社団法人 情報処理学会　研究報告
IPSJ SIG Technical Report

2004－ＦＩ－ 76　(12)
2004－ＮＬ－163　(12)
2004／9／17

動詞節に修飾された名詞句の係り受け解析

竹信伸介† 徳久雅人†
村上仁一† 池原悟†

本稿では「動詞節＋名詞 A ＋の＋名詞 B」型名詞句において，動詞節が名詞 A と名詞 B のどちらに係るかを決定する手法を提案する．本手法は，5 つの方式を成功率の高い順に適用する形態をとる．特に，動詞が名詞を修飾する際，被修飾語が動詞の格関係にあることに着目して，「結合価文法」及び「格要素と動詞の共起情報」を利用した点に特徴がある．新聞記事における当該名詞句について解析実験を行ったところ，クローズドテストで 89%，オープンテストで 84%の正解率であった．本手法は正解率が 57%というベースライン評価を大きく上回ることから，有効性が確認できた．

The dependency analysis of the noun phrase modified by the verb clause

Shinsuke Takenobu,† Masato Tokuhisa,† Jin'ichi Murakami†
and Satoru Ikehara†

This paper proposes a dependency analysis method to select which noun in such a noun phrase as "verb clause + noun A + *no* + noun B" is modified by the verb clause. This method contains 5 sub-methods that are applied in the sequence of their precision. Especially, a valency grammar based sub-method and a collocation based sub-method between a case and a verb are included, since the modified noun is related to the verb as a case element. The accuracy of the closed test is 89% and that of the open test is 84%, whose test data are collected from news paper. Comparing that the accuracy of the baseline test is 57%, this method is effective.

1. はじめに

日本語文の構文解析において，名詞句の構造の曖昧性解消が大きな問題点の一つとなっている．

これまで，「の」で名詞を接続した「名詞 A の名詞 B」型名詞句に関する構文解析は，先行する品詞毎に判断する方法が提案されてきた．「名詞 N の A の B」型名詞句における名詞 N の係り先を決定する問題に対して，益田らは，名詞を 13 種類の品詞に下位分類し，一般的にそれらの品詞が「の」の左右に現れる頻度を求め，この頻度を接続強度として係り先の判定条件とした[1]．中井らは，係り受け関係にある 2 つの名詞の意味属性をマトリクスでとらえて判定条件とした[2]．「形容詞 AJ+A の B」型名詞句において形容詞 AJ の係り先を決定するために，森内らは，形容詞と名詞の意味属性の組み合わせ頻度を用いて判定した[3]．白井らは，コーパスから自動学習により，単語間の意味的な制約を反映した決定リストを生成して用いた[4]．「形容動詞 AJV+A の B」型名詞句では，美野らが同じく決定リストを用いた方法を提案した[5]．

このように「の」型名詞句の関連研究は多い．しかし，動詞節に修飾される名詞句，すなわち，「動詞節 V+A の B」型名詞句における有効な係り受け解析方法は提案されていない．そこで，本稿では「V+A の B」型名詞句において，動詞節 V が名詞 A もしくは名詞 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ことを目的とする．

動詞節が名詞を修飾する際の特徴として，格関係が挙げられる．それは「内の関係」と「外の関係」と呼ばれ，前者は動詞節と係る名詞との間に格関係のある場合，後者は動詞節と係る名詞との間に格関係のない場合として区別される[6]．したがって，従来研究で用いられた名詞の意味属性および単語の出現頻度とは異なる手がかりとして，「結合価文法」および「格要素と動詞の統計的共起情報」が有効であると予想される．

そこで，結合価文法，格要素と動詞の共起情報，名詞の意味属性，および，表記上の特徴を手がかりとする判断方式を 5 つ作成し，それらを組み合わせて，「V+A の B」型名詞句の係り受け解析を行う手法を構築する．

本稿の構成は次のようになる．まず，第 2 章では本稿で用いる標本データについて述べる．次に，第 3 章で係り受け解析の 5 つの方式を提案し，第 4 章でこれらの方式を統合する．第 5 章で評価実験を行い，第 6 章で考察を述べ，第 7 章で今後の課題をあげる．最後に第 8 章でまとめを述べる．

2. 対象とする名詞句と標本データ

2.1 動詞節に修飾された名詞句

本稿では，「動詞節 V ＋名詞 A の名詞 B」型の名詞句を対象とする．なお，以降ではこの型の名詞句を単に名詞句と呼

† 鳥取大学工学部
Faculty of Engineering, Tottori University
{takenobu,tokuhisa,murakami,ikehara}@ike.tottori-u.ac.jp

