概要

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 6 つの方式を解析成功率の高い順に適用する形態をとる．その 1 つの方式は,「結合価文法」に基づく方式である．動詞が名詞を修飾する際に被修飾語が動詞の格関係にあること，すなわち,「内の関係」に着目した方式である．また，任意の格要素や「外の関係」のように，結合価文法では原理上解析が困難な場合がある．そこで，先行研究における他の型の名詞句解析や英文における PP-attachment の解析の方式を参考にして，名詞 A,B の意味属性および動詞と格要素の統計的共起情報が用いた方式を本稿の手法に取り入れる．なお,「いずれの方式でも係り先が定まらないならば，名詞 A にかかる」とするデフォルトルールを用いる．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，検査テストで 89%，評価テストで 84%の正解率であった．本手法は評価テストにおいて正解率が 57%というベースライン評価（デフォルトルールによる判定）を大きく上回ることから，提案手法の有効性が確認できた．

目 次

図 目 次

表 目 次

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，動詞と格要素の関係を明示している「結合価文法」で，動詞節の係り先が解析できる可能性がある．

ただし，結合価文法では動詞の語義を同定するために必須の格要素を記載しているが,任意となる格要素については記載がない．英語文の構文解析では，PP-attachment の問題が，任意の格要素の係り先解析の問題に近い．PP-attachment の問題を解決するために，手がかりとしている情報は,「v np1 p np2」型動詞句において，単語品詞と前置詞句の係り先の組についての頻度情報 [7] がある．したがって,「格要素と動詞の統計的共起情報」を考慮に入れる．

本稿で述べる解析手法は,「外の関係」の解析を明示的に実行する手段ではないが，上述した日本語の名詞句の解析手法を参考に,「名詞 A, B の意味属性の共起関係」を考慮に入れると，格関係を問わないため「内／外の関係」に依らず使用可能であることが期待される．

本稿では，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章では本稿で対象とする名詞句と着眼点について述べる．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着眼点

2.1　対象とする名詞句

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．以降ではこの型の名詞句を単に名詞句と呼び,「Ａの Ｂ」型名詞句を「の」型詞句と呼ぶ．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　「の」型名詞句の解析についての関連研究

「の」型名詞句の解析に関する先行研究を示す.

2.2.1　名詞間の接続強度を用いた「の」型名詞句構造解析

名詞間の接続強度を用いた「の」型名詞句構造解析 [1] では,「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手法を提案している．益田らは，名詞を 13 種類の品詞に分類し，それぞれの左右出現頻度を求めた．図１に係り受けの例を，表１に出現頻度を示す.

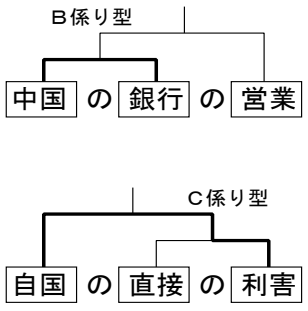

図 1: 係り受けの例

表１に示す頻度情報を元に，名詞Ｎの係り先を判定したところ，91%の解析成功率であるという.

表 1: 各品詞が「の」の左右に現れる頻度

品詞	左側出現頻度	右側出現頻度
普通名詞	289820	296367
サ変名詞	64695	115761
動作名詞	11600	23520
状態名詞	12513	7493
形容詞転生名詞	448	2726
形容動詞転生名詞	56	498
連体詞性名詞	31290	8508
数詞	9488	5409
時詞	37867	18590
副詞型名詞	15402	7723
固有名詞	20913	4860
形式名詞	12150	19441
代名詞	7749	3095

2.2.2 「の」型名詞句に対する形容詞の係り先解析

「の」型名詞句に対する形容詞の係り先解析 [3] では,「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手法を提案している．森内らは，以下に示す手順で，形容詞の係り先を決定している．

手順 1: 「AJ+N」型名詞句の名詞を意味属性に置き換え，その頻度統計をとる．

手順 2: 「AJ+A の B」型名詞句の名詞を意味属性に置き換える．

手順 3: 手順 2 で得られた結果を「形容詞＋名詞意味属性 α」,「形容詞＋名詞意味属性 β」に分ける．

手順 4: 手順 1 の頻度統計より，それぞれの頻度 (α， β の頻度) を調べ，頻度の大きい方を係り先と決定する．

以上の方法により，約 90%の正解率が得られるという．

2.2.3 Prepositional Phrase Attachment through a Backed-Off Model

Prepositional Phrase Attachment through a Backed-Off Model[7] では，前置詞句が文のどの品詞に係るのかを決定する手法を提案している．Collins らは，以下に示す手順で，前置詞句の係り先を決定している．

