点数計算が不適切だった場合 . . . . . 30
7.1.3 格の変化に対応しきれなかった場合 . . . . . 31
7.1.4 形態素解析のミスがあった場合 . . . . . 31
7.2 全体的な特徴 . . . . . 32

**8 結論** **33**

# 1　はじめに

機械翻訳では複数の訳語を持つ語の訳語選択が問題となる。解決の手法として結合価パターンは訳語選択の精度向上に有効である [1]。しかし、実際の文は語順変更、省略等により現状の登録パターンで結合価パターンがそのまま適用できない場合がある。特に省略の場合、正しい結合価パターンが導き出せない場合が多数あるため難しい問題となる。よって候補パターンを出来る限り絞り、中から正しい結合価パターンを導き出す事が、結合価パターンを用いる機械翻訳では重要となる。

過去の研究では、入力文を辞書登録パターンに対応させ、パターンの照合については格要素が最も一致するパターンを選択する方式と、それとは逆に辞書登録パターンを入力文に合わせる方式が提案されている [2]。辞書登録パターンを入力文に合わせる方式は、具体的には入力文の様相機能語や助詞に着目して特殊構文の可能性を検出し、辞書登録パターンから可能な展開形を複数派生させ、展開形と入力文と照合する方法である。(参考文献 [2] より引用) 利点として、「れる、られる」の自発、尊敬、受身、可能の判定や二重主格文の結合価パターン絞り込みの精度向上を実現している。しかしながら、語順変更、省略は考慮されていない。

本研究では、入力文を辞書登録パターンに対応させる方式を用い、パターン照合については名詞の意味属性、格助詞に着目した点数計算によるパターン照合を行う。そして、語順変更、省略に対応した精度の高い結合価パターン対選択方式の考案を目的とする。基本として単文での入力文の形態素解析を行い、既存のパターンで入力文との照合を行ってパターンを絞る。照合の際には、語順変更に対応できるよう、語順を入れ換えた場合の照合も行う。また、結合価パターンは受身、使役に対応していない事から、格変化を用いて一時的に平叙文に戻す方法を使用する。さらに候補パターンを一意に絞るため、文を格要素ごとに区切り、格要素の持つ名詞の意味属性の深さや格助詞の種類に着目する。深さ、格助詞への点数付けを行い、格要素ごとの点数の積の合計によって結合価パターンの候補を絞る事で、省略部分の重要性を見極め、正しいパターンを導き出す方法を提案、検証する。

点数付け、受身、使役に対応した格変化を用いて単文 300 文について実験を行った結果、94 %の精度で正しい結合価パターンを選択する事が出来た。しかしながら、受身、使役に対応した格変化はまだ改良の余地があり、「は」「も」等の複数の格の性質を持つ格助詞の扱いについても別法則を用いる必要がある事が分かった。

本稿では、第 2 章で結合価パターンについて述べ、第 3 章では具体的な結合価パターンの照合方法について述べる。第 4 章では受身等の追加法則について示す。第 5 章では詳細な実験の方法を述べる。そして第 6 章では実験の結果について述べ、第 7 章で得られた結果の考察を述べる。

# 2　結合価パターンによる意味的制約

## 2.1　結合価パターン

結合価パターンは、用言と格要素 (名詞+格助詞) の意味的関係を記述したもので、用言と格要素間に意味的な制約をつける事で訳語選択の精度を高める。

本研究では、日本語語彙大系 [3] にある「構文意味辞書」の結合価パターンを基本にする。結合価パターンは一般表現と慣用表現を合わせ、約 16,000 件の日本語文型パターンにまとめている。結合価パターンの例を表 1 に示す。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。慣用表現の場合は直接単語の字面で表現するので番号は付かない。

表 1: 結合価パターンの例 ( 一般表現と慣用表現 )

| | 結合価パターン | 対応する英文 |
|---|---|---|
| 一般表現 | $N1$(2483 性質) が大きい | $N1$ be remarkable |
| | $N1$(3 主体) が $N1$(4 人) を絞る | $N1$ grill $N2$ |
| 慣用表現 | $N1$ が (名) を売る | $N1$ make $N1$-self famous |
| | $N1$(4 人) が (知恵) を絞る | $N1$ rack $N1's$ brains |

一般表現は用言と一つ以上の格要素で規定する。慣用表現は、名詞が意味属性でなく直接単語の字面で規定され、主に特殊な表現で使う。

## 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 1) を用いる。一般名詞意味属性とは、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 1 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。

