動詞節に修飾された名詞句の係り受け解析

竹信伸介　池原悟　村上仁一
鳥取大学工学部知能情報工学科
{takenobu,ikehara,murakami}@ike.tottori-u.ac.jp

1　はじめに

機械翻訳における問題点の一つに、言語表現の構造の曖昧性がある。日本語では特に、名詞句の構造が曖昧である。これまでの名詞句の研究としては、並列型名詞句における形容詞の係り受け解析 [1]、助詞「の」が結ぶ名詞の意味関係の解析 [2]、フレーム構造を用いた名詞句「A の B」の意味解析 [3] などがある。しかし、動詞に修飾される名詞への係り受け解析方法が提案されていない。

そこで本研究では、「動詞 + 名詞 A + の + 名詞 B」の名詞句において、動詞が名詞 A もしくは B のどちらに係るかを決定する方法を考案し、その精度を評価する。

2　結合価文法による係り受け解析

2.1　結合価文法

本研究の係り受け規則は、日本語語彙大系 [4] の結合価文法を使用する。結合価文法とは、用言と格要素の意味的関係を記述したもので、これは日本語解析で発生する意味上の多義の解消を目的をして開発されたものである。表 1 に結合価パターンの例を示す。なお本研究では以下、パターンとは結合価パターンである。

表 1: 結合価パターンの例 (括弧内は意味属性)

$N1$(4 人 2306 物象) が $N2$(4 人) を 起こす
$N1$(3 主体) が $N2$(2059 事件) を 起こす
$N1$(3 主体) が $N2$(302 番人 1476 見積り) を 立てる
$N1$(3 主体) が $N2$(533 具体物) を 立てる

2.2　一般名詞意味属性体系

結合価パターンでは、一般名詞意味属性が使用されている。一般名詞意味属性体系とは、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、図 1 のように 12 段の木構造 (属性数 約 2700) になっている。また、上位属性は下位属性を内包する。

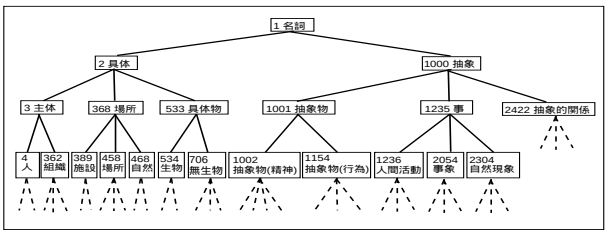

図 1: 一般名詞意味属性体系 (上位部位)

2.3　研究で使用する標本

本研究では、95 年度版毎日新聞データより「動詞 + 名詞 A + の + 名詞 B」を含む 500 文を使用する。

2.4　解析原理

文において名詞と動詞の間に格関係が成り立てば、動詞が名詞に係ると考えられる。よって本研究では、動詞に対する結合価パターンの格要素に名詞が対応する場合、動詞節の係り先と判断する。

2.5　解析の流れ

本研究の係り受け解析は、3 つの段階に分けられる。

1. 動詞に対する結合価パターンを語彙大系から抽出
2. 標本文に出てくる名詞の意味属性を語彙大系で調べ、標本文に適合する意味属性を抽出
3. 標本文がパターンに当てはまるか比較

係り受け解析の例をあげる。
例 1 : 事故で死亡した ドライバー の 遺品
例 1 の「死亡する」のパターンは P1 になる。
P1:$N1$(人) が $N2$(災難) で死亡する
例 1 に適合する名詞の意味属性を抽出すると、事故 (事件)、ドライバー (運転手)、遺品 (持ち物) となる。
例文がパターンに当てはまるか比較する。

- で格「事故」の意味属性は (事件) で、パターンので格「$N2$(災難)」の下位属性なので内包される。
- ドライバーの意味属性は (運転手) で、パターンのが格「$N1$(人)」の下位属性なので内包される。
- 遺品の意味属性は (持ち物) で、「$N1$(人)」に内包されない。

$N1$ に対応するのは「ドライバー」であり、「死亡する」は「ドライバー」に係ると判断する。

2.6 比較方法

2.5 節の 3. の比較は、基本的に結合価パターンの適合性に準じて行う。ただし、標本中では「 」に囲われる名詞は動詞節の係り先となっているので、その場合はパターンを使用しない。また、パターンが適用できない場合は意味属性で判断する。具体的には以下の 7 つの段階に分けることができる。