手順 1: 「v np1 p np2」型動詞句において，前置詞句の係り先についての頻度をとる．

手順 2: 同じく「v np1 p」,「v p n2」,「n1 p n2」についての頻度をとる．

手順 3: 「v p」,「n1 p」,「p n2」についての頻度をとる．

手順 4: 「p」だけの頻度もとる．

手順 5: 手順 1〜4 で得た頻度情報を元に，前置詞句の係り先を決定する．

以上の方法により，84.5%の正解率が得られるという．

2.3 内と外の関係

ある語や句や節が名詞を修飾する場合を「連体修飾」という．修飾される名詞を「底」というが，この修飾節と底との関係によって，2 通りに分けられる．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．

2.3.1 内の関係

「内の関係」である場合，動詞節と名詞との間に格関係が存在すれば，係り受け関係があると言える．動詞と格要素の関係は,「結合価文法」で明示されている．

結合価文法を用いるためには，日本語語彙大系 [8] の，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，以下で概説する．

一般名詞意味属性

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 2 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．

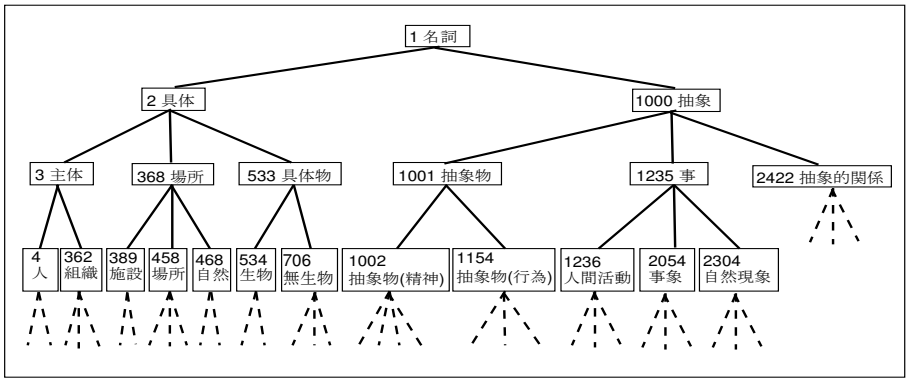

図 2: 一般名詞意味属性体系（上位部位）

結合価パターン

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 2 に具体例を示す.

表 2: 結合価パターンの例（括弧内は意味属性）

$N1$（人 物象） が $N2$（人） を 起こす
$N1$（主体） が $N2$（事件） を 起こす
$N1$（主体） が $N2$（番人 見積り） を 立てる
$N1$（主体） が $N2$（具体物） を 立てる

表 2 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,「$N1$（人） が $N2$（人） を 起こす」のパターンに適合する．これは，文中の「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，かつ，文中の「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.

内の関係の例

内の関係の例をあげる．

例 3：経営が破綻した組合

例 3 では，動詞節「経営が破綻した」が名詞「組合」に係る．両者の間には，格関係が存在し，格要素「の」で繋がるので,「組合の経営が破綻する」という形の文に変形することができる．

節を使った連体修飾文は，二つの文が結合してできあがっている．内の関係では，底となる名詞が元の二つの文のそれぞれに現れる．

組合の経営が破綻した

組合は解散した

→経営が破綻した組合は解散した（底の名詞は組合）

2.3.2 外の関係

「外の関係」である場合，動詞節と名詞の間に格関係が存在しないため，係り受け関係の判定に結合価文法を用いることができない．以下に外の関係の例をあげる．

例 4：ガラスが落ちる危険性

例 4 では，動詞節「ガラスが落ちる」は名詞「危険性」に係るが，両者の間に格関係は存在しないため，例 3 のように文の形を変形することができない．

また，外の関係の連体修飾文は，底となる名詞が一方の文だけにしか現れない．

ガラスが落ちる

その危険性がある

→ガラスが落ちる危険性がある（底の名詞は危険性）

2.4 解析方法に対する着眼点

名詞句が「内の関係」であれば，動詞と修飾される名詞の間には，格関係が存在する．そこで，動詞節の修飾先を決定する方法として，格関係の有無を利用する方法が考えられる．本稿では，動詞と格要素の関係を明示している「結合価文法」を，動詞節の係り先解析に用いる．