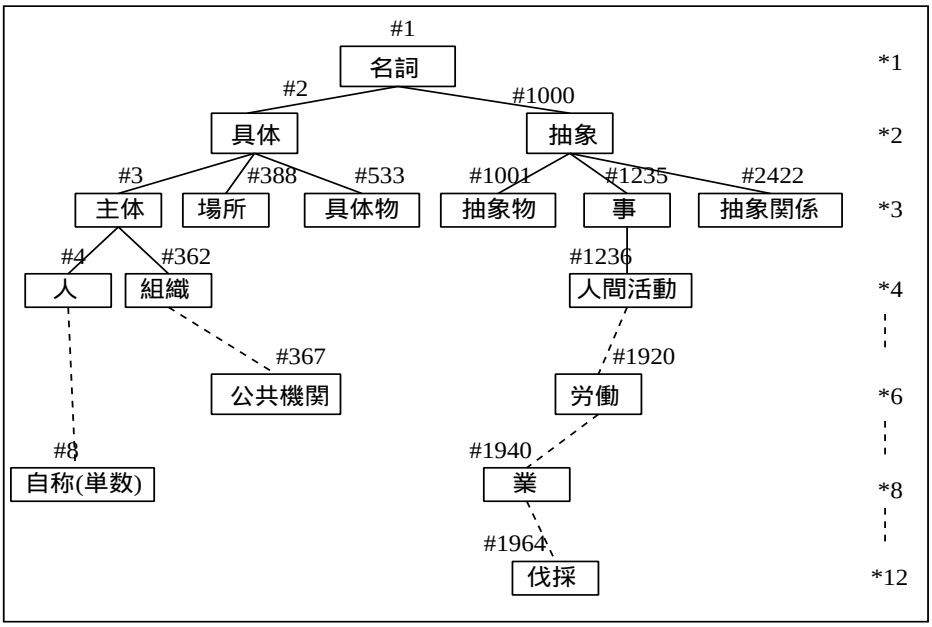


図 1: 一般名詞意味属性体系の一部

注：＃ 1 ~ 1964 は名詞意味属性番号、付随する単語は名詞意味属性。＊は名詞の意味属性の深さ。

## 2.3　単文の意味的整合性の判定

文の意味的整合性を判定する方法を、具体例によって示す。判定は結合価パターンと、名詞意味属性体系の上位の意味属性が下位の意味属性を包含する事を利用する。

以下の結合価パターンの例について考える。

$N1$(3 主体) が $N2$(4 人) を絞る。
注：() 内の番号は名詞の意味属性番号、単語は意味属性。

上記の例に対して次の 2 文を考える。

例文 **1)** (私) が (人員) を絞る。
例文 **2)** (彼) が (頭) を絞る。

$N1$ にあてはまる名詞は、例文 1 においては (8：自称 (単数))、例文 2 においては (22：他称 (単数)) であり、共に (4 人) 以下の属性なので正しい答となる。しかし $N2$ にあてはまる名詞は、例文 1 においては (4 人) 以下の属性の名詞 (119 ：成員) なので正答になるのに対し、例文 2 では (554 頭) という人以下の属性の名詞ではないので別の意味のパターンであると判定される。

# 3　結合価パターンの照合

## 3.1　格助詞の点数付け

本研究では、複数出現する結合価パターンの候補を絞るために格要素に着目し、名詞と付随する格助詞を利用する事で省略に対応する照合方法を提案する。省略のある文は、省略部分によって重要性が変わる。例えば「が」格なら省略しても文全体の意味に影響しない事が多い。しかし、「を」格は目的格となる場合が多く、省略すると文全体の意味が曖昧になり、最悪の場合意味不明の文となる。よって、格助詞の種類によって、重要性が違うと仮定し、格助詞に点数付けを行う。また、格助詞が付随する名詞は、意味属性の深さが深ければ深い程名詞の意味を制約する働きがあると考えられる。意味を制約する事で候補パターンを絞り込む事ができる事から名詞の深さを点数として用い、格助詞との積で点数を決定する。候補パターンの中で格要素の点数合計が最高得点のパターンを最適なパターンとする事で候補文を一意に絞る。具体的には以下の計算式を用いる。

$$D = \Sigma\ (A * B) \qquad \cdots (1)$$

$D$：パターンの点数 $A$：名詞の意味属性深さ $(\mathbf{1} \sim \mathbf{12})$ $B$：格助詞の点数
式 $(1)$ は、名詞の意味属性の深さが深ければ深いほど意味的制約が強くなる事を利用する。また、名詞に付随する格助詞の種類が文の意味に大きく関わると見て、格助詞にも点数付けを行う。表の格要素の点数付けについては、[3] にある結合価パターンでの格助詞の使用頻度によって点数を規定する。例えば「が」格は頻繁に使用される割に省略しても文章全体の意味に影響する事は少ないので点数は低くする。基本的に使用頻度の低い物を重要度が高いとし、点数付けを行う。また、各パターンにおいての省略部分については減点を行う。

### 3.2 格助詞の配点について

格助詞の使用頻度調査は [3] を利用する。[3] に登録してあるパターン数が 10 個の自動詞と他動詞から、各 5 種類を選び、用いられている格助詞の使用頻度を調査する。調査した格助詞の使用頻度を表 2 に示す。

表 2: 格助詞の出現頻度

| が格 | に格 | を格 | で格 | へ格 | の格 | と、から格 |
|---|---|---|---|---|---|---|
| 100 % | 56 % (56/100) | 51 % (51/100) | 16 % (16/100) | 11 % (11/100) | 7 % (7/100) | 3 % (3/100) |

注：その他に「まで」格、「より」格、格要素無しが 1 % 有り。

次に、格助詞の使用頻度を元に算出した格助詞の加点、減点を表 3 に示す。「が」「に」「を」格以外は使用頻度が低い事と、使用頻度の高い格助詞でも重要度が高い場合がある事を考慮し、点数の調整は点数の開きが少ないように行う。減点については、この加点での最低点数が減点の最高点数と同じになると、減点による弊害が発生するため、調整をする。また、減点によって誤った結合価パターンを出来る限り導き出さないように最悪の場合を想定して配点をする。加点、減点はどちらも名詞の属性の深さと表 3 の格助詞の積で一つの格要素分の点数となり、全ての格要素点の合計が 1 パターンの点数となる。

表 3: それぞれの格助詞の配点

| | が格 | に格, を格 | で格 | へ格 | の格 | その他 |
|---|---|---|---|---|---|---|
| 加点 | 3 点 | 3.5 点 | 4 点 | 4 点 | 4.5 点 | 5 点 |
| 減点 | 0.5 点 | 0.5 点 | 0.8 点 | 1 点 | 1.2 点 | 1.5 点 |

