概要

日本語複文 (埋め込み文) は，英語に翻訳する場合，関係節，同格節，準動詞，前置詞句など様々な表現に訳されることが多い．埋め込み文を正しく訳し分けることは，日英機械翻訳の重要な課題の一つとなっている．その課題の 1 つが「内と外の関係」である．

そこで本稿では，動詞及び節が底の名詞 (英語では先行詞に相当する) を修飾する，連体節を含む文を取り上げ，「底の名詞の意味属性」及び「修飾部の述部の結合価パターン」を考慮することによって，「内と外の関係」を計算機で判断する方法を提案し，その精度を評価した．

その判断規則の有効性を，アンカー和英辞典 (約 46000 文) より抽出した複文 200 文を用いて検証した結果，クローズドテストにおいて内の関係で 82.2%，外の関係で 83.4%の正解率を得た．さらに，アンカー和英辞典より複文 400 文を抽出してオープンテストを行った結果，内の関係で 84.9%，外の関係で 71.1%の正解率を得た．

この結果より，底の名詞の意味属性および修飾部の述部の結合価パターンを考慮した内と外の関係の判断規則の有効性が示された．

目 次

1 まえがき **1**

2 内と外の関係 **2**

2.1 日本語複文の構造 2

2.1.1 複文 (研究対象) 2

2.1.2 重文 (研究対象外) 2

2.2 内と外の関係 4

3 日本語語彙大系 **5**

3.1 意味属性体系 5

3.2 構文体系 6

4 内と外の関係の判断方法 **7**

4.1 底の名詞の意味属性 7

4.2 述部の結合価パターン 8

4.3 固有名詞と一般名詞意味属性 9

4.4 補助動詞 10

4.5 格の変更と制限 11

5 本手法の判定方法 **12**

6 実験と結果 **14**

6.1 クローズドテスト 14

6.2 オープンテスト 15

7 正解と失敗の例 **16**

7.1 正解例 (評価が◎ の場合) 16

7.2 失敗例 (評価が× の場合) 18

7.3 評価が○ の例 20

8 考察 **21**

8.1 失敗となる原因 (オープンテスト) 21

9　あとがき　25

図 目 次

1 日本語複文の構造 2
2 非埋め込み文 (研究対象外) 2
3 「内の関係」と「外の関係」 4
4 一般名詞意味属性体系 (上位属性) 5
5 結合価パターンの例 6
6 「内と外の関係」の決定規則の流れ 13

表 目 次

1 「女性」の意味属性 7
2 「覚悟」の意味属性 7
3 一般名詞と固有名詞の意味属性対応表 (一部) 9
4 「伊吹山」の固有名詞意味属性 9
5 「伊吹山」の一般名詞意味属性 9
6 補助動詞と「内と外の関係」 10
7 「夢」の一般名詞意味属性 10
8 評価結果 14
9 評価結果 (クローズドテスト) 14
10 評価結果 15
11 評価結果 (オープンテスト) 15
12 複数存在する底の名詞の意味属性の割合 (評価○) 15
13 形態素解析の結果 (例文 18) 16
14 意味属性 (例文 19) 16
15 結合価パターン (例文 19) 16
16 形態素解析の結果 (例文 20) 18
17 意味属性 (例文 20) 18
18 結合価パターン (例文 20) 18
19 形態素解析の結果 (例文 21) 19
20 意味属性 (例文 21) 19
21 結合価パターン (例文 21) 19
22 形態素解析の結果 (例文 22) 20
23 意味属性 (例文 22) 20
24 結合価パターン (例文 22) 20
25 形態素解析の結果 (例文 23) 21
26 意味属性 (例文 23) 21
27 結合価パターン (例文 23) 22
28 形態素解析の結果 (例文 24) 23
29 意味属性 (例文 24) 23
30 結合価パターン (例文 24) 23

31　形態素解析の結果 (例文 25)　. 24

1　まえがき

日本語複文 (埋め込み文) は，英語に翻訳する場合，関係節，同格節，準動詞，前置詞句など様々な表現に訳されることが多い．そのため，埋め込み文を正しく訳し分けることが困難となっており，日英機械翻訳の重要な課題の一つとなっていた．その課題の 1 つに「内と外の関係」がある．

従来の複文の分類方法として，底の名詞 (英語における先行詞) と修飾部の格関係の種類に応じて複文を分類する方法 [1] や，底の名詞が連体節内で格関係を有するか否かに着目して分類を行う方法 [2] がある．しかし，いずれの方法も，計算機で判断するのは困難となっていた，

昨年の研究で，表 [3] は，底の名詞の意味属性に着目して，底の名詞の意味属性が「具体」配下の場合を「内の関係」，底の名詞の意味属性が「抽象」配下の場合を「外の関係」と考え，「内と外の関係」を判断した．しかし，「底の名詞の意味属性」だけでは，正解率がそれぞれ内の関係で 94.9%，外の関係で 63.8%であり，外の関係の正解率が，内の関係の正解率に比べ低い結果となった．