しかし，結合価文法では動詞の語義を同定するために必須の格要素を記載しているが，任意となる格要素については記載がない．また，名詞句が「外の関係」である場合，動

詞節と名詞の間に格関係が存在しないため，結合価文法を解析に用いることができない．この問題を解決するために，関連研究で用いられる手法を参考にする．関連研究では，名詞 A と B の組み合わせの頻度，名詞とそれを修飾する品詞の組み合わせの頻度，一定の型における係り先頻度など，主に単語間の頻度情報が用いられ，大きな成果をあげている．本稿の対象とする名詞句でも，名詞を意味属性に汎化し，係り先の頻度を取ることで，以下の例のような解析方式が提案できると推測される．

例 5：経営が破綻した二つの組合

例 6：大阪に住む板前の太田さん

例 5 の名詞 A「二つ」の意味属性は数量である．数量のような意味属性は，動詞節の係り先になることは少ないと思われるので，ある特定の意味属性の名詞は，動詞節の係り先にならないという方式が提案できる．

例 6 の名詞 A「板前」の意味属性は職人，名詞 B「太田さん」の意味属性は敬称である．職人といった意味属性は，もう一方の名詞の意味属性が主体配下である場合，それを修飾する性質があり，動詞節の係り先にはならない．このように，名詞 A と B の意味属性の組み合わせによって，動詞節の係り先を判定するという方式が提案できる．

そこで,「格要素と動詞の統計的共起情報」,「名詞 A,B の意味属性の共起関係」を考慮することで，結合価文法を用いた手法で解析できないものをカバーする．図 3 に標本と手法の対応関係を示す.

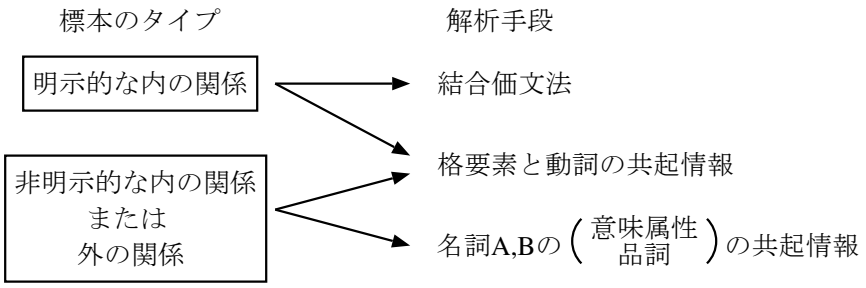

図 3: 標本と手法の対応

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．3.1 節で予備調査について述べ，格関係に関して 2 つの解析方式を 3.2 節および 3.3 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，3.4 節から 3.6 節に，提案する解析方式で結果が得られない場合に適用するデフォルトルールについて，3.7 節で述べる．

3.1　予備調査

まず，標本を集め,「V ＋ A の B」型名詞句の感触を探る必要がある．本稿では，95 年度版毎日新聞 [9] から標本を抽出する．その中で，分析・調査を行った検査用標本 1,000 件を，解析規則作成に使用する．標本には，正解ラベル（「A 係り」/「B 係り」），および，名詞の意味属性を付与する．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，本章で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（全標本データを付録 1 に，入力形式のデータを付録 2 に示す）．なお，抽出したデータを調査したところ，全て内の関係の名詞句であり，外の関係となる名詞句が無かった．以後，本稿で扱う標本データは，全て内の関係の名詞句である．

3.2　結合価文法による係り受け解析（VPM 方式）

動詞節と内の関係にある名詞は，動詞節の係り先となる．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．

3.2.1　VPM 方式の解析手順

「V+A の B」型名詞句が与えられた場合，以下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」とし，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」とする.

以下に具体例を示す.

例 7 : 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.

3.2.2 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 [10].

[10] 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．

本稿で扱う適合度の計算式は，次のようになる．

$$\text{《適合度》} \;=\; \sum_{i} D_{A_i} S_i$$

i はパターンの格要素の数，D_{A_i} は属性 A_i の日本語語彙大系で定義される深さを示し，S_i は格要素の種類ごとの得点を示す.

得点 S 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，検査用標本 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 3 にその結果を示す.