### 3.3　点数計算の例

点数計算の例を示す。

例 **1**：勉強会を開く。

候補パターン 1：(3 主体) が (1 全て) を開く。
候補パターン 2：(3 主体) が (1223 式、行事) を開く。(正解)
パターン 1 の計算 (深さ 1 * 3.5 点) − (深さ 3 * 0.5 点)=2 点
パターン 2 の計算 (深さ 6 * 3.5 点) − (深さ 3 * 0.5 点)=19.5 点
どちらの候補パターンも、「が」格が省略のために、「が」格の分だけ減点となる。加点は「を」格の分で加点、減点が 1 つずつとなる。点数計算では 2 のパターンが正解となる。1 のパターンは「~を開ける」という意味に対し、2 のパターンは「会を開く」という意味に取れる事からも 2 が正解となる。

次に、減点の有効性を示す例を述べる。

例 **2**：大会開催が決まった。

候補パターン 1：(1 全て) が決まる。
候補パターン 2：(1 全て) が (1 全て)[から、より、で] 決まる。
パターン 1 の計算 (深さ 1 * 3 点) = 3 点
パターン 2 の計算 (深さ 1 * 3 点) − (深さ 1 * 1.5 点) = 1.5 点
減点は、余分な格要素が入っている場合を除外できるため、有効である。ただし、省略という事を考慮する場合、パターン 2 もありうる。本研究では、出来る限り省略が無く、合致するパターンを正答とする。

# 4　受身、使役文への対処方法

今回使用した結合価パターンは能動態で登録されている。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。本研究では入力文の格助詞を変更して平叙文で認識する。そして、英文に変換する際には元に戻して対応する。一時的に文章を平叙文に直す事により、入力文は結合価パターンに対応し、受身、使役文に対しても結合価パターンを用いる事ができる。

平叙文に直す作業は基本的に格要素の格助詞を変える事で行う。表 4 の格変化対応表は受身、使役文を平叙文に直す際に最もよく使う格変化を基本とする。ただし、「に」格についてはかかる名詞が主体 (注釈あり) 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。

例：契約が交わされている。

→契約を交わす。(「が」格→「を」格に変化)

表 4: 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 |

注：一部例外あり。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。() の数字は名詞の意味属性番号で単語は名詞の意味属性である。

# 5　評価実験の方法

語順変更、省略、受身、使役文に対応した点数計算の結合価パターン対応精度を調べる。方法として、新聞記事データより文章を抽出し、点数計算する事で適切な結合価パターンを導き出す精度を実験的に調べる。

## 5.1　実験の手順

以下に実験の手順例を示す。

１：新聞記事の文を単文に加工する。

弱いように見えるものが強いものを従え柔らかいものが固いものを征服するという老子の予言が胸にずっしり響く。→予言が胸にずっしり響く。

２：単文を形態素解析する。
予言が胸にずっしり響く。
1. 予言 (1220,[1503,1436,1025])/が (7410)
2. 胸 (1100,[578,1238])/に (7430)
3. ずっしり (4110)
4. 響く (2116)/。([P]0110)

３：形態素解析済みの文をプログラムにかける。
ID.No 101818 = 6.500000 (候補パターンの点数)
(2) 予言が/胸に/ずっしり/響く。 [ 101818 ] (入力文と候補パターン ID)
101816 : 2108 1064 が 578 863 2610 に から響く
101817 : 1349 2354 922 1064 が 388 2610 に で響く
101818 : 1 が 1 に響く (結合価パターン照合において候補に上がったパターン)
540107 : 1148 が響く
540086 : 2354 が 874 884 に響く
(上の 5 文は結合価パターン ID と結合価パターン)

４：人手で正解かどうか判断する。今回の場合は候補パターンが一つしかない。101818 のパターン「(1 全て) が (1 全て) に響く」は意味的に入力文と合致しているため、正解となる。

## 5.2　用いる文のルール

　また、本研究では以下のルールを加える。
・副詞、感動詞は結合価パターンとは関係が無いので除外する。

例：凄く、おや、等。この系統の語は除外する。

・可能、尊敬については考慮しない。ただし、可能においては受身と意味が似通った物は受身として判別する。

例：位置や水の流れに月ごとの変動が見られた。
例の場合、可能が受身と似通っているので受身の意味とする。

・「は」「も」格は、基本的に「が」格扱いとする。

例：ロシア軍は激しい空爆と砲撃を加えた。
例の場合等で、「は」格を「が」格として認識する。

## 5.3　人手による評価基準

評価は入力文と結合価パターンの意味を見て比べ、意味が概ね同じ物であったなら正答とする。本研究においては、正答であったとしても、複数の場合がある事が認められたので、いくつかの場合に分けて評価する。

**1)** ◎ ：導き出したパターンの意味が正しい場合
**2)** ○ ：導き出したパターンは正しいが、類似のパターンが他にも存在する場合
**3)** △ ：導き出したパターンの意味で正しいが、別の意味のパターンでも意味が通る場合
**4)** * ：形態素解析ミスによる意味の取り違いの場合、もしくは格助詞が複数の意味を持つ為に正しい結合価パターンを導きだせなかった場合
**5)** × ：点数計算で正しいパターンを導き出せなかった場合