ぶ．以下に例を示す．

例 1：住宅を失った被災者の支援
例 2：廃棄物を積んだ英国の輸送船

例 1 では動詞節「住宅を失った」は名詞 A「被災者」を修飾し，例 2 では動詞節「廃棄物を積んだ」が名詞 B「輸送船」を修飾している．本稿では，動詞節 V が名詞 A に係るものを「A 係り」，名詞 B に係るものを「B 係り」と呼ぶ．

2.2 標本データ

本稿では，95 年度版毎日新聞[7] から標本データを抽出する．抽出した「V + A の B」の表現 1,000 件をクローズドデータとして，同じく 400 件をオープンデータとして使用する．ただし，表現の幅が広い動詞「ある」,「する」☆,「なる」を含む名詞句，および，修飾する動詞節が受け身になっている名詞句は本稿の対象外とする．

クローズドデータ，オープンデータともに，正解ラベル（「A 係り」/「B 係り」），および，名詞の意味属性（後述）を付与する．正解ラベルは，標本を複数の人間が見て判断し，判断の異なる点は協議して一方に定める．意味属性は，後に提案する判定方式の入力情報となる．本稿では，意味属性の解析は対象外であるので，意味属性の曖昧性解消が終了した段階の情報を入力する．

クローズドデータの正解ラベルの割合を調べたところ,「A 係り」が 505 件,「B 係り」が 495 件であり,「A 係り」が多かった．したがって，本解析タスクのベースライン評価は,「A 係り」の判定で正解率が 50.5%となる．第 3 章以降で，解析方式を提案するが，いずれの方式でも判定不能の場合がある．本稿では，その場合「A 係り」と決定するものとし，これを「デフォルトルール (DR)」による決定という☆☆．

3. 係り受け解析方式

本章では，5 種類の係り受け解析方式を提案する．主に格関係に関して 2 つの解析方式を 3.1 節および 3.2 節で述べる．次に，格関係では判断しづらい場合に備え，名詞句の事例分析により得られた解析方式を，3.3 節から 3.5 節にかけて述べる．

3.1 結合価文法による係り受け解析

動詞節と内の関係にある名詞は，動詞節の係り先である．内の関係は動詞と名詞の間の格関係から判定される．格関係は，動詞の語義ごとに格要素にとりうる名詞を制限する結合価文法を用いて解析される．よって本節では，結合価文法を用いて格関係の認められる名詞を係り先と判定する方式 (VPM 方式:Valency Pattern Matching 方式) を提案する．

☆ 「〜をする」や「〜とする」など「する」が文法上の主動詞となる場合は対象外とするが,「サ変名詞＋する」(勉強する等) は対象とする．

☆☆ 文献 1) によると,「A の B の C」型名詞句で，名詞 A が名詞 B に係る割合が約 73%，名詞 A が名詞 C に係る割合が約 27%と報告されている．本稿で扱う標本は,「A 係り」と「B 係り」の割合が，ほぼ 1:1 であり，デフォルトルールによる係り先決定の精度は期待できない．

3.1.1 結合価文法と一般名詞意味属性

本稿は，結合価文法を用いるために，日本語語彙大系[8] から，文型パターン（結合価パターンという）および一般名詞意味属性を使用するので，ここで概説する．

一般名詞意味属性は名詞の語義を表すラベルである．一般名詞意味属性体系では，約 2,700 種類が定義され図 1 のように 12 段の木構造のノードに対応している．木構造のノードは上位のノードの語義を継承している．単語は，最下位のノードに限らず，適切な抽象度の語義を表すノードに対応し，また，複数の語義を持つ単語は，複数のノードに対応する．

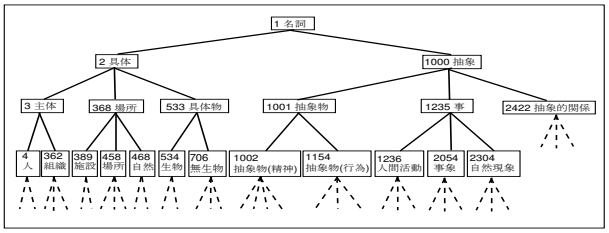

図 1 一般名詞意味属性体系 (上位部位)