表 3: 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 D 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，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仮説に基づき，設定した．

以下に適合度の計算例を示す．

例 8：首相に謝罪を求める農民の手紙

例 8 に対し，次に示す 2 つのパターン P1 および P2 が適合するとする．

P1: $N1$(主体) が $N2$(抽象) を $N3$(主体) に 求める

P2: $N1$(主体) が $N2$(助力, 援助) を $N3$(主体) に/へ 求める

例 8 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 P1 の $N2$ に内包されるが，P2 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．

《P1 の適合度》

$$\begin{aligned} &= (D_{\text{主体}} * 5.0) + (D_{\text{抽象}} * 4.0) + (D_{\text{主体}} * 3.5) \\ &= (3 * 5.0) + (2 * 4.0) + (3 * 3.5) \\ &= 33.5 \end{aligned}$$

《P2 の適合度》

$$\begin{aligned} &= (D_{\text{主体}} * 5.0) - (((D_{\text{助力}} + D_{\text{援助}})/2) * 1.2) + (D_{\text{主体}} * 3.5) \\ &= (3 * 5.0) - (((9 + 9)/2) * 1.2) + (3 * 3.5) \\ &= 14.7 \end{aligned}$$

P1 の方が高いので，例 8 では P1 を選択する．

3.3 動詞と名詞の共起に注目した係り受け解析（VCC 方式）

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，標本に適合しない場合がある．つまり，登録されているパターンに，標本の語義が完全に一致しない．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 element Co-occurrence 方式）を導入する．

3.3.1 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．ただし，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ようとした場合，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，現在のところ網羅性が十分であるコーパスが得られるのは困難である．

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ととし，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．この組の集合を共起データベースと呼ぶ．なお，データベースに登録する一般名詞意味属性は，それぞれの名詞が持つ，全ての属性を対象とする．

表 4 に共起データベースの例を示す．共起データベースに登録された動詞は，およそ 9,700 件である．

表 4: 共起データベース（括弧内は共起頻度）

動詞	共起する名詞の意味属性
読む	本 (49), 息子 (2), 童話 (1), ...
書く	手紙 (10), 友人 (2), ...
見る	犯人 (6), 住民 (2), ...
...	...

3.3.2 VCC 方式の解析手順

「V+A の B」が与えられるたとき，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．

3.3.3 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，検査用標本を満足していない．そこで，検査用標本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．

3.3.4 解析の具体例

例 9：被災した朝鮮の人達

手順 1: A=「朝鮮 (領土)」，B=「接辞 (人間／複数)」となる．

手順 2: 動詞「被災する」に対応する共起データから，A と B の意味属性の共起頻度を検索，A の共起頻度は 1，B の共起頻度は 22 となる．

手順 3: 手順 2 より，$1 < 22$ から，共起頻度の高い B を係り先として選択し,「B 係り」と判定する．

3.4　特殊記号を手がかりとした係り受け解析（IPS 方式）

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）を提案する．検査用標本 1,000 件の分析による特殊記号の出現回数と正解数を表 5 に示す.（IPS 方式が適用可能な標本を付録 3 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 5: 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「」で囲まれているため，動詞節「K 社が販売する」は「A 係り」と判定する．

例 10：K 社が販売する「ネオン」の開発

3.5　特定の意味属性を根拠とした係り受け解析（EPA 方式）

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）を提案する．検査用標本 1,000 件の分析によると，表 6 に示す 9 つの属性があげられる．(EPA 方式が適用可能な標本を付録 4 に示す)

なお，名詞 A,B ともに表 6 に示す属性に内包されない場合は,「解析不能」と出力する．

表 6: 係り先とならない意味属性

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

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．

例 11：二日にハワイで行う 50 周年の記念式典

3.6　名詞 AB 間の意味関係による係り受け解析（AAC 方式）

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 Co-occurrence 方式）を提案する．属性の組み合わせの種類と係り先のルールを検査用標本 1,000 件の分析に基づき，44 種類作成した．表 7 に一部の例を示す.（AAC 方式が適用可能な標本を付録 5 に示す）

なお，名詞 A と B の意味属性の組み合わせが，表 7 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 7: 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	組織	人 (職業・地位・役割)	B	34	22
2	場所	施設	B	17	17
3	人 (職業・地位・役割)	敬称	B	14	14
4	人間 (親族関係)	人間 (親族関係)	B	2	2
	...	...			
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）に内包されるので，動詞節「バッテリーを組む」は「B 係り」と判定する．