そこで本稿では，日本語複文の底の名詞の意味属性，修飾部の述部の結合価パターンに着目することによって，「内と外の関係」の計算機による判断規則を作成し，精度の向上を目指す．その結果，クローズドテストにおいては，「内の関係」で 82.2%，「外の関係」で 83.4&の正解率が得られた．また，オープンテストにおいては，「内の関係」で 84.9%，「外の関係」で 71.1%の正解率が得られ，判断規則の有効性が示された．

以下，第 2 章では，内と外の関係，第 3 章では，日本語語彙大系，第 4 章では，内と外の関係の判断方法，第 5 章では，本手法の判断規則，第 6 章では，実験結果，第 7 章では，正解と失敗の例，第 8 章では，考察，第 9 章では，今回の結論を述べる．

2　内と外の関係

本研究では，日本語複文を「内と外の関係」に分類するため，今回対象とする複文の構造について述べる．

2.1　日本語複文の構造

2.1.1　複文 (研究対象)

図 1 に示すように，底の名詞 (英語の先行詞に相当する) に係る修飾部 (動詞及び節) が文中に埋め込まれた，連体節を含む日本語複文を本研究の対象とする．

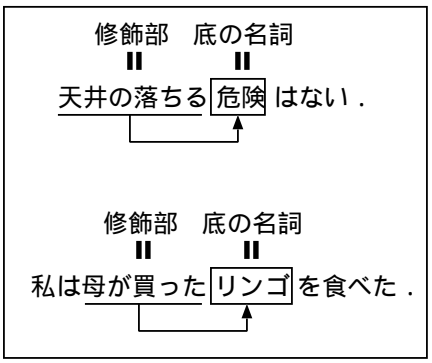

図 1: 日本語複文の構造

2.1.2　重文 (研究対象外)

また，埋め込み文でない文としては，図 2 に示すように，主節と従属節が等位関係，従位関係にある文を接続助詞で結んだ重文がある．

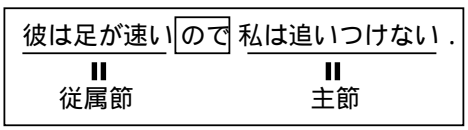

図 2: 非埋め込み文 (研究対象外)

なお，底の名詞が以下に示す形式名詞については，今回対象外とする．

1. 「とき」,「ため」

   前の節に対しては連体修飾される名詞として働くが，後の節に対しては副詞的な修飾として働くという二面性をもっている．名詞としての「~とき」「~ため」と，接続助詞としての「~とき」「~ため」と考えられるため今回は取り扱わない．

   (例文 1)　事故が起きた<u>とき</u>、飛行機には 150 人の旅客が乗っていた。

   (例文 2)　彼の店は不便な所にある<u>ため</u>、はやっていなかった。

2. 「こと」,「はず」,「もの」,「の」

   上と同様に考えて，助動詞としての「~ことがある」「~はずだ」と考えられるため，今回は取り扱わない．

   (例文 3)　この計画は変更される<u>こと</u>がある。

2.2　内と外の関係

本研究では，「内と関係」と「外の関係」を判断するために，寺村 [2] の「内と外の関係」法を用いる．(図 3 参照)

寺村 [2] の「内と外の関係」は，英語の「関係節」と「同格節」への対応に用いられる．寺村は，図 3 に示すように，底の名詞が連体節内で格関係 (ガ格・ヲ格・ニ格・デ格等) を有するか否かに着目して，複文を「内の関係」と「外の関係」に分類した．ここで，底の名詞と修飾部の間に格関係を有するものを「内の関係」と呼び，これは，英語の「関係節」に対応する．そして，格関係を含まず，修飾部が底の名詞の具体的陳述であるものを「外の関係」と呼び，これは，英語の「同格節」に対応する．