結合価パターンとは，述語の意味的用法を区別するために，共起する格要素に一般名詞意味属性による制約を与えた文型パターンである．構文体系では，約 14,800 種類の結合価パターンが定義されている．表 1 に具体例を示す．

表 1 結合価パターンの例 (括弧内は意味属性)

*N*1(人 物象) が *N*2(人) を 起こす
*N*1(主体) が *N*2(事件) を 起こす
*N*1(主体) が *N*2(番人 見積り) を 立てる
*N*1(主体) が *N*2(具体物) を 立てる

表 1 に示すように，一つの動詞に対して，複数の結合価パターンが存在する．例えば,「母が息子を起こす」という文は,「*N*1(人) が *N*2(人) を 起こす」のパターンに適合する．これは,「が格」の「母」の意味属性が，パターンの「が格」の意味属性 (人) の下位属性であり,「を格」の「息子」の意味属性が，パターンの「を格」の意味属性 (人) の下位属性となるからである．本稿では，x が y の下位属性になることを「y が x を内包する」と呼ぶ．

3.1.2 VPM 方式の解析手順

「V+A の B」型名詞句が与えられると次の手順で解析を行う．

手順 1: V に適合する結合価パターンを構文体系から抽出する．

手順 2: 一般名詞意味属性体系から A および B の意味属性を求める．

手順 3: 1 の結合価パターンの格要素に，2 で求めた意味属性を持つ名詞が代入可能であるか判定する．

手順 4: 制約を満たす名詞が A ならば「A 係り」，B ならば「B 係り」となり，いずれも満たすならば「AB 係り」，いずれも満たさないならば「判定不能」となる．

以下に具体例を示す．

例 3：事故で**死亡した** ドライバー の 遺品

手順 1. 「事故で死亡する」と「N1(人) が N2(災難) で死亡する」が適合する．

手順 2. A=「ドライバー (運転手)」，B=「遺品 (持ち物)」となる．

手順 3. N1(人) に代入可能なのは A であり，B は不可能であることがわかる．

手順 4. 「A 係り」と判定する．

3.1.3 VPM 方式のトレーニング

通常，動詞節に対応する結合価パターンは複数存在するため，VPM 方式では適合パターンの選択が問題となる．そこで，本稿は，吉田らの方法を採用して，適合パターンの選択を行う[9]．

9) の方法は，結合価パターンの格要素の使われ方，および意味属性の制約の深さに着目して適合度を計算する方法で，入力の 1 文に適合する複数の結合価パターンに対して，適合度に則した順位付けを行うことができる．

具体的にまず，格要素の使われ方については，格要素の種類ごとに，格要素が使われた場合の加点，および，使われなかった場合の減点，という 2 つの基準値を使う．そこで本稿では，クローズドデータ 1,000 件を用いて，動詞節の係り先が正しく判定できるように，加点/減点の基準値を設定した．表 2 にその結果を示す．

表 2 格の種類による得点付け

	が格	を格	に格, で格	の格	と格	その他
加点	5.0 点	4.0 点	3.5 点	3.3 点	3.0 点	2.5 点
減点	1.5 点	1.2 点	1.0 点	0.8 点	0.6 点	0.5 点

次に，意味属性制約の深さについては，結合価パターンで深いレベルの意味属性が制約である場合は，その格要素に代入可能な名詞の具象度が高く，述語の語義を定める影響度が高いといえるため，重要視すべきであるという考えがなされている．

それでは以下に適合度の計算例を示す．

例 4：首相に謝罪を**求める** 農民 の 手紙

例 4 に対し P4.1,P4.2 の 2 つのパターンがあるとする．

P4.1:$N1$**(主体) が** $N2$**(抽象) を** $N3$**(主体) に 求める**

P4.2:$N1$**(主体) が** $N2$**(助力, 援助) を** $N3$**(主体) に/へ 求める**

上の 2 つのパターンにおいて，例 4 の「に格」の「首相」はそれぞれのパターンの $N3$ に内包され，「を格」の「謝罪」は P4.1 の $N2$ に内包されるが，P4.2 の $N2$ には内包されない．さらに名詞 A「農民」はそれぞれのパターンの $N1$ に内包されるので，この各パターンの適合度は次のように計算する．ここで，D_A とは属性 A の深さを示す．

[P4.1 の適合度]

$= (D_{\text{主体}} * 5.0) + (D_{\text{抽象}} * 4.0) + (D_{\text{主体}} * 3.5)$

$= (3 * 5.0) + (2 * 4.0) + (3 * 3.5)$

$= 33.5$

[P4.2 の適合度]