例 12：バッテリーを組む主将の鶴岡君

3.7　デフォルトルールによる係り受け解析（DR）

検査用標本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．本章で解析方式を提案したが，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．

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 である．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ことから，本研究の重要性が指摘できる．

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，検査テストにおける各方式の個別成功率を求め，方式を用いる優先順位を定める．そして，評価テストにおいては，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．

評価〇： 解析結果が一つに定まり，正解ラベルと一致する場合

評価×： 解析結果が正解ラベルと一致しない場合

評価*1： 解析結果が「A 係りおよびＢ係り」である場合

評価*2： 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，評価〇または評価×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，評価〇または評価×における評価〇の割合である．

検査用標本 1,000 件に対する各解析方式の解析結果を表 8 に示す.

表 8: 個別解析結果

方式	評価○	評価×	評価*1	評価*2	カバー率	成功率
VPM	54.2%	6.5%	13.3%	26.0%	60.7%	89.3%
VCC–	63.7%	27.9%	0.9%	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%
AAC	30.2%	5.1%	-	64.7%	35.3%	85.5%
(VCC	75.0%	23.9%	1.1%	0.0%	98.9%	75.8%)

ここで,「VCC–」方式とは，本章のみで使う方式で，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，検査用標本 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，評価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（3.6節参照）に従い,「A係り」として出力する（図4）．

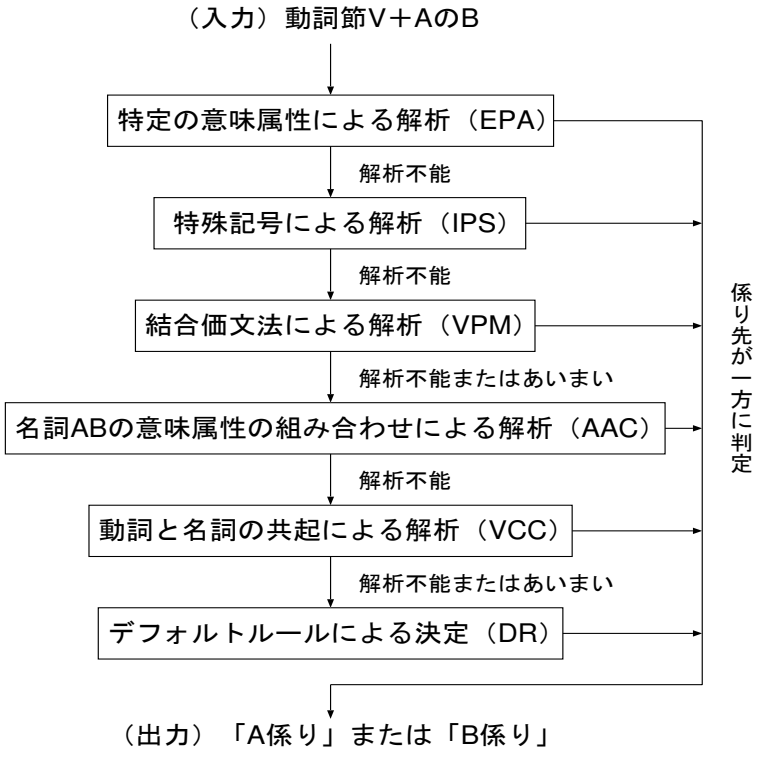

図 4: 係り受け解析の順序

5　実験

5.1　評価データ

本稿では，3.1 節で解説した検査用標本データと同じく，実験用の評価データとして，95 年度版毎日新聞から抽出した「V+A の B」の表現 400 件を使用する．ただし，表現の幅が広い動詞「ある」,「する」[1],「なる」を含む名詞句は，動詞ごとの個別の規則が必要になってくるため，本稿の対象外とし，修飾する動詞節が受け身になっている名詞句は，能動態に変形すると，格助詞の種類が変化することがあるため，対象外とする．

5.2　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 ALL，5 つの方式のみを統合した手法を ALL-DR，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評価データ 400 件である．比較のため，統合した解析手法による，検査用標本 1,000 件の精度も求める．

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．

[1]「〜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．

5.3 解析例

5.3.1 5 方式により係り先が断定できた場合

〇の例 例 13：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 13 では次のパターンが選択される．
[$N1$(乗り物, 人) が $N2$(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 $N2$(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 $N1$(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，$N1$(乗り物) に内包される．