正答率としては、△ までを正答とし、それ以外は誤りとしている。

**1)** ◎ の例：勉強会を開く。

候補パターン 1：(3 主体) が (1 全て) を開く。
候補パターン 2：(3 主体) が (1223 式、行事) を開く。(正解)
パターン 1 の計算 (深さ 1 * 3.5 点) - (深さ 3 * 0.5 点)=2 点
パターン 2 の計算 (深さ 6 * 3.5 点) - (深さ 3 * 0.5 点)=19.5 点
点数計算上ではパターン 2 の方がパターン 1 より高いのでパターン 2 を正しいパターンとしている。パターン 1 は (箱、扉等の開ける物) を開くという意味に対しパターン 2 は (会議、会合等) を開く (開催する) という意味なので意味的に入力文と合致するのはパターン 2 となる。よって点数計算でパターン 2 を選出するのは正しいという事になり、◎ の評価となる。

**2)** ○ の例：大会開催が決まった。

候補パターン 1 : (1 全て) が決まる。
候補パターン 2 : (1 全て) が (1 全て)[から、より、で] 決まる。
3.2 章で述べたパターンで、類似したパターンがある場合にあたる。1、2 とも、どちらの意味とも取れる。

**3)** ▵ の例：軍隊は全滅する

候補パターン 1 : (1 全て) が全滅する (3 点)
候補パターン 2 : (3 主体 535 動物) が (1 全て) を全滅する (8.5 点)
この文の場合、候補パターン 2 を正解としており、確かにパターン 2 の意味で取る事は可能である。しかし、本来は別の意味での候補パターン 1 である可能性もある。○ とは異なり、明らかに意味が違うが、両方の意味で取れる場合のパターンである。

＊、× については考察で述べる。

# 6　実験の結果

## 6.1　格変化、点数計算の有効性

まず、正答率についての結果を表 5 に載せる。本研究の成果を明確にするために、比較材料として、3 種類のデータを用意する。

**1**：格変化、点数計算無し：受身に対して何の処置も行わず、候補パターンが一つだけならそのまま正答となるが複数の候補パターンが出た場合には候補の中に一つだけ正解があるとして正解を選ぶ確率を加算した場合。
**2**：点数計算のみ：受身に対しての処置を行わず、複数の候補パターンが出た場合は点数計算によって候補パターンを一意に絞る場合。
**3**：格変化、計算使用：受身に対して格変化を行い、複数の候補パターンが出た場合は点数計算によって候補パターンを一意に絞る場合。

ただし、この正答率は、全てパターン無し (パターンが登録されていない) と判定された文は除き、慣用表現を用いている文についても除く。パターン無しの文は 58 文、慣用表現の文は 8 文なので、正答率の判定に用いた文は 234 文である。

表 5: 正答率比較 1

| 格変化、点数計算無し | 点数計算のみ | 格変化、計算使用 |
|---|---|---|
| 81.9 % ((161+30.54)/234) | 86.8 % (203/234) | 94.0 % (220/234) |

注：30.54 は複数の候補パターンが残った場合、正解を選ぶ確率。複数のパターンが残った文数は 59 文で、59 文の中で正解が出るのは 30.54 文という事になる。

また、表 5 について例文を用いて以下に示す。

**1**：格変化、点数計算無し

例文：大会開催が決まった。

候補パターン 1：(1 全て) が決まる。(正解)
候補パターン 2：(1 全て) が (1 全て)[から、より、で] 決まる。
格要素の照合によって候補パターンは 2 つに絞られる。だが、点数計算をしないため、無作為に正答を選ぶとすると正答を選ぶ確率は 1/2。よって、2 回に 1 回は正答を選ぶとして 0.5 を正答数に加算する。

**2**：点数計算のみ

例文：大会開催が決まった。

候補パターン 1：(1 全て) が決まる。(3 点)(正解)
候補パターン 2：(1 全て) が (1 全て)[から、より、で] 決まる。(1.5 点)
点数計算を使用するのでパターン 1 を選択する。意味的にもパターン 1 が正解となり、正答数に 1 加算する。

**3**：格変化、計算使用
例文が同じ場合は点数計算のみと同じなので別の例で示す。

例文：契約が交わされている。

候補パターン 1：(3 主体) が (抽象 1000 具体物 533) を (3 主体) と交わす。(1 点)
候補パターン 2：(3 主体) が (1717 契約) を交わす。(30 点)(正解)
格変化を使用するので、「 契約を交わす 」として扱う。格変化を使用しないと「 契約が交わす 」となり、認識できない。格変化と格要素照合により、候補パターンが 2 つに絞れたので点数計算をして、パターンを一意に絞る。パターン 2 の方が点数が高く意味的にも正しいため、正答となる。

## 6.2 格助詞、意味属性の深さの有効性

格助詞と名詞属性の両方を用いる事の有効性を示すために、参考として以下の 3 通りの計算を比較する。

1：格助詞の点数のみでの計算

2：名詞の意味属性の深さのみでの計算

3：名詞の意味属性の深さと格助詞の点数の積による計算

配点については前述した点数をそのまま用い、受身、使役に対応した格変化も用いる。ここで複数候補の正答率とは、複数の候補パターンが残った時に点数計算によって正解を導き出した正解率を表す。ただし、点数計算だけの成功率とする。

表 6: 正答率比較 2

| | 格助詞のみ | 意味属性の深さのみ | 格助詞、意味属性の深さ併用 |
|---|---|---|---|
| 複数候補の正答率 | 68.9 % (42/61) | 93.4 % (57/61) | 96.7 % (59/61) |
| 全体の正答率 | 86.8 % (203/234) | 93.2 % (218/234) | 94.0 % (220/234) |

格助詞のみを用いた場合は、格助詞単独用に点数の調整をしていないため、精度はあまり高くはない。だが、意味属性の深さを用いる場合は、両方を用いた場合に近い結果となる。結果より格助詞、意味属性はパターン選択精度向上に有効である事が分かる。