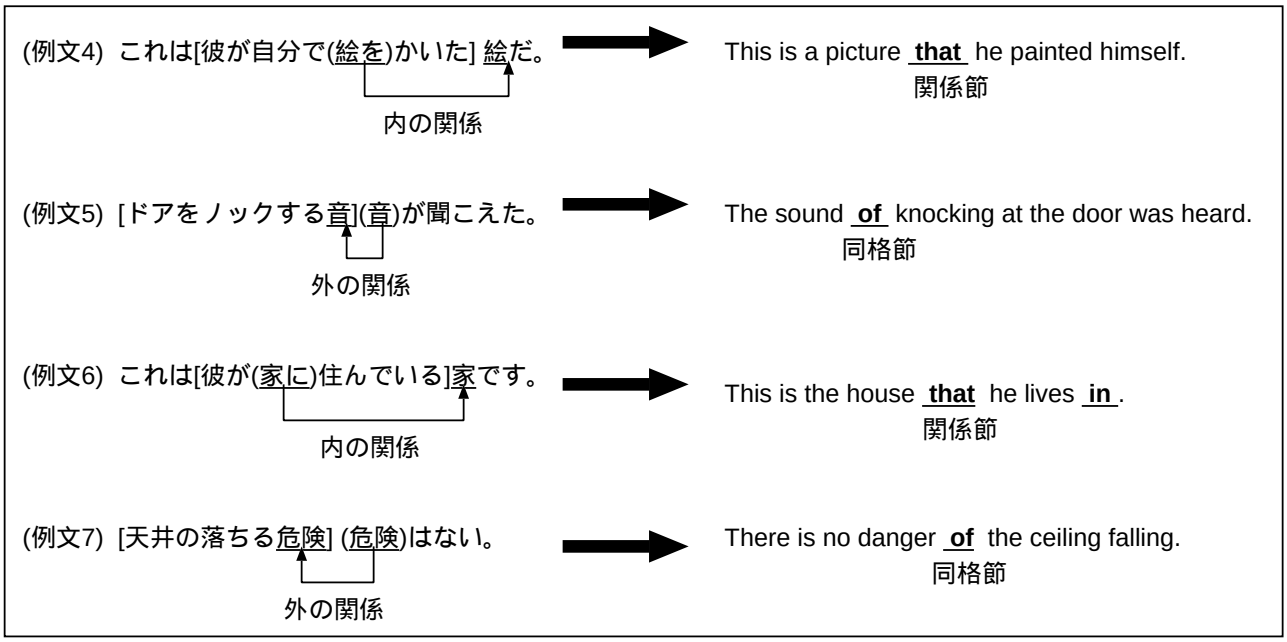

図 3: 「内の関係」と「外の関係」

3　日本語語彙大系

「内と外の関係」については，従来人手で判断していたため，計算機では判断できないという問題点がある．そこで，以下に本手法で「内と外の関係」を計算機で判断する方法について述べる．

「内と外の関係」を計算機で判断するにあたって，本手法では，日本語語彙大系 [4] を用いて格関係の有無の判断を行う．

池原らによって作成された日本語語彙大系 [4] は，「意味体系」「単語体系」「構文体系」より構成されている．本研究では，その中の「意味体系」と「構文体系」を用いる．

3.1　意味属性体系

「意味体系」とは，単語の見方，捉え方に着目して，名詞の意味的用法を整理，体系化したものであり，その中で約 40 万語の一般名詞は，2710 の意味属性に分類されている．そして，それぞれのノードは，上位の属性が下位の属性を内含する性質を持ち，最大 12 段の木構造にまとめられている．(図 4 参照)

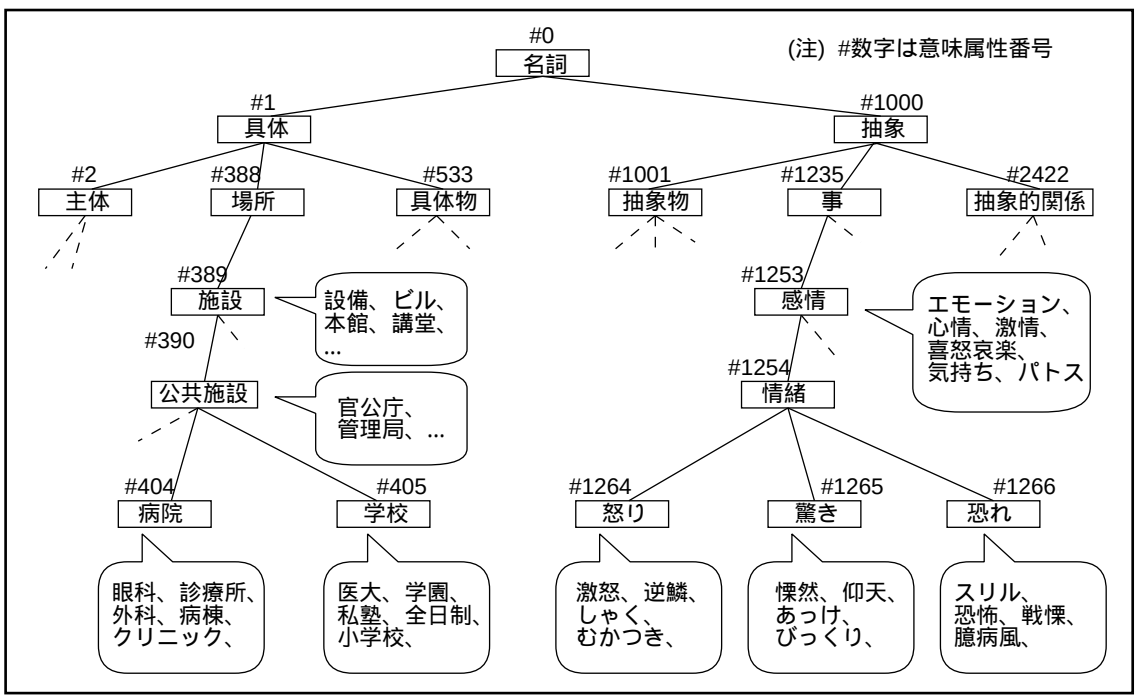

図 4: 一般名詞意味属性体系 (上位属性)