$= (D_{\text{主体}} * 5.0) - (((D_{\text{助力}} + D_{\text{援助}})/2) * 1.2) + (D_{\text{主体}} * 3.5)$

$= (3 * 5.0) - (((9 + 9)/2) * 1.2) + (3 * 3.5)$

$= 14.7$

P4.1 の方が高いので，例 4 では P4.1 を選択する．

3.2 動詞と名詞の共起に注目した係り受け解析

VPM 方式では，結合価パターンに登録されていない動詞が出現した場合，解析ができない．また，パターンが存在しても，標本に適合するとは限らない．この問題に対して，大量のデータに基づいた統計的手法でカバーする方法が考えられる．

そこで，格要素の名詞と動詞の共起頻度は，通常の文からあらかじめ作成するものとして，「V+A の B」型名詞句において，V と A の共起する頻度と，V と B の共起する頻度を比較して，頻度の高い方を係り先と判定するという方式（VCC 方式:Verb and Case Collocation 方式）を提案する．

3.2.1 共起情報の作成

共起情報を作成するために，第 2 章で述べた 95 年度版毎日新聞データ（約 100 万文）を使用する．新聞記事の文において，「格要素（名詞＋格助詞）の連続と動詞」となる部分から，名詞と動詞の組を抽出する．したがって，2.2 節で集めた 1,400 件の名詞句は使用されない．

ここで，動詞と名詞の共起情報を字面の一致で集めると，基本的な動詞の数（約 6 千件）および名詞の数（約 6 万件）から考えて，現在のところ十分なコーパスが得られるとは断言しがたい．そこで，名詞については，一般名詞意味属性に抽象化する．

こうして，名詞の一般名詞意味属性と動詞の組を共起情報として集計する．この組の集合をを共起データベースと呼ぶ．

表 3 に共起データベースの例を示す．

表 3 共起データベース (一部)

動詞	共起する名詞の意味属性
読む	息子, 本, 童話, ...
書く	手紙, 友人, ...
見る	住民, 犯人, ...
...	...
	(合計 9700 件)

3.2.2 VCC 方式の解析手順

「V+A の B」が与えられると VCC 方式では次の手順で

解析する．

手順 1: 名詞 A および B の意味属性を一般名詞意味属性体系から検索する．

手順 2: V の動詞と A の意味属性の共起頻度，および，同じく B の共起頻度を共起データベースから検索する．共起データベースの検索条件は，「A や B の意味属性が共起データベースの意味属性に内包される」，「共起データベースの意味属性が A や B の意味属性に内包される」，または，「A や B の意味属性が共起データベースの意味属性と一致する」とする．

手順 3: 手順 2 で得た頻度を比較して，高い方を係り先として出力する．ただし，A と B の両方とも頻度が 0 である場合は「解析不能」と出力する．

3.2.3 VCC 方式のトレーニング

3.2.1 項で作成した共起データベースは，クローズドデータを満足しているとはいえない．そこで，クローズドデータの動詞節に含まれる動詞と，動詞節の係り先となる名詞の組み合わせを，共起データベースに追加する．

3.3 特殊記号を手がかりとした係り受け解析

新聞記事では，名詞が特殊記号などで強調されている場合，その名詞が係り先となることが多い．よって「」もしくはダブルコーテーション (”) で囲まれている名詞を，動詞節の係り先と判断する方式（IPS 方式:Inclusion by Particular Symbol 方式）を導入する．クローズドデータ 1,000 件の分析による，特殊記号の出現回数と正解数を表 4 に示す．

なお，特殊記号が使われていない場合は，「解析不能」と出力する．

表 4 特殊記号の出現回数と正解数

#	特殊記号	出現回数	正解数
1	「」	55	50
2	””	2	2

次の例では，名詞 A「ネオン」が「」で囲われているため，動詞節「K 社が販売する」は「A 係り」と判定する．

例 5 : K 社が**販売する** 「ネオン」 の 開発

3.4 特定の意味属性を根拠とした係り受け解析

名詞の意味属性が「時間」や「上下」などの下位属性になっている場合，その名詞が係り先にならないことが多い．よって名詞 A または B が，特定の意味属性の下位属性となる場合，もう一方の名詞を動詞節の係り先と判断する方式（EPA 方式:Exclusion by Particular Semantic Attribute 方式）を導入する．クローズドデータ 1,000 件の分析によると，表 5 に示す 9 つの属性があげられる．