(1) 特殊な場合

名詞 A 又は B に「 」もしくは”“で囲われている場合、動詞節の係り先と判断する。

例 2：地方自治を考える「列島ロジー」の 第一部
名詞 A の「列島ロジー」が記号「 」で囲われているため、「考える」は「列島ロジー」に係ると判断する。

(2) 結合価パターン

例文が結合価パターンに適合するか比較する。
例 3：寄付金を扱う 自治体 の 巨大イベント
「を格」の格要素「寄付金」の意味属性は (金銭)
名詞 A の「自治体」の意味属性は (行政機関)
名詞 B の「イベント」の意味属性は (催し)
例 3 の「扱う」のパターンは P2 になる。
P2:$N1$(主体, 機械) が $N2$(*) を扱う
$N1$ の意味属性に内包されるのは名詞 A の方なので、「扱う」は「自治体」に係ると判断する。

(3) 適合パターンが複数存在し格要素数が異なる場合

当てはまる結合価パターンが複数存在することもある。その場合、パターンに含まれる格要素数が多い方の結果を優先する。例えば P3 と P4 の両方で当てはまるとする。
P3:$N1$ が $N2$ に ~ する
P4:$N1$ が $N2$ を $N3$ に ~ する
この場合、格要素が 3 つある後者の評価を結果とする。

(4) 適合パターンが複数存在し格要素数が同じ場合

当てはまる結合価パターンが複数存在し、格要素数が同じ場合は、名詞 A 又は B に対応する格要素の種類に応じて優先度を決め、優先度の高い方を結果とする。格要素の種類別優先度を図 2 に示す。

が > を に > で へ > の > と から まで より

図 2：格要素の優先度

格要素の優先度は、標本データでの頻出度に応じて作成した。例えば P5 と P6 の両方のパターンに当てはまるとする。
P5:$N1$ が $N2$ に ~ する
P6:$N1$ が $N2$ で ~ する
名詞 A 又は B が $N2$ に対応していたとすると、「に格」の方が「で格」よりも優先度が高いので、前者の評価を結果とする。

(5) 省略を含むパターン

元の結合価パターンから $N1$(が格) を除いたパターンと比較する。

例 4：収容された ロシア兵捕虜 の 姿
名詞 A の「捕虜」の意味属性は (囚人・捕虜)
名詞 B の「姿」の意味属性は (形, 姿)
例 4 の「収容する」のパターンで (1) ~ (4) の方法に当てはまらず、P7 から $N1$ を省略したパターンに当てはまる。
P7:$N1$(施設) が $N2$(主体, 動物) を収容する
$N2$ の意味属性に内包されるのは名詞 A の方なので、「収容する」は「捕虜」に係ると判断する。

(6) 全ての意味属性が * のパターン

意味属性が全て受けつけるパターンと比較する。

例 5：企業収益を維持する 経営トップ の 姿勢
「を格」の格要素「収益」の意味属性は (利益)
名詞 A の「トップ」の意味属性は (等級)
名詞 B の「姿勢」の意味属性は (態度)
例 5 では「維持する」のパターンで (1) ~ (5) の方法にあてはまらず、P8 に当てはまる。
P8:$N1$(*) が $N2$(*) を維持する
$N1$、$N2$ ともに意味属性が全て受けつける属性なので、「維持する」は「トップ」、「姿勢」の両方に係ると判断する。

(7) パターンの Nx の意味属性と名詞 **A**、**B** の意味属性が適合

結合価パターンにおいて、$N1$ の意味属性から順に名詞 A、B の意味属性と比較していく。比較する Nx の格要素の種類によって優先度を決めておき、より優先度の高い格要素での評価を結果として採用する。格要素の種類での優先度は (4) と同様である。

例えば「$N1$ が $N2$ から $N3$ に ~ する」において、「に格」$N3$ で名詞 B が当てはまったら、「から格」$N2$ で名詞 A が当てはまっても、「に格」の方が優先度が高いので、結果は名詞 B に係ると判断する。

例 6：視察に行く 議員 の 選考方法
「に格」の格要素「視察」の意味属性は (検査)
名詞 A の「議員」の意味属性は (政治家, 人 < 地位 >)
名詞 B の「方法」の意味属性は (方法)
例 6 では「行く」の結合価パターンは (1) ~ (6) の方法に当てはまらないので、各結合価パターンについて、Nx の意味属性と名詞 A、B の意味属性を比較する。この結果、P9 が選択される。

P9:$N1$(主体, 動物) が $N2$(道路) を行く
$N1$ に名詞 A が内包され、「が格」は最優先なので、「行く」は「議員」に係ると判断する。

比較評価の優先度は (1) が最も高く、(7) が最も低い。(1) の評価から順に行い、比較評価での結果が出た時点で解析は終了し、以降の比較評価はスキップする。

3 評価

2 章の係り受け解析の精度を、①解析規則を作成するのに使用した 95 年度版毎日新聞データの 500 文、②機能試験文集と基本語用例辞典の 200 文で評価する。正解は標本文を複数の人間で見て作成し、以下の 3 種類に分ける。

a. 動詞節が名詞 A に係る
b. 動詞節が名詞 B に係る
c. 動詞節が名詞 A と B の両方に係る

評価の分類は 5 つに分ける。

◎ ：結果が正解と一致する場合
○ ：結果が A で、正解が A 及び B の場合
　　結果が B で、正解が A 及び B の場合
△ ：結果が A 及び B で、正解が A の場合
　　結果が A 及び B で、正解が B の場合
× ：結果が正解と一致しない場合
＊ ：結果が出力されない場合