3.2 構文体系

また，「構文体系」の結合価パターン (図 5 参照) は，用言と格要素の意味的関係を記述したものである．これにより用言と底の名詞との間に意味的な制約が生まれ，この制約を利用して，「内と外の関係」の判断にも応用できると考えられる．

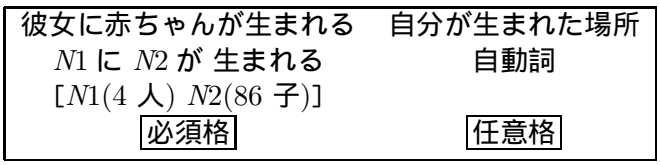

図 5: 結合価パターンの例

「構文体系」において，図 5 のように登録されている結合価パターンの格関係は「必須格」とされ，これは，それぞれの用言にとって必要とされる格関係である．しかし，図 5 に示すように，格関係のなかには「時」「場」などのように用言によってはなくてもよいものもある．このような格関係は「任意格」とされ「構文体系」の中には登録がされていないが，日本語複文において底の名詞となることもあるため，格関係とみなすことができる．

以上のことより，日本語語彙大系 [4] を用いることで格関係の有無を決定できると言える．

4　内と外の関係の判断方法

4.1　底の名詞の意味属性

ここでは，前章をふまえて「内の関係」と「外の関係」に分類する方法について考える．「内の関係」と「外の関係」の大きな違いは「修飾部と底の名詞の間に格関係を有するか否か」である．昨年の表[3]の研究では，アンカー和英辞典 (約 46000 文) より埋め込み文を 244 文抽出して，底の名詞により「内と外の関係」の分類を行った．その結果，底の名詞の意味属性が「名詞/具体/... 」である場合，約 99% が「内の関係」となる．また，底の名詞の意味属性が「名詞/抽象/... 」となる場合，約 65%が「外の関係」となることが報告されている．

(例文 8) 彼女は国会議員に選ばれた最初の 女性 である．(内の関係)

表 1: 「女性」の意味属性

[/1 名詞/2 具体/3 主体/4 人/5 人間/46 人間〈生物学的特徴〉/47 人間〈男女〉/49 女]
[/1 名詞/2 具体/3 主体/4 人/5 人間/111 人間〈対人関係〉/112 人間〈交際関係〉/ 124 友・なじみ/127 恋人/129 恋人 (女)]

(例文 9) 彼らはカナダに骨を埋める 覚悟 で移住を決意した．(外の関係)

表 2: 「覚悟」の意味属性

[/1 名詞/1000 抽象/1235 事/1236 人間活動/1237 精神/1351 意向/1370 執着・諦め/ 1374 断念]
[/1 名詞/1000 抽象/1235 事/1236 人間活動/1237 精神/1351 意向/1353 決心・ためらい/ 1354 決心]

また，修飾部に「という」を含む場合，統計的に「外の関係」となる．ここで ,「... 」はその属性の配下を示す．

(例文 10) 彼らはサハラ砂漠を車で横断するという 偉業 を達成した．(外の関係)

(例文 11) その外国人は困ったという 身振り をした。(外の関係)

4.2 述部の結合価パターン

4.1 節より，底の名詞の意味属性が「具体」配下の名詞は，「内の関係」とみなすことができる．しかし，底の名詞の意味属性が「抽象」配下の場合は，「内と外の関係」が決定できない．そこで，底の名詞が「抽象」配下の場合の判定方法について考える．

(例文 12) しょせん地球上に住む 人々 は運命共同体なのだ．(内の関係)

(例文 13) あの山には鬼が住んでいるという 言い伝え がある．(外の関係)

例文 12・例文 13 のように，底の名詞の意味属性が「抽象」配下の場合，底の名詞が同じであっても動詞によっては，「内の関係」にも「外の関係」にもなりうることがある．これは，動詞によって格関係をとることができる名詞の意味属性が異なるからである．具体的には，「住む」という動詞は，「N1(4 人 535 動物) が N2(388 場所 2610 場 863 建造物 384 社会) に 住む」という結合価パターンを持つ．例文 12 の場合，底の名詞「人々」が「/1 名詞/2 具体/3 主体/4 人/5 人間」の意味属性を持つために，「住む」の結合価パターンに当てはまり，底の名詞と格関係を持つので「内の関係」と言える．

また，例文 13 の場合では，底の名詞「言い伝え」の意味属性が「/1 名詞/1000 抽象/1235 事/1236 人間活動/1237 精神/1495 言動/1540 報道/1549 伝達・報知/1550 伝達」であり，「住む」の結合価パターンに当てはまらないので格関係がないとし，「外の関係」と言える．

4.3 固有名詞と一般名詞意味属性

4.1 節より名詞の意味属性が一般名詞の場合，意味体系より意味属性を知ることができるが，固有名詞の場合は一般名詞意味属性を適用できない．そこで，表 3 を用いて，固有名詞意味属性を一般名詞意味属性に置き換えることにする．