なお，名詞 A,B ともに表 5 に示す属性に内包されない場合は，「解析不能」と出力する．

次の例では，名詞 A「50 周年」の意味属性が，意味属性（時間）に内包されるため，動詞節「二日にハワイで行う」は「B 係り」と判定する．

表 5 係り先とならない意味属性

#	属性	出現回数	正解数
1	時間	35	34
2	数量	152	150
3	風・観・姿	13	13
4	内外	33	32
5	上下	15	15
6	左右	1	1
7	遠近	5	5
8	色	2	2
9	過不足	2	2

例 6 : 二日にハワイで**行う** 50 周年 の 記念式典

3.5 名詞 AB 間の意味関係による係り受け解析

名詞 A と B の意味属性の組み合わせによっては，動詞の種類に関わらず係り先が判定できる場合がある．よって，特定の意味属性（下位属性含む）の組み合わせにより，動詞節の係り先を判定する方式（AAC 方式:Attribute-Attribute Collocation 方式）を導入する．属性の組み合わせの種類と係り先のルールをクローズドデータ 1,000 件の分析に基づき，44 種類作成した．表 6 に一部の例を示す．

なお，名詞 A と B の意味属性の組み合わせが，表 6 に示すような組み合わせとならない場合は，「解析不能」と出力する．

表 6 属性の組み合わせと動詞節の係り先 (一部)

#	名詞 A	名詞 B	係り先	出現回数	正解数
1	人間 (親族関係)	人間 (親族関係)	B	2	2
2	人 (職業・地位・役割)	敬称	B	14	14
3	組織	人 (職業・地位・役割)	B	34	22
4	場所	施設	B	17	17
	…	…			
44	製造	人工物	B	2	2

次の例では，名詞 A「主将」の意味属性が，意味属性（人 (職業・地位・役割)）に内包され，名詞 B「鶴岡君」の意味属性が，意味属性（敬称）の内包されるので，動詞節「バッテリーを組む」は「B 係り」と判定する．

例 7 : バッテリーを**組む** 主将 の 鶴岡君

4. 解析方式の統合

本稿では，5 つの解析方式 (VPM, VCC, IPS, EPA, AAC) を統合して解析する手法をとる．そこで，クローズドテストにおける各方式の個別正解率を求め，方式を用いる優先順位を定める．そして，オープンテストにおいては，定めた順位で方式を用いて最終的な解析結果を出力する．4.1 節で各方式の解析性能を求め，4.2 節に解析順序を示す．

4.1 各方式の解析性能

各方式の係り受け解析結果の評価の分類は，係り受け解析結果と正解ラベルの組み合わせによるもので，次の 4 つに分

かれる．

○： 解析結果が一つに定まり，正解ラベルと一致する場合
△： 解析結果が「A 係り」および「B 係り」であり，正解ラベルが「A 係り」の場合，もしくは，
　解析結果が「A 係り」および「B 係り」であり，正解ラベルが「B 係り」の場合
×： 解析結果が正解ラベルと一致しない場合
＊： 解析不能の場合

なお,「A 係りおよび B 係り」とは,「A 係り」と「B 係り」の両方を解析結果として出力した場合である．

解析精度は，カバー率と成功率で表すことにする．カバー率は，入力に対して，解析方式が適用され，係り先が 1 つに絞り込まれた割合，すなわち，評価○または評価×の割合である．成功率は，係り先が 1 つに絞り込まれた場合の成功の割合，すなわち，評価○または評価×における評価○の割合である．

クローズドデータ 1,000 件に対する各解析方式の解析結果を表 7 に示す．

表 7 個別解析結果

方式	評価○	評価△	評価×	評価＊	カバー率	成功率
VPM	54.2%	13.3%	6.5%	26.0%	60.7%	89.3%
VCC–	63.7%	0.9%	27.9%	7.5%	91.6%	69.5%
IPS	5.2%	-	0.5%	94.3%	5.7%	91.2%
EPA	25.4%	-	0.4%	74.2%	25.8%	98.4%
ACC	30.2%	-	5.1%	64.7%	35.3%	85.5%
(VCC	75.0%	1.1%	23.9%	0.0%	98.9%	75.8%)