## 6.3 格変化、点数計算使用データの詳細

次に、格変化、点数計算を使用した場合の詳細を表 7 に示す。

一意正答 (1)：結合価パターン自体が一種類しか存在せず、そのまま対応し、正答と判定できる場合。
一意正答 (複)：結合価パターンが複数存在するが、名詞の意味属性、格助詞の一致判別によって一意になり、正答と判定できる場合。
複数正答：結合価パターンが複数存在し、名詞の意味属性、格助詞によって候補パターンが絞られたが、一意に絞れず、点数計算によって一意に絞られ正答となった場合。

表 7: 格変化、計算を使用した場合の詳細 (1)

| 一意正答 (1) | 一意正答 (複) | 複数正答 | 慣用表現 | パターン無し | 誤り |
|---|---|---|---|---|---|
| 19.3 % (58/300) | 34.3 % (103/300) | 19.7 % (59/300) | 2.7 % (8/300) | 19.3 % (58/300) | 4.7 % (14/300) |

また、表 7 について以下に示す。

一意正答 (1) の例：老子が若返った。

「若返る」のパターン：(4 人) が若返る。
登録パターンが 1 つしかなく、入力文に適合する場合である。今回の場合、老子は人属性であるので、パターンに適合する。

一意正答 (複) の例：英文をもとに翻訳を進めた。

(以下、進めるの登録パターン)
登録パターン 1：3 が 3 535 を 388 2610 318 2585 から 388 2610 318 2585 に へ進める
登録パターン 2：3 が 1001 1236 を進める
登録パターン 3：2355 838 が 1357 を進める
登録パターン 4：4 が 984 を進める
……
登録パターン 10：3 が 1422 を進める
10 個のパターンの中で例文に対応するパターンはパターン 2 だけである。複数のパターンがあるが、適合するパターンは 1 つしかない場合である。
複数正答の例：ロシア軍は激しい空爆と砲撃を加えた。
候補パターン 1：(3 主体 533 具体物) が (1 全て) を (1 全て) に加える
候補パターン 2：(3 主体) が (1762 攻撃 1858 懲罰 2058 害 1242 痛み 暴力) を (3 主体) に加える
候補パターン 1 の点数 = 12.000000
候補パターン 2 の点数 = 39.000000
照合によって結合価パターン候補が 2 つ以上に絞られるが、一意ではないために点数計算によって一意に正しいパターンが求まる場合である。

慣用表現の例：予算の編成に知恵を絞る。

「知恵を絞る」は慣用表現となるため、自動的に慣用表現のパターンが選択される。

パターン無しの例：見解を表明する。

「表明する」という動詞は登録パターンにあるが、該当するパターンが無いためにパターン無しとなる。

誤りの例：首相訪米は予定通り行う。

「は」格が「を」格の性質を持つため、「が」格として認識すると、「首相訪米」は主格にならないので誤りとなる。

次に、全ての正答と誤りについての詳細なデータについて表 8 に示す。パターン無しと慣用表現を除いているため、総文数は 234 文となっている。

表 8: 格変化、計算を使用した場合の詳細 (2)

| ◎ | ○ | △ | * | × |
|---|---|---|---|---|
| 85.0 % (199/234) | 5.6 % (13/234) | 3.4 % (8/234) | 5.1 % (12/234) | 0.9 % (2/234) |

完全に正答と言えるのは◎ だが、○ や△ も正しい答えではあるので総合的には正答としている。(よって正答は 220/234)

また、表 8 について以下に示す。
**1)** ◎ ：導き出したパターンの意味が正しい場合
**2)** ○ ：導き出したパターンは正しいが、類似のパターンが他にも存在する場合
**3)** △ ：導き出したパターンの意味で正しいが、別の意味のパターンでも意味が通る場合
**4)** * ：形態素解析ミスによる意味の取り違いの場合、もしくは格助詞が複数の意味を持つ為に正しい結合価パターンを導きだせなかった場合
**5)** × ：点数計算で正しいパターンを導き出せなかった場合

**1)** ◎ の例：英語の本を携える。

照合パターン：(4 人) が (533 具体物 1002 抽象物 (精神)) を (594 手) に携える
今回の例では、パターンは一つしかなく、意味的にも正しいので正答となる。

**2)** ○ の例：ドゥダエフ政権部隊の兵士が、南方に脱出している。

候補パターン 1：(3 主体 535 動物 760 人工物 1000 抽象) が (388 場所 1000 抽象) を脱出する (7.5 点)
候補パターン 2：(3 主体 535 動物 760 人工物 1000 抽象) が (388 場所 1000 抽象) から より脱出する (4.5 点)
点数計算上ではパターン 1 が正答となるが、ほぼ同じ意味でパターン 2 の意味としても取れる。

**3)** △ の例：ソ連は消滅した。

候補パターン 1：(1 全て) が消滅する (3 点)
候補パターン 2：(3 主体 526 宇宙 533 具体物 2305 非生命現象) が (1 全て) を消滅する (8.5 点)
点数計算上ではパターン 2 が正答となるが、別の意味としてパターン 1 であった可能性もある。

**4)** * の例：関根容疑者が住民表を移した。

候補パターン 1:$N1$(3 主体) が $N2$(2 具体) を $N3$(388 場所 2610 場 533 具体物)[から、より]$N4$(388 場所 2610 場 533 具体物)[に、へ] 移す。(正解)
候補パターン 2：$N1$(3 主体) が $N2$(1000 抽象) を $N3$(2051 実施) に移す。
パターン 1 の計算：(深さ 3*3)+(深さ 2*3.5)-(深さ 4*1.5)-(深さ 4*1)=6
パターン 2 の計算：(深さ 3*3)+(深さ 2*3.5)-(深さ 8*0.5)=12
パターン 1 が意味的に正解であるが、点数計算上ではパターン 2 となり誤りとなる。