例えば，固有名詞「鳥取」の場合，固有名詞意味属性は「/1 固有名詞/2 地名/... 」となるが，表 3 を用いることにより，一般名詞意味属性の「/1 名詞/2 場所/... 」と置き換えることができる．

対応表の詳細を付録 1 の (1) に示す．

表 3: 一般名詞と固有名詞の意味属性対応表 (一部)

固有名詞意味属性	→	一般名詞意味属性
姓	→	人間
組織名	→	組織
国名	→	領土　国家
地域	→	場所

(出典:日本語語彙大系)

(例文 14) 雪を頂いた 伊吹山 はみごとだ．

表 4: 「伊吹山」の固有名詞意味属性

[/1 固有名詞/2 地名/46 自然名/47 地形名/48 地形名 (陸)/49 陸上地形名]

例文 14 の場合，「伊吹山」の意味属性が固有名詞のため一般名詞意味属性が適用できない．そこで，表 3 を用いて固有名詞意味属性を一般名詞意味属性に置き換えると意味属性は，表 5 のようになる．

表 5: 「伊吹山」の一般名詞意味属性

[/1 名詞/2 具体/388 場所/468 自然/469 地勢/470 地形/471 陸地]

4.4 補助動詞

結合価パターンを持つ動詞については,「内と外の関係」を決定することができる．しかし,「となる」・「にする」・「である」・「になる」のような補助動詞は結合価パターンが登録されていない．そこで，補助動詞の場合の決定方法を考える．底の名詞が「抽象」配下で補助動詞を含む文 (17 文) を対象に「内と外の関係」の分類を行った．その結果を表 6 に示す．

表 6: 補助動詞と「内と外の関係」

	抽象物 (文)	事 (文)	抽象的関係 (文)
内の関係	1	1	6
外の関係	3	3	3

表 6 から「抽象的関係」の場合，約 7 割が「内の関係」となる .「抽象物」・「事」の場合，約 8 割が「外の関係」となる．そこで本手法では，動詞が補助動詞の場合,「抽象的関係」配下のものを「内の関係」,「抽象物」・「事」配下のものを「外の関係」とする．

(例文 15) プロゴルファーになる夢が実現した．(外の関係)

表 7: 「夢」の一般名詞意味属性

[/1 名詞/1000 抽象/1235 事/1236 人間活動/1237 精神/1239 感覚/1249 睡眠・覚醒/1252 夢]
[/1 名詞/1000 抽象/1235 事/1236 人間活動/1237 精神/1351 意向/1361 願望・失望/1363 希望]
[/1 名詞/1000 抽象/1001 抽象物/1002 抽象物（精神）/1003 知的生産物（思考・学習）/ 1019 論理・意味等/1023 真偽/1025 偽]

例文 15 の場合，底の名詞「夢」の意味属性は「事」配下と「抽象物」配下となるため，「外の関係」となる．

4.5 格の変更と制限

日本語複文では「火のついた油に水をかけても無駄だ」のように修飾部内の述部の前にくる格要素「の」が「が」の代わりにくることがある．これは「の」が主格なのか所有格なのかを計算機では判断できないため「結合価パターン」を用いる際に妨げとなる．そのため，次のように変更する．

(例文 16) 火のついた油に水をかけても無駄だ．
⇒ 火がついた油に水をかけても無駄だ．

(例文 17) わたしの訪れたその小さな町はうらぶれた感じだった．
⇒ わたしが訪れたその小さな町はうらぶれた感じだった．

5　本手法の判定方法

4 章での考えをもとに内と外の関係の判断規則を作成する．以下に判断規則の手順を説明する．

1. 入力文の解析

   (a) 入力文は，「修飾部 + 底の名詞」の形を含む文とする．
   (例) 彼の仕事は歴史に残る偉業だ。

   (b) 入力文を形態素解析にかけ，「述部」「底の名詞」に該当するものを調べる．
   (ここで，修飾部が「という」となる場合は，「外の関係」とする。)