ここで，VCC–方式とは，3.2.3 項でのトレーニング結果を反映していない場合である．つまり，クローズドデータ 1,000 件において正解ラベルの示す係り先の名詞と動詞節の動詞の共起関係を共起データベースに登録していない段階の評価結果である．本章では，各方式を統合する順番を決定するために評価を行っている．3.2.3 項のトレーニング後の VCC 方式は，事例に exact マッチするため，成功率の変動が激しいことが予想される．よって，オープンテストにおいて同様の値が期待できないので，方式の統合順序を検討するためには用いない．なお，参考として，VCC 方式の結果も掲載する．

4.2 解 析 順 序

解析方式の統合は，5 つの方式を成功率の高い順に適応する形態をとる．具体的には，ある段階の方式で係り先が 1 つに断定できるならばその係り先を解析結果として出力し，そうでない場合は，次の段階の方式で判定する．そして，5 つの方式のいずれも 1 つに断定できない場合には，デフォルトルール（2.2 節参照）に従い,「A 係り」として出力する（図 2）．

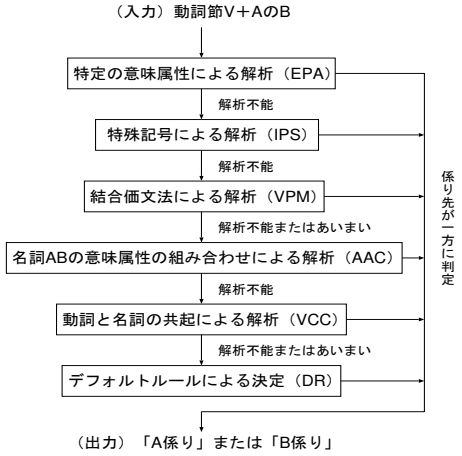

図 2 係り受け解析の順序

5. 実　　験

5.1 実験の目的および方法

第 3 章で示した方式を第 4 章のとおり統合した解析手法の解析精度の評価を目的とする．比較のため，5 つの方式およびデフォルトルールを統合した手法を ALL6，5 つの方式のみを統合した手法を ALL5，および，デフォルトルールのみによる判定手法を DR と称して実験する．

評価対象は，第 2 章で述べた新聞記事データのオープンデータ 400 件である．比較のため，クローズドデータ 1,000 件を統合した解析手法の精度も求める．

評価方法は，提案手法に「V+A の B」を入力し「A 係り」および「B 係り」を自動で判別させ，4.1 節と同じ評価の分類を行う．そして，係り先が一意に決まらない場合は誤りとみなし，正解率を計算する．

5.2 解　析　例

5.2.1 5 方式により係り先が断定できた場合

○ の 例

例 8：関西国際空港に**着陸した** シンガポール発 の 航空機
(正解：B 係り)
「に格」の格要素「空港」の意味属性は (空港)
名詞 A の「発」の意味属性は (出発)
名詞 B の「航空機」の意味属性は
　　(乗り物 (本体 (移動 (空圏))))
例 8 で選択されるパターンは P8 になる．

P8:*N*1(乗り物, 人) が *N*2(場所, 場) に/へ着陸する

- 「に格」の「空港」の意味属性は (空港) で，パターンの「に/へ格」の *N*2(場所) に内包される．
- 「発」の意味属性は (出発) で，パターンの「が格」の *N*1(乗り物, 人) に内包されない．
- 「航空機」の意味属性は (乗り物 (本体 (移動 (空圏)))) で，*N*1(乗り物) に内包される．

*N*1 に対応するのは「航空機」であり，解析結果は「B 係り」となる．正解も「B 係り」であるので，評価は○となる．

× の 例

例 9：同誌に**抗議した**「生活保障連絡会議」の 三沢代表
(正解：B 係り)

IPS 方式により,「」の付いている名詞 A が優先され,「A 係り」と判定される．正解は「B 係り」であるので，評価は×となる．

5.2.2 デフォルトルールで判定を補助した場合

△ の 例

例 10：昨年暮れに**起きた** 三陸はるか沖地震 の 余震
(正解：A 係り)

「に格」の格要素「暮れ」の意味属性は (終り)
名詞 A の「地震」の意味属性は (地震)
名詞 B の「余震」の意味属性は (地震)
例 10 で選択されるパターンは P10 になる．

P10:$N1$(事) が $N2$(主体) に起きる

- 「に格」の「暮れ」の意味属性は (終り) で，パターンの「に格」の $N2$(主体) に内包されない．
- 「地震」の意味属性は (地震) で，パターンの「が格」の $N1$(事) に内包される．
- 「余震」の意味属性は (地震) で，$N1$(事) に内包される．