**5)** × の例：3 日も首都空爆を行った。

「も」格が「に」格の意味を持つので誤りとなる。

## 6.4　パターン無しの内訳

パターン無しとなった場合の内訳を結合価パターンとして登録されていなかった動詞、今回導入した受身、使役文への対応精度について表 9 に示す。

表 9: パターン無しの内訳

| パターンが無い | 動詞自体が登録されていない |
|---|---|
| 63.7 % (37/58) | 36.2 % (21/58) |

また、表 9 について以下に示す。

パターンが無い

例文：詩人が一節を訳した。

「訳す」のパターン (一種類)：(3 主体 962 機械) が (1002 抽象物 (精神) 1161 掟 1237 精神) を (1062 言語) から (1062 言語)[に、へ] 訳す
パターンが無いというのは、動詞自体は結合価パターンとして登録されているが、入力文に適合するパターンが存在しない場合である。

動詞自体が登録されていない

例文：ロシア軍が首都から敗走した。

動詞自体が登録されていないというのは、結合価パターンとして動詞自体が未登録である場合である。

## 6.5　登録されていなかった動詞

今回調べた中で結合価パターンとして登録されていなかった動詞を表 10 に示す。

表 10: 今回登録されてなかった動詞

| | | | | | | |
|---|---|---|---|---|---|---|
| 敗走する | 追いかける | 進攻する | 拠出する | 持って来る | 大切にする | 詰めかける |
| 死守する | 印象づける | 覆す | 新たにする | 取りやめる | 積み重ねる | |
| 三唱する | 飛来する | 画像処理する | 明らかにする | 踏み外す | カバーする | |

以下に登録のなかった動詞の文全てを示す。
ロシア軍が首都から敗走した。
強さを追いかける。
ロシア軍は首都に進攻した。
一千万円を拠出する。
工事一件は持って来てもらう。
手つかずの自然を大切にする。
現代生活の中で、人間同士の触れ合いを大切にする。
ラグビーファンら約 2 万人が詰め掛けた。
数十人の報道陣や近所の人たちが詰め掛けた。
官邸周辺地域はチェチェン側が死守している。
敗北を強烈に印象づけた。
ロシア政府の発表を覆す。
株式市場の再生に決意を新たにした。
村山富市首相は伊勢神宮参拝を取りやめた。
育成策を積み重ねた。
全員で万歳を三唱した。
ジョンマッキロップさんの家に多くの野鳥が飛来する。
データをコンピューターで画像処理する。
大出俊郵政相は一九九五年の主要施策を明らかにした。
冒頭陳述は手続きを踏み外した。
増加コスト分全部を料金でカバーする。

## 6.6　受身、使役文への対応精度

今回導入した受身、使役文への対応精度について表 11 に示す。

表 11: 受身、使役文での比較

| | 正答 | 対応不可 | パターン無し | 正答率 |
|---|---|---|---|---|
| 受身、使役対応有り | 85.2 % (23/27) | 3.7 % (1/27) | 11.1 % (3/27) | 95.8 % (26/27) |
| 受身、使役対応無し | 22.2 % (6/27) | 66.6 % (18/27) | 11.1 % (3/27) | 25.0 % (6/24) |

注：正答率はパターン無しを除く。

また、表 11 について以下に示す。

受身、使役対応有り

正答例：委員会の事務局が置かれた。

「が」格が「を」格に変化し、「事務局を置く」として認識するために、パターンが適合する。

対応不可例：花を咲かせる。

使役文に対する誤りである。「を」格が「が」格に変化しなくてはならないため、現在の格変化規則では対応できない。

パターン無しの例：欄外に招致委の名称が印刷された。

受身文であるが、適合するパターンが存在しない場合である。

受身、使役対応無し

正答例：数字が羅列されていた。

「数字を羅列する」の意味で取るはずではあるが、「数字が羅列する」という意味で取る事によって、別の意味で正答となる。

対応不可例：委員会の事務局が置かれた。

受身、使役対応有りでは正答となるが、対応無しでは「が」格が変化しないため、パターンと適合しない。

パターン無しの例：欄外に招致委の名称が印刷された。

パターン無しについては、受身、使役対応有りと共通である。

## 6.7　実験に使用したデータ

実験に使用したデータの一部について記載する。全てのデータについては付録として添付する。データは 95 年度毎日新聞記事を単文に加工しており、計 300 文である。データは全て形態素解析済みである。

---

詩人が一節を訳した。
1. 詩人 (1100,[241])/が (7410)
2. 一 ([P]1610,[2586], 1 )/節 (6220,[2595,2588])/を (7430)
3. 訳し (2334, 訳す)(2433, 訳する)/た (7216)/。([P]0110)

---

英訳本を手にした。
1. 英訳 (1220,[1501,1111])/本 (6280,[920,1119])/を (7430)
2. 手 (1100,[594,1035,1166,592])/に (7430)
3. し (2433, する)/た (7216)/。([P]0110)

---

老子が若返った。
1. 老子 (1900,[183,189,186,48])/が (7410)
2. 若返っ(2184, 若返る)/た (7216)/。([P]0110)