   (c) 「底の名詞」については「一般名詞意味属性」を日本語語彙大系 (意味体系) より調べる．

   (d) 「述部」については「結合価パターン」を日本語語彙大系 (構文体系) より調べる．

2. 「内と外の関係」の判断

   (a) 「底の名詞の意味属性」と「結合価パターン」より内と外の関係を調べる。

   (b) 底の名詞の意味属性が「具体」配下の場合を「内の関係」とする。

   (c) 底の名詞の意味属性が「抽象」配下となるものについては述部の結合価パターンを調べ，格要素となるものを「内の関係」とする。

   (d) 格要素にならない場合は任意格により格関係を決定する。

   (e) 格関係がある場合を「内の関係」，格関係のない場合を「外の関係」とする。

図 6 に本研究で提案した判断規則を示す．

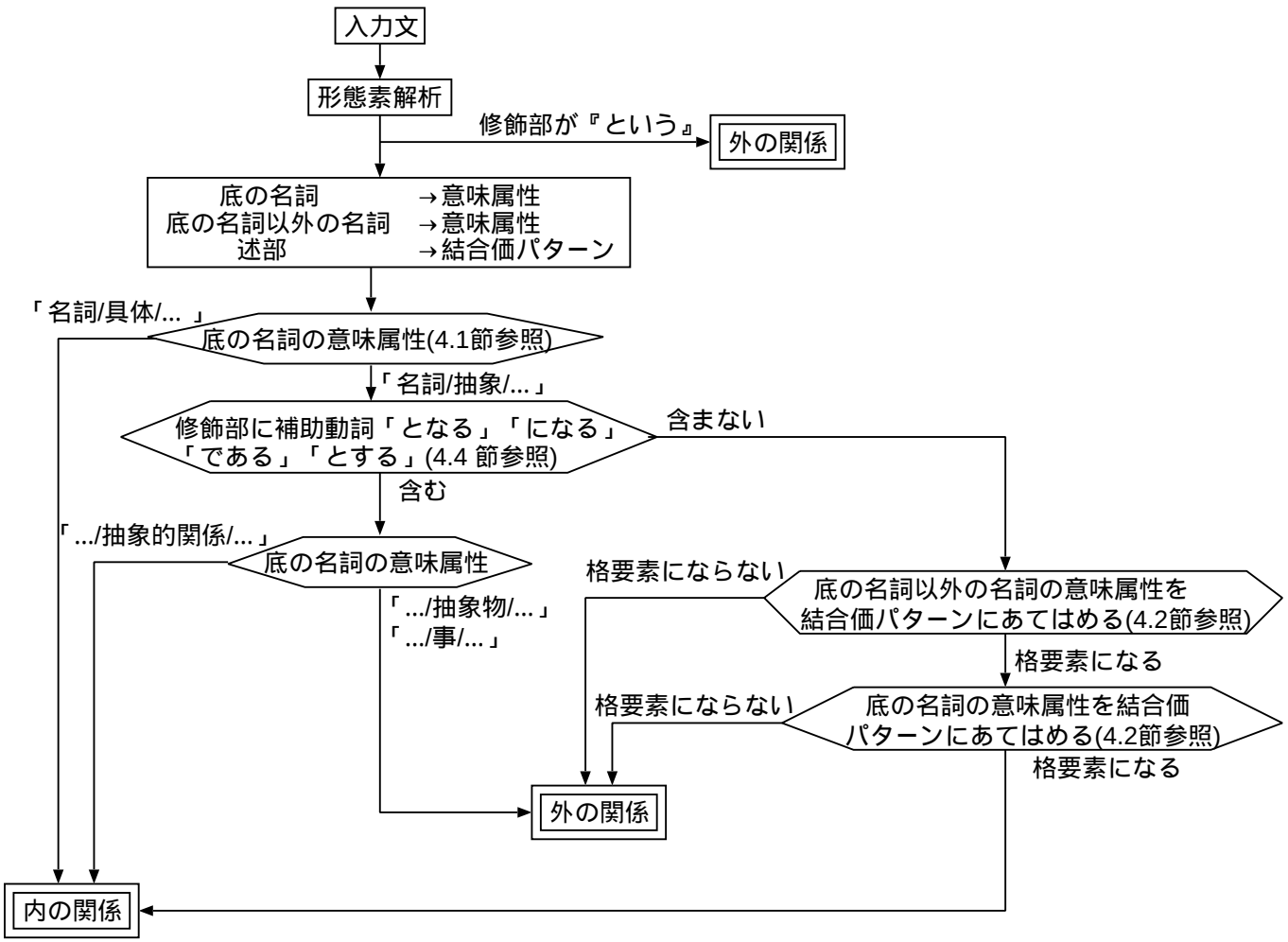

図 6: 「内と外の関係」の決定規則の流れ

6　実験と結果

本手法で提案した判断規則の有効性を，「アンカー和英辞典」(約 46000 文) より抽出した複文 200 文を用いて人手で検証した．なお，本実験は決定規則作成と同一のデータを用いたため，クローズドテストとなる．

次に「アンカー和英辞典」より複文 400 文を抽出して人手でオープンテストを行った．評価は，内と外の関係が一意に決まる場合を◎ ，底の名詞の意味属性が複数存在し，候補の中に正解を含む場合を○ ，不正解の場合を× とした．さらに，底の名詞の意味属性が複数存在し，評価が○ の場合を，意味属性により正解となる場合 (*) と不正解となる場合 (△) の 2 通りで評価した．

6.1　クローズドテスト

クローズドテスト (「アンカー和英辞典」より抽出した複文 200 文) の結果を表 8 に示す．

表 8: 評価結果

	クローズドテスト (200 文)		
◎	**70.5%(141 文)**		
○	**21.5%(43 文)**	*	**36.7%**
		△	**63.7%**
×	**8.0%(16 文)**		

その結果，クローズドテストにおいては 78.1%の正解率を得た．

次に，クローズドテストにおける「内の関係」と「外の関係」の正解率を表 9 に示す．

表 9: 評価結果 (クローズドテスト)