$N1$ に「地震」も「余震」も対応してしまい，他の解析方式でも係り先を絞れないので，解析結果は「A 係りおよび B 係り」である．正解は「A 係り」であるので，評価は△となる．

＊ の 例

例 11：議場で**うたた寝する** 国会議員 の 写真(正解：A 係り)

動詞「うたた寝する」に対する結合価パターンが存在しない為，解析できない．他の解析方式でも係り先を判定できないので，結果は解析不能となる．正解が「A 係り」で結果が出ないので，評価は＊となる．

5.3 実 験 結 果

実験結果を表 8 にまとめる．本稿の提案手法 (ALL6) の正解率は，デフォルトルール (DR) による判定より 27.2%上回る結果となった☆.

オープンテストにおいて，一意に係り先を断定できない，もしくは，5 つの方式では解析不能となった名詞句が，オープンデータの 400 件中，10 件であった．ALL6 と ALL5 の差は 1.2% とデフォルトルールによる判定の依存度が低いことがわかる．

また，クローズドテストにおける正解率と比較すると 4.4% の差である．デフォルトルールのみによる判定の差は 6.5%であることをみると，本手法は安定した解析が行われていることが伺える．

以上より，本手法の有効性が確認できた．

☆ ベースライン評価の参考として，汎用的な係り受け解析ツールである cabocha[10] を用いて，クローズドデータ 1,000 件を対象に解析実験を試みたところ，正解率は 59.6%であった．なお，cabocha のトレーニングは行っていない．

表 8 各手法の正解率

対象	手法	正解率
open	ALL6	84.2%
open	ALL5	83.0%
open	DR	57.0%
closed	ALL6	88.6%
closed	ALL5	87.7%
closed	DR	50.5%

6. 考 察

本提案手法の方式限界について考察する．まず，本手法では，カバー率と成功率をみると，結合価文法による解析方式 (VPM) の性能が重要であるので，結合価パターンの追加を試みる．次に，結合価パターンでは原理的に解析できない事例を示す.

6.1 結合価パターンの追加

日本語語彙大系に収録されている結合価パターンの意味属性は，相互矛盾の無いように付与されており，安易に変更してしまうと，他への影響が懸念される．そこで，結合価パターンを新たに追加して，カバー率の向上をねらう．

クローズドテスト（表 7）において一意に解析できなかった 393 件の標本データに関して結合価パターンの追加を試みた．たとえば，5.2 節の例 11 に対しては，動詞「うたた寝する」の結合価パターンとして，P11 を追加した．

P11:$N1$(人) が $N2$(場所, 場) でうたた寝する

「で格」の格要素「議場」の意味属性は (席)
名詞 A の「国会議員」の意味属性は (政治家)
名詞 B の「写真」の意味属性は (写真・画像)

- 「議場」の意味属性は $N2$ の意味属性に内包される．
- 「国会議員」の意味属性は $N1$ の意味属性に内包される．
- 「写真」の意味属性は $N1$ の意味属性内包されない．

よって「うたた寝する」は「国会議員」に係る「A 係り」と判断する．

このように，203 件のパターンを新たに作成することで，393 件中 245 件の標本がカバーできた．

こうして強化した結合価パターンによる解析方式 (VPM+) のみによるクローズドテスト，および，統合手法 (ALL6+) によるクローズドテスト，および，オープンテストを行ったところ表 9，表 10 の結果を得た．

表 9 追加パターンを使用した VPM 方式の結果

方式	評価○	評価△	評価×	評価＊	カバー率	成功率
VPM+	78.7%	15.8%	2.9%	2.6%	81.6%	96.4%

6.2 結合価パターンでは解析できない標本

名詞 A と名詞 B の意味属性に違いが見られない場合，結合価パターンでの解析が困難だった．

表 10 追加パターンを使用した正解率

対象	手法	正解率
open	ALL6	84.2%
closed	ALL6	94.0%

以下にパターン追加できない標本の例をあげる．

例 12：昨年暮れに**起きた** 三陸はるか沖地震 の 余震
(正解：A 係り)
「に格」の格要素「暮れ」の意味属性は (終り)
名詞 A の「地震」の意味属性は (地震)
名詞 B の「余震」の意味属性は (地震)
名詞 A と B の意味属性が同じであるので，どんなパターンを追加しても，評価は△となってしまうため，パターンを追加することができない．