なお A 及び B とは動詞節の係り先が A と B の両方である。

4 結果

3 章での解析結果を表 2 と表 3 に示す。なお解析規則を用いた場合、どのくらい効果があるかを調べるため、デフォルトルール (結果を全て A とする) での評価も示す。括弧内はそれぞれの度数である。

表 2: 解析結果 (標本内テスト)

評価	① 毎日新聞データ (標本内)				
	解析規則	規則+デフォルト	デフォルト	独自パターン	独自+デフォルト
◎	48.8%(244)	71.2%(356)	52.2%(261)	84.0%(420)	91.2%(456)
○	3.2%(16)	5.6%(28)	11.4%(57)	3.4%(17)	4.0%(20)
△	19.8%(99)	0.0%(0)	0.0%(0)	7.4%(37)	0.0%(0)
×	8.2%(41)	23.2%(116)	36.4%(182)	2.0%(10)	4.8%(24)
*	20.0%(100)	0.0%(0)	0.0%(0)	3.2%(16)	0.0%(0)

表 3: 解析結果 (標本外テスト)

評価	② 機能試験文集 基本語用例辞典 (標本外)		
	解析規則	規則+デフォルト	デフォルト
◎	48.0%(92)	72.5%(145)	75.5%(151)
○	1.0%(2)	3.0%(6)	7.5%(15)
△	15.0%(30)	0.0%(0)	0.0%(0)
×	17.0%(34)	24.5%(49)	17.0%(34)
*	19.0%(38)	0.0%(0)	0.0%(0)

解析規則の結果が正解である文 (◎ 又は○) は、標本内データで約 52 %の文の名詞が正しい係り先として選択された。また、評価が△ 又は＊になった文にデフォルトルールを適用した場合、正解の文が約 77 %に上昇し、デフォルトルールのみを用いた場合より約 13 %の向上が見られた。

標本外データでは、解析規則の結果が正解である文は約 49 %であった。解析規則での評価が△ 又は＊になった文にデフォルトルールを適用した場合、正解の文が約 75 %に上昇した。しかし、デフォルトルールのみを用いた場合よりも約 8 %低い結果となった。

4.1 ◎ の例

例 7 : 寄付金を扱う 自治体 の 巨大イベント
　　(正解は名詞 A)
「を格」の格要素「寄付金」の意味属性は (金銭)
名詞 A の「自治体」の意味属性は (行政機関)
名詞 B の「イベント」の意味属性は (催し)
例 7 の「扱う」のパターンは P10 になる。
P10:$N1$(主体, 機械) が $N2$(*) を扱う
2.6 節の比較方法 (2) に当てはまり、P10 の $N1$ の意味属性に内包されるのは名詞 A の方なので、「扱う」は「自治体」に係ると判断する。正解が名詞 A で結果も名詞 A なので、評価は◎ となる。

4.2 ○ の例

例 8 : 研究調査を受託した 防災研究室 の
　　佐藤隆雄室長 (正解は名詞 A 及び B)
「を格」の格要素「調査」の意味属性は (調べ, 測定)
名詞 A の「研究室」の意味属性は (部屋)
名詞 B の「室長」の意味属性は (長)
例 8 の「受託する」のパターンは P11 になる。
P11:$N1$(主体) が $N2$(人工物, 制度) を $N3$(主体) から受託する
2.6 節の比較方法 (7) に当てはまり、P11 の $N1$ の意味属性に内包されるのは名詞 B の方なので、「受託する」は「室長」に係ると判断する。正解が名詞 A 及び B で結果が名詞 B なので、評価は○ となる。

4.3 △ の例

例 9 : 利益が出る プライスキャップ方式 の 導入
　　(正解は名詞 A)
「が格」の格要素「利益」の意味属性は (利得, 利益)
名詞 A の「方式」の意味属性は (方法)
名詞 B の「導入」の意味属性は (入れ)
例 9 の「出る」のパターンは P12 になる。
P12:$N1$(抽象) から出る
2.6 節の比較方法 (5) に当てはまり、P12 の $N1$ の意味属性に名詞 A も名詞 B も内包されるので、「出る」は「方式」、「導入」の両方に係ると判断する。正解が名詞 A で結果が名詞 A 及び B なので、評価は△ となる。