---

英訳をもとに翻訳を進めた。
1. 英訳 (1220,[1501,1111])/を (7430)
2. もと (1100,[769,2590,1198,1188,2446], 元)
(1100,[2617,2636], 下)(1500,[2707,2099], 元)/に (7430)
3. 翻訳 (1220,[1501,1111])/を (7430)
4. 進め (2413,[1625,2156,2158,2164,2302], 進める)/た (7216)/。([P]0110)

---

軍隊は全滅する。
1. 軍隊 (1100,[369])/は (7530)
2. 全滅する (2236,[2079])/。([P]0110)

---

花を咲かせる。
1. 花 (1100,[698,1675,2299])/を (7430)
2. 咲か (2112, 咲く)/せる (7136)/。([P]0110)

---

# 7　考察

## 7.1　誤ったパターンを導き出した場合

今回の実験で、正答率は高い数値を得る事が出来た。しかし、一部誤ったパターンを導き出した場合があったので、以下に詳細について表 12 と共に述べる。7.1 は、パターン未検出も含む。

表 12: 誤ったパターンを導き出した場合

| | 原因 | 文数 |
|---|---|---|
| 7.1 | 「は」「も」格が別の性質を持つ場合 | 3.8 % (9/234) |
| 7.2 | 点数計算が不適切だった場合 | 0.9 % (2/234) |
| 7.3 | 格の変化に対応しきれなかった場合 | 0.9 % (2/234) |
| 7.4 | 形態素解析のミスがあった場合 | 0.4 % (1/234) |

7.1.1 ~ 7.1.4 で 7.1 ~ 7.4 の失敗パターンを全て示す。前述した * の例は 7.1.1、7.1.3、7.1.4 に対応し、× は 7.1.2 に対応する。

### 7.1.1 「は」「も」格が別の性質を持つ場合

失敗例 **1-1**：首相訪米は予定通り行う。

出力はパターン無しとして判定される。
正解パターン：(3 主体 962 機械) が (1395 思考 1002 抽象物 (精神) 1560 行為 1479 明示・暗示 2064 変動) を行なう。
「は」格が「を」格の性質を持つため、「が」格として認識してしまうと、「首相訪米」は主格にならないので誤りとなる。対処としては「は」「も」格に対して新たな法則を見出す必要がある。あてはまる全ての格助詞の性質を持たせてみたが、「が」格でしかあてはまらない場所に「を」格が当てはまるなど、誤作動を起こす事より、全ての性質を持たせる方法は困難であると思われる。

失敗例 **1-2**：宣伝用ビデオに二億円もかけた。

出力はパターン無しとして判定される。
正解パターン：(3 主体) が (1 全て) に (2590 値・額) を掛ける。
「も」格の部分は「を」格の意味となっている。「も」格の部分を「が」格と認識しているため、パターンは検出されず、誤りとなる。二億円が主格になる事はないので登録パターンには存在しない。

失敗例 **1-3**：夕刊は 4 日から発行する。

出力はパターン無しとして判定される。
正解パターン：(3 主体) が (1109 文書) を発行する。
「は」格が「を」格の意味であり、夕刊が主格となる事は無いのでパターン無しとなる。「から」格は認識の役には立っていない。

失敗例 **1-4**：3 日も首都空爆を行った。

出力はパターン無しとして判定される。
正解パターン：(3 主体 962 機械) が (1395 思考 1002 抽象物 (精神) 1560 行為 1479 明示・暗示 2064 変動) を行なう。
「も」格が「に」格の意味を持つので誤りとなる。

失敗例 **1-5**：正月休みは、飲み食いとゴロ寝で大方を過ごした。

出力はパターン無しとして判定される。
正解パターン：(4 人) が (酒 1236 人間活動) を過ごす。
「正月休みは」の「は」格が「を」格の性質を持つので誤りとなる。1-5 の例では、「大方を」が「を」格であるため、認識が難しい。

失敗例 **1-6**：第 50 会びわ湖毎日マラソンは 3 月 5 日、滋賀県で開催します。

出力はパターン無しとして判定される。
正解パターン：(3 主体) が (1154 抽象物 (行為) 1236 人間活動) を開催する。
「は」格が「を」格の性質を持つため、誤りとなる。

失敗例 **1-7**：訪問先はヨーロッパが多い。

出力は、判定不能。正解パターン：(1 全て) が (388 場所 533 具体物 1000 抽象) に多い。
「は」格が「に」格の性質を持つので誤りとなる。

失敗例 **1-8**：天下り先でのポストは圧倒的に営業関係が多い。

出力パターン：(1 全て) が (388 場所 533 具体物 1000 抽象) に多い。
正解パターン：(1 全て) が (3 人 962 機械) に多い。
「は」格が「に」格の性質を持つので誤りとなる。

失敗例 **1-9**：伊勢参拝は首相代理で河野洋平副総理外相が行った。

出力は、判定不能。
正解パターン：(4 人) が (388 場所 2610 場)[に、へ、まで] 行く。
「は」格が「を」格の性質を持つので誤りとなる。

### 7.1.2 点数計算が不適切だった場合

失敗例 **2-1**：産廃処理工場に男の声の電話を二回受けた。

候補パターン 1：$N1$(3 主体) が $N2$(1 全て) を $N3$(3 主体)[から、より]$N4$(813 衣料 904 容器) で受ける。(正解に近い)
候補パターン 2：$N1$(1037 創作物 1670 芸 1987 出版等 1991 興行等) が $N2$(3 主体) に受ける。
パターン 1 の計算：(深さ 1*3.5)-(深さ 3*0.5)-(深さ 3*1.5)-(深さ 8*0.8)=-8.9
パターン 2 の計算：(深さ 3*3.5)-(深さ 11*0.5)=5
意味上では候補パターン 1 が正解となるが、2-1 の場合、省略が多いとみなされ、大幅に減点となるため候補パターン 2 が選出されてしまう。今回の実験の場合、格要素が多いパターンでは、大幅に減点されてしまうという欠点が残る。もちろん、点数を調整すれば改善は可能であるが、今回のパターンだけにしか対応できず、逆に対応できなくなるパターンが現れると言う弊害がある。ただ、電話を受けるという意味では 1 の方が近いというだけで、実際はさらに適したパターンを登録すべきだと考える。