7. 今後の課題

7.1 新方式の要請

AAC 方式や EPA 方式は，動詞節に依存せずに係り先を解析する方式である．特に AAC 方式では，3.5 節の表 6 に示したように，類似する意味属性が並んだときにでも係り先を識別することに成功している．しかし，クローズドデータの 1,000 件についてでしか分析をしていないので，より多くの標本の分析に基づき，意味属性の条件を洗練する余地が残っている．

さらに，6.2 節で示した問題点に対しては，AAC 方式も EPA 方式も名詞 A,B の意味属性を手がかりとしているので，例 12 のような事例においていつも正しく判定できるという原理的な根拠に乏しい．そこで，今後は，意味属性とは異なる観点で名詞 A,B 間の関係を識別する方式が望まれる．

7.2 対象外への対応

本稿では，2.2 節で述べたとおり次の 2 点について対象外として，標本に含めなかった．

- 「ある」,「する」,「なる」を主動詞とする場合
- 「受身形」の場合

今後は，このような対象外とした標本に対し，今回のルールがどの程度有効か，検証する必要がある．

また，解析のための入力は，次の条件を課している．

- 名詞 A および名詞 B の意味属性は，一意に決定されている

しかし，実際には，はじめから意味属性が一意に決定されているわけではない．

オープンテストにおいて，意味属性が一意に決定されていない段階での情報を入力とした場合，正解率は 81.2%となり，意味属性が一意に決定された場合と比べ，正解率が 3%低かった．より精度の高い解析を行うためにも，名詞の意味属性を自動的に絞り込む方法を考案する必要がある．

8. おわりに

本稿では,「動詞節 V ＋名詞 A ＋の＋名詞 B」名詞句において，動詞節の係り先が名詞 A と B のどちらになるのか，自動的に判定するため，結合価文法，格要素と動詞の共起情報，名詞の意味属性，および，表記上の特徴を手がかりとする判断方式を 5 つ作成し，それらを組み合わせて，係り受け解析を行う手法を構築した．

解析方式を作るクローズドデータとして，95 年度版毎日新聞から抽出した「V+A の B」型名詞句 1,000 件を利用した．また，評価用のオープンデータとして，同じく 95 年度版毎日新聞から抽出した「V+A の B」型名詞句 400 件を利用した．考案した 5 つの解析法式を組み合わせて，標本データに適用したところ，クローズドテストでは約 89%，オープンテストでは約 84%の正解率となり，デフォルトルールに比べて 30%前後の向上が見られ，本手法の有効性が確認された．

参 考 文 献

1) 益田裕也, 宮崎正弘 : 名詞間の接続強度を用いた「の」型名詞句構造解析, 言語処理学会第 9 回年次大会, pp.238-241(2003).
2) 中井慎司, 伊藤真樹, 池原悟, 白井諭 : 名詞間係り受け解析に必要な単語意味属性の組の最適化, 情報処理学会第 57 回全国大会, Vol.2, pp.233-234(1998).
3) 森内昭雄, 中井慎司, 池原悟, 大西真理子 :「の」型名詞句に対する形容詞の係り先解析, 情報処理学会第 57 回全国大会, Vol.2, pp.237-238(1998).
4) 白井清昭, 橋本泰一, 西館耕介, 徳永健伸, 田中保積 : 決定リストを利用した形容詞の修飾先の決定, 言語処理学会第 7 回年次大会, pp.253-256(2001).
5) 美野秀弥, 橋本泰一, 徳永健伸, 田中保積 : 決定リストを利用した形容動詞の修飾先の決定, 言語処理学会第 8 回年次大会, pp.411-414(2002).
6) 寺村秀夫 : 日本語シンタクスと意味 I〜III, くろしお出版 (1982〜1991).
7) 毎日新聞社 : CD-毎日新聞'95 データ集, 日外アソシエーツ (1996).
8) 池原悟, 宮崎正弘, 白井諭, 横尾明男, 中岩浩巳, 小倉健太郎, 大山芳史, 林良彦 : 日本語語彙大系 1. 意味体系, 岩波書店 (1997).
9) 吉田真司, 池原悟, 村上仁一 : 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, B2-6, pp299-302 (2002).
10) 工藤拓, 松本裕治 : チャンキングの段階適用による日本語係り受け解析, 情報処理学会論文誌, pp1834-1842 (2002).