4.4 × の例

例 10：休日となる 土曜日 の 生徒指導方法
（正解は名詞 A）
「と格」の格要素「休日」の意味属性は (日, 休業)
名詞 A の「土曜日」の意味属性は (日)
名詞 B の「方法」の意味属性は (方法, 案)
例 10 の「なる」のパターンは P13 になる。
P13:$N1$(抽象物, 人間活動) がなる
2.6 節の比較方法 (2) に当てはまり、P13 の $N1$ の意味属性に内包されるのは名詞 B の方なので、「なる」は「方法」に係ると判断する。正解が名詞 A で結果が名詞 B なので、評価は× となる。

4.5 * の例

例 11：建物から銃撃する チェチェン側兵士 の 姿
（正解は名詞 A）
動詞「銃撃する」に対する結合価パターンが存在しないので、解析できない。正解が名詞 A で結果が出ないので、評価は * となる。

5 考察

5.1 独自パターン

解析規則のみの評価は、標本文に適合する既存の結合価パターンが存在しない場合や、結合価パターンが登録されていない動詞があるため、低い正解率になった。独自の結合価パターンを追加して再評価した結果、標本内テストの解析規則の正解率は約 87 %、解析規則にデフォルトルールを適用する方法では約 95 %に上昇した。

例 12：建物から銃撃する チェチェン側兵士 の 姿
（正解は名詞 A）
動詞「銃撃する」の結合価パターンが存在しないため、P14 を追加する。
P14:$N1$(主体) が $N2$(場所) から銃撃する
「から格」の格要素「建物」の意味属性は (施設)
名詞 A の「兵士」の意味属性は (兵卒, 軍人)
名詞 B の「姿」の意味属性は (形, 姿)
・「建物」の意味属性は $N2$ の意味属性に内包される。
・「兵士」の意味属性は $N1$ の意味属性に内包される。
・「姿」の意味属性は $N1$ の意味属性に内包されない。
よって「銃撃する」は「兵士」に係ると判断する。正解は名詞 A で、結果も名詞 A なので、評価は◎ となる。

5.2 標本外データ

4 章で述べたように、標本外テストでは解析規則よりデフォルトルールでの正解率が高くなる結果になった。これは、基本語用例辞典より抽出した標本文に、格要素の存在しないデータが多かったため、結合価パターンに適合せず、間違った結果が出たためである。

例 13：死んだ 子供 の 事（正解は名詞 A）
名詞 A の「子供」の意味属性は (子)
名詞 B の「事」の意味属性は (事)
例 13 に適合する「死ぬ」の結合価パターンが無く、P15 から $N1$ を省略したパターンに当てはまる。
P15:$N1$(人) が $N2$(事) で死ぬ
$N2$ の意味属性に内包されるのは名詞 B の方なので、「死ぬ」は「事」に係ると判断する。正解は名詞 A で結果は名詞 B なので、評価は× となる。

また、標本外テストにおいても、標本内テストと同様に独自の結合価パターンを登録する事で、正解率の上昇が期待できる。

6 結論

本研究において、結合価パターンを使用することにより、ある程度の精度で動詞が名詞 A もしくは B のどちらに係るかを選択できることが分かった。また、既存の結合価パターンでは全ての文をカバーするには不十分であると分かった。
今後の課題として 2 つあげられる。

(1) 独自パターンが追加できない文への考慮

動詞に対する既存の結合価パターンが存在しない場合や、文に適した結合価パターンがない場合については独自パターンを追加した。しかし、動詞「対する」のように、動詞と名詞との間に格関係が存在しない場合、独自パターンの追加ができない。名詞の意味的関係について調べ、考慮する必要があると考えられる。

(2) 格要素の無い文への規則作成

5.2 節で述べたように、文に格要素が存在しない場合、結合価パターンに正しく適合しない事がある。格要素が存在しない文への新たな規則を作成する必要があると考えられる。また、2.6 節の比較方法 (5) では間違った結果が出力される事がある。結合価パターンに完全に適合しない場合の新たな規則を作り直す必要があると考えられる。

参考文献

[1] 大西, 池原, 村上：並列型名詞句における形容詞の係り受け解析, 鳥取大学卒業論文 (Feb. 1999)

[2] 島津：助詞「の」が結ぶ意味関係の解析, 計量言語学会 15 巻第 7 号,pp.247-266 (1986)

[3] 横山：フレーム構造を用いた名詞句「A の B」の意味解析, 電子情報通信学会技術研究報告書,NLC97-3,pp.17-24 (1997)

[4] 池原, 宮崎, 白井, 横尾, 中岩, 小倉, 大山, 林：日本語語彙体系 1．意味体系, 岩波書店 (1997)