失敗例 **2-2**：関根容疑者が住民表を移した。

候補パターン 1:$N1$(3 主体) が $N2$(2 具体) を $N3$(388 場所 2610 場 533 具体物)[から、より]$N4$(388 場所 2610 場 533 具体物)[に、へ] 移す。(正解)
候補パターン 2：$N1$(3 主体) が $N2$(1000 抽象) を $N3$(2051 実施) に移す。
パターン 1 の計算：(深さ 3*3)+(深さ 2*3.5)-(深さ 4*1.5)-(深さ 4*1)=6
パターン 2 の計算：(深さ 3*3)+(深さ 2*3.5)-(深さ 8*0.5)=12
2-2 の場合も格要素の省略が多いために誤ったパターンを選択している。

### 7.1.3　格の変化に対応しきれなかった場合

失敗例 **3-1**：花を咲かせる。

出力ではパターン無しと判定される。
正解パターン：(671 植物) が咲く。
使役文ではあるが、「花を咲かせる」から「花が咲く」に変わらなければならず、前述した法則では例外にあたる。よって、誤りとなる。

失敗例 **3-2**：市場の実態が把握できていない。

出力ではパターン無しと判定される。
正解パターン：(3 主体 965 作業機) が (3 主体 533 具体物 1000 抽象) を把握する。
可能の文だが、「が」格は「を」格に変化しなくてはならない。可能の文は形態素解析の品詞コードでは認識できないため (「れる、られる」は受身文として認識可能)、新たな法則を用いる必要がある。

### 7.1.4　形態素解析のミスがあった場合

失敗例 **4**：規模が均質に近い。

「均質に」の部分が「均質」と「に」というように分けて取られず、「均質に」という塊で取られるため、判定が出来ない。

## 7.2 全体的な特徴

高い正答率が得られ、単文においてはほぼ点数計算と受身、使役対応が有効である事が判明したが、全体として以下の特徴が見られた。
1) 結合価パターンは、比較的数字に関する名詞への対応は弱い。
2) 固有名詞に対しての対応はほとんど出来ない。
3) 格要素の多い文に対しては省略が多すぎると対応が難しい。
4) 登録パターンは不足である。
5) 格助詞の用法によっては対応が出来ない物がある。
6) 二重主格文 (主語が 2 つ存在する文) に関しては適応しにくい。
7) 受身に対して格変化規則が必要なパターンが多い。
7 であるが、今回使用したパターンが能動態主体であったために起こった問題で、実際は受身用のパターンを登録すれば解決は可能であると思われる。また、失敗例であるが、現在の状況から見るに、現状以上の精度向上は難しい問題となる。現在のバランス付けを変えたとしても、正答率はほぼ変わらない。単文である故に、高い正解率を誇るとも思うので、複文、重文には、法則の追加が必要である。二重主格文に関しても、別法則が必要となる。「は」格の格変化で解消できる問題と考え、今回の場合「は」格での変換ミスと同じ扱いにしてある。「は」格での変換ミスは、二重主格文か、もしくは二重主格文の省略形である文が殆どだからである。正解となった場合というのは 2 つある主格の内、前の主格を無視して選択しての正解であり、二重主格文自体への対応には至っていない。また、格変化による対応は難しいと考えている。パターン無しについては、「~する」という系統の動詞、または複数動詞を組み合わせた動詞が無い場合が目立った。「~する」については、名詞と「する」を格助詞で分けて「~をする」という様な形式にすると認識精度上がると思われる。

# 8　結論

本研究の目的は省略に対応した精度の高い結合価パターン照合方式の考案にあった。実験結果より、格の深さに注目した点数計算は、省略において出現した複数の結合価パターン候補を絞る事に対し、有効である事が分かり、精度の高い照合を可能としたと言える。しかし、複数の性質を持つ格や受身、使役での格変化の対応についてさらに改良の必要があり、特殊な文においても別法則を用いる必要がある。主に「は」格を用いる二重主格文への対応やさまざまな格に変化する「も」格に対しての対処を行えばさらに 2、3 %程度の精度向上が可能だと思われる。また、単文に対しては高い精度で正しい結合価パターンを得る事が出来たが、重文、複文については失敗パターンが増える可能性があるので将来的には対応できるようにしたい。

# 参考文献


[1] 金出地 :「 結合価パターン辞書を用いた訳語選択の精度 (2001)

[2] 白井,Francis, 野沢, 佐々木, 上田 :「 入力文と結合価パターン対辞書の照合に関する一手法 」, 言語処理学会第 5 回年次大会発表論文集, pp.80-83 (1999)

[3] 池原, 宮崎, 白井, 横尾, 中岩, 小倉, 大山, 林 : 日本語語彙大系, 岩波書店 (1997)

[4] 日外アソシエーツ株式会社 : 毎日新聞 95 年度版 CD-ROM(1995)

[5] NTT アドバンステクノロジ株式会社 : 形態素解析パッケージ ALTJAWS http://www.kecl.ntt.co.jp/icl/mtg/index-j.html

[6] 池原, 村上, 吉田 : 入力文に対する結合価パターン対の選択方法について (2002)