$N1$ 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，評価は〇となる．

×の例 例 14：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．

5.3.2　デフォルトルールで判定を補助した場合

ALL-DR での*1 の例　例 15：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 15 では次のパターンが選択される．
[$N1$(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*1 となる．

ALL-DR での*2 の例　例 16：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*2 となる．

5.4 実験結果

実験結果を表 9 にまとめる．対象の open は評価データを，closed は検査用標本を示す．手法は 5.1 節で解説した 3 つの手法であり，正解率は評価○となったものの割合である．

評価テストにおいて，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，評価データの 400 件中，10 件であった．ALL と ALL-DR 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，検査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．

本稿の提案手法 (ALL) 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．詳しい実験方法は次節で述べるが，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[11] を用いて，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．また，VCC 方式の標本数をより拡大した方式として，インターネット検索サイトを利用する方式が考えられる．ランダムで選んだ標本 100 件を，動詞と名詞 A,B を各種格要素で繋ぎ，検索サイト Google で検索ヒット数による共起判定を行ったところ，カバー率は 98%，成功率は 60%であった．

以上より，本手法の有効性が確認できた．

表 9: 各手法の正解率

対象	手法	正解率
open	ALL	84%(337/400)
open	ALL-DR	83%(332/400)
open	DR	57%(228/400)
closed	ALL	88.6%(886/1000)
closed	ALL-DR	87.7%(877/1000)
closed	DR	50.5%(505/1000)

5.5 比較実験

5.5.1 cabocha による解析実験について

本稿で提案する手法との比較のため，係り受け解析ツール cabocha での解析を行う．評価対象は検査用標本 1,000 件である．以下に例を示す.

例 17：予算に盛り込んだ都道府県会議員の海外視察費

例 17 を cabocha にかけると，出力は次のようになる.

```
0 1D 0/1 1.28626164
予算 ヨサン 予算 名詞-一般
に ニ に 助詞-格助詞-一般
1 3D 0/1 0.73579348
盛り込ん モリコン 盛り込む 動詞-自立 五段・マ行 連用タ接続
だ ダ だ 助動詞 特殊・タ 基本形
2 3D 2/3 0.00000000
都道府県 トドウフケン 都道府県 名詞-一般
会議 カイギ 会議 名詞-サ変接続
員 イン 員 名詞-接尾-一般
の ノ の 助詞-連体化
3 -1O 2/2 0.00000000
海外 カイガイ 海外 名詞-一般
視察 シサツ 視察 名詞-サ変接続
費 ヒ 費 名詞-接尾-一般
```

ここで,「0 1D 0/1 1.28626164」のパラメータの内，左隅の 0 が文節番号，次の 1D の数字部分が係り先の文節番号である．他の値については，本稿の内容とは関係がないので省略する.

動詞の係り先文節パラメータは 3D となっているので，係り先は文節 3 の「海外視察費」であり，B 係りとなる.

cabocha の解析結果から，動詞が名詞 A と B のどちらに係っているのかを人手で集計したところ，正解数は 596 件（正解率 59.6%）であった．本稿の手法による検査テストの正解率は 88.6%であり，cabocha による解析結果を約 30%上回った.（全解析結果を付録 6 に示す）

5.5.2 検索サイトを利用した解析実験について

本稿で提案したVCC方式との比較のため，検索サイトGoogleを利用した共起判定を行う．評価対象は検査用標本からランダムに選んだ100件である．手順は次のようになる．

手順 1: 標本の動詞と名詞A,Bそれぞれを，表10に示す，各種格要素で繋ぐ．

手順 2: 1で接続したものを，全てGoogle検索にかけ，検索ヒット数を調べる．

手順 3: ヒット数の合計が多い方を，係り先として判断．

表 10: 格要素の種類

が格	を格	に格	で格	へ格	の格	と格	から格	より格	まで格

以下に例を示す．

例18：隣接する共和国の大統領

手順 1: 「共和国が隣接する」,「共和国を隣接する」，...,「共和国まで隣接する」,「大統領が隣接する」,「大統領を隣接する」，...,「大統領まで隣接する」のように，動詞と名詞A,Bを各種格要素で接続．

手順 2: 1で接続したものをGoogle検索する．ヒット数を集計する（表11）．

手順 3: 表11より，検索ヒット数は名詞Aの方が多いので，A係りと判断する．

表 11: Google検索ヒット数（例18の場合）

	が格	を格	に格	で格	へ格	の格	と格	から格	より格	まで格	合計
名詞A	2	0	155	0	0	9	40	1	0	0	207
名詞B	0	0	0	0	0	0	0	0	0	0	0

評価対象100件中，カバー率は98%，成功率は60%となり，本稿で提案した手法が大きく上回っていることを確認した．また，VCC方式単体と比較しても，Google方式よりも約10%上回っている．(全解析結果を付録7に示す)

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，格関係の解析が重要であり，結合価文法による解析方式 (VPM) のカバー率と成功率の向上が必要であるので，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.