	クローズドテスト (200 文)					
	内の関係 (140 文)			外の関係 (60 文)		
◎	**69.3%(97 文)**			**73.4%(44 文)**		
○	**24.5%(35 文)**	*	**53.5%**	**13.3%(8 文)**	*	**75.0%**
		△	**46.5%**		△	**25.0%**
×	**5.7%(8 文)**			**13.3%(8 文)**		

その結果「内の関係」については 82.4%「外の関係」については 83.4%の正解率を得た．

6.2 オープンテスト

オープンテスト (「アンカー和英辞典」より抽出した複文 400 文) の結果を表 10 に示す．

表 10: 評価結果

	オープンテスト (400 文)		
◎	**71.0%(284 文)**		
○	**17.5%(70 文)**	*	**54.7%**
		△	**45.3%**
×	**11.5%(46 文)**		

その結果，オープンテストにおいては 80.1%の正解率を得た．

次に，オープンテストにおける「内の関係」と「外の関係」の正解率を表 11 に示す．

表 11: 評価結果 (オープンテスト)

	オープンテスト (400 文)					
	内の関係 (275 文)			外の関係 (125 文)		
◎	**72.7%(200 文)**			**67.2%(84 文)**		
○	**21.1%(58 文)**	*	**57.5%**	**9.6%(12 文)**	*	**59.3%**
		△	**42.5%**		△	**40.7%**
×	**6.2%(17 文)**			**23.2%(29 文)**		

その結果「内の関係」については 85.2%「外の関係」については 73.1%の正解率を得た．

また，評価が○ の場合について複数存在する底の名詞の意味属性の割合を調べた．その結果を表 12 に示す．

表 12: 複数存在する底の名詞の意味属性の割合 (評価○)

意味属性の総数 (個)	最大意味属性数 (個)	平均意味属性数 (個)
226	6	3.2

その結果，底の名詞の意味属性の総数は 226 個となり，意味属性の最大数は 6 個となった．また，意味属性の平均数は 3.2 個となった．意味属性の平均数が高いため，今後一意に決定する必要があると考えられる．

7　正解と失敗の例

7.1　正解例 (評価が◎ の場合)

(例文 18) 銀行は人から預かった 金 で利潤を得る。(評価 ◎)(クローズドテスト)

表 13: 形態素解析の結果 (例文 18)

1. 銀行 (1100,[374,428])/は (7530)
2. 人 (1100,[2606,5,33])/から (7410)
3. 預かっ(2384, 預かる)/た (7217)
4. 金 (1100,[715,710])(1100,[1190,934])/で (7410)
5. 利潤 (1100,[1179])/を (7430)
6. 得る (2416,[1868])/。 ([P]0110)

表 14: 意味属性 (例文 19)

人 (意味属性)
[/1 名詞/1000 抽象/2422 抽象的関係/2585 数量/2601 組/2604 単複/2606 複数]
[/1 名詞/2 具体/3 主体/4 人/5 人間]
[/1 名詞/2 具体/3 主体/4 人/5 人間/6 人間〈人称〉/31 自他/33 他人]
金 (意味属性)
[/1 名詞/2 具体/533 具体物/706 無生物/707 自然物/712 物質(本体)/713 固体/
714 金属/715 貴金属]
[/1 名詞/2 具体/533 具体物/706 無生物/707 自然物/708 物質 (部分)/710 元素]
[/1 名詞/1000 抽象/1001 抽象物/1154 抽象物(行為)/1155 制度/1168 制度 (経済)/
1187 資本・金銭等/1190 金銭]
[/1 名詞/2 具体/533 具体物/706 無生物/760 人工物/893 道具/925 目印・象徴物/934 貨幣]

表 15: 結合価パターン (例文 19)

○ 預かる (結合価パターン)
$N1$ が $N2$ を $N3$ から 預かる
[$N1$(3 主体) $N2$(760 人工物 1168 制度 (経済)) $N3$(3 主体)]
[$N1$(3 主体) $N2$(3 主体 534 生物 389 施設 458 地域 2537 立場) $N3$(3 主体)]
$N1$ が $N2$ を 預かる
[$N1$(3 主体) $N2$(" 家計 ")]
[$N1$(3 主体) $N2$(" 喧嘩/ケンカ " 1754 紛争)]

例文 19 の場合，形態素解析の結果と意味属性，結合価パターンは各々表 13・表 14・表 15 のようになる．底の名詞「金」の意味属性が「具体」配下のときは，「内の関係」となり正解となる．また，「金」の意味属性が「/1 名詞/1000 抽象/1001 抽象物/1154 抽象物 (行為) /1155 制度/1168 制度 (経済)/1187 資本・金銭等/1190 金銭」のときは，底の名詞以外の名詞「人」の意味属性が「預かる」の結合価パターン「$N1$ が $N2$ を $N3$ から 預かる」の $N3$ に当てはまる．次に，底の名詞「金」の意味属性が $N2$ に当てはまるため，「内の関係」となり正解となる．

7.2 失敗例 (評価が× の場合)

(例文 20) 城の堀は敵の侵入を防ぐ目的でつくられた。(評価 ×)(オープンテスト)