検査テスト（表 8）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 16 に対しては，動詞「うたた寝する」の結合価パターンとして，次のパターンを追加した.

[$N1$(人) が $N2$(場所, 場) でうたた寝する]

- 例 16 の「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+) のみによる検査テスト，および，統合手法 (ALL+) による検査テスト，および，評価テストを行ったところ表 12，表 13 の結果を得た．VPM+では，VPM に比べて評価○の割合が 20%以上上がり，カバー率も 20%，正解率も 7%向上した．ALL+では，評価テストの結果は変わらなかったが，検査テストでは正解率が 5.4%向上した．評価テストで結果が変わらなかったのは，数万におよぶ結合価パターンに対し，200 件程度の結合価パターンを追加しても，カバー率に影響を及ぼさないためである．

表 12: 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*1	評価*2	カバー率	成功率
VPM+	78.7%	2.9%	15.8%	2.6%	81.6%	96.4%

表 13: 追加パターンを使用した正解率

対象	手法	正解率
open	ALL+	84.2%
closed	ALL+	94.0%

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．

以下にパターン追加できない標本の例をあげる．

例 19：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*1 となるため，パターンを追加できない．

また，意味属性に違いがないため，VPM 方式だけでなく，VCC 方式，EPA 方式での解析も不能である．例 19 のような名詞句は，特殊記号が使われている場合では IPS 方式，それ以外では AAC 方式での対応が可能である．名詞 A と名詞 B の意味属性が同じ標本は，全部で 9 件存在した．

7　今後の課題

7.1　外の関係

本稿の方式は，外の関係の「V+A の B」型名詞句が，新聞データから抽出できなかったため，外の関係に特化した解析方式がない．外の関係の名詞句に対し，本稿で述べた手法が有効かどうか，いくつかの作例で検証してみた．

例 20：ガラスが落ちる危険の回避

(正解：A 係り)

例 21：議員がうたた寝する写真の値段

(正解：A 係り)

例 20 は VCC 方式で，例 21 は EPA 方式でともに A 係りという結果が出力され，どちらも評価○となり，本稿の手法が外の関係にも適用可能だと分かった．しかし，標本が新聞記事から収集できなかったため，外の関係の名詞句を集める必要がある．また，検証に使用した外の関係の名詞句は，係り先の正解が「A 係り」となるものばかりであった．標本数を増やし，正解が「B 係り」となるものがあるかどうか調べる必要がある．新聞記事のような実用文では，外の関係の名詞句は希少だが，言語学的には当然存在するものである．今後は，小説などのデータから標本を集める必要がある．

7.2　名詞意味属性の絞り込み

本稿の解析方式のための入力は，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，解析時に意味属性が一意に決定されていない．

評価テストにおいて，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，ALL の正解率は 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．

7.3　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

例 22：真相が明らかになった盛岡地裁の例
(正解：A 係り)
例 23：トラックの荷台に積まれたタンクの中
(正解：A 係り)

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．

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．

解析方式を作る検査用標本として，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つの解析法式とデフォルトルールを組み合わせて，標本データに適用したところ，検査テストでは約 89%，評価テストでは約 84%の正解率となり，ともにデフォルトルールよりも正解率が約 30%上回ることから，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] M.Collins, J.Brooks : Prepositional Phrase Attachment through a Backed-Off Model, In Proceedings of the Third Workshop on Very Large Corpora (1995).

[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] 毎日新聞社：CD-毎日新聞'95 データ集, 日外アソシエーツ (1996).

[10] 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).

[11] 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).

付録

付録 1: 検査用標本，評価データ

付録 2: 検査用標本，評価データ（入力形式）

付録 3: IPS 方式適用可能な標本

付録 4: EPA 方式適用可能な標本

付録 5: AAC 方式適用可能な標本

付録 6: cabocha での検査用標本解析結果

付録 7: Google 検索による共起判定