表 16: 形態素解析の結果 (例文 20)

1. 城 (1100,[449])/の (7410)
2. 堀 (1100,[502,451,497])/は (7530)
3. 敵 (1100,[123])/の (7410)
4. 侵入 (1210,[1766,2185])/を (7430)
5. 防ぐ (2327)
6. 目的 (1100,[2456,2651])/で (7410)
7. つくら (2382, 創る)(2382, 造る)(2382, 作る)/れ (7113, れる)/た (7216)/。([P]0110)

表 17: 意味属性 (例文 20)

侵入
[/1 名詞/1000 抽象/1235 事/1236 人間活動/1560 行為/1681 交際/1752 争い/
1764 討伐・守備/1766 侵略]
[/1 名詞/1000 抽象/1235 事/2054 事象/2064 変動/2182 出・入/2183 出・入り/2185 入り]

目的
[/1 名詞/1000 抽象/2422 抽象的関係/2443 関連/2454 理由・目的等/2456 目的]
[/1 名詞/1000 抽象/2422 抽象的関係/2610 場/2649 方向/2651 方面]

表 18: 結合価パターン (例文 20)

○ 防ぐ
$N1$ が $N2$ を 防ぐ
[$N1$(3 主体) $N2$(2521 危険 1762 攻撃 1766 侵略)]
[$N1$(*) $N2$(533 具体物 2306 物象 2373 風〔かぜ〕)]
[$N1$(*) $N2$(*)]

例文 20 の場合，本来は底の名詞と修飾部の間に格関係がなく，「外の関係」となるべきところ，「防ぐ」の結合価パターンの「$N1$(*) が $N2$(*)」に当てはまってしまい，本手法では，格関係があるとし「内の関係」と判断した．ここで，(*) はどの意味属性でも当てはまることを意味する．

(例文 21) この町を走るバスはすべてワンマンだ。(評価×)(オープンテスト)

表 19: 形態素解析の結果 (例文 21)

1. この (4200)
2. 町 (1100,[465,464,364,418])/を (7430)
3. 走る (2187)
4. バス (1100,[2669])/は (7530)
5. すべて (1110,[2599], 凡て)
6. ワンマン (1100,[323,335,167,2488])/だ (7256)/。([P]0110)

表 20: 意味属性 (例文 21)

町
[/1 名詞/2 具体/388 場所/458 地域/465 都市]
[/1 名詞/2 具体/388 場所/458 地域/464 行政区画]
[/1 名詞/2 具体/3 主体/362 組織/363 機関/364 行政機関]
[/1 名詞/2 具体/388 場所/389 施設/390 公共施設/417 交通路/418 道路]

バス
[/1 名詞/1000 抽象/2422 抽象的関係/2610 場/2669 経路]

表 21: 結合価パターン (例文 21)

○ 走る
N1 が N2 を 走る
　[N1(“ 稲妻 ” 2345 光) N2(532 空)]
①[N1(989 乗り物 (本体 (移動 (水圏)))) N2(388 場所 2610 場)]
②[N1(4 人 535 動物 988 乗り物 (本体 (移動 (陸圏))) 417 交通路 1148 評判)
N2(388 場所 2610 場)]
　[N1(4 人 535 動物) N2(“ 距離 ”)]
　[N1(1239 感覚) N2(552 動物 (部分))]

N1 が N2 に 走る
　[N1(3 主体) N2(2046 犯罪)]
　[N1(3 主体) N2(“ 党利党略 ” 1179 利益 1890 利得)]
　[N1(1242 痛み) N2(552 動物 (部分))]

N1 が N2 から/より N3 に/へ/まで 走る
　[N1(4 人 535 動物 954 車 417 交通路) N2(388 場所 2610 場) N3(388 場所 2610 場)]

例文 21 の場合，底の名詞以外の名詞「町」の意味属性が「走る」の結合価パターンの①・②の $N2$ に当てはまり，結合価パターンが①・②に絞られる．しかし，底の名詞「バス」の意味属性が $N1$ に当てはまらなかったため，本来「内の関係」と判断すべきところを誤って「外の関係」と判断した．

7.3　評価が○ の例

(例文 22) 日給で働く人の生活は不安定だ。(評価○)(オープンテスト)

表 22: 形態素解析の結果 (例文 22)

1. 日給 (1100,[1194])/で (7410)
2. 働く (2117)(2317)
3. 人 (1100,[2606,5,33])/の (7410)
4. 生活 (1210,[1581,1171])/は (7530)
5. 不安定だ (3226,[2535])/。([P]0110)

表 23: 意味属性 (例文 22)

日給
[/1 名詞/1000 抽象/1001 抽象物/1154 抽象物 (行為) /1155 制度/1168 制度 (経済)/1193 賃金・料金・利子等/1194 給与]

人
①[/1 名詞/1000 抽象/2422 抽象的関係/2585 数量/2601 組/2604 単複/2606 複数]
②[/1 名詞/2 具体/3 主体/4 人/5 人間]
③[/1 名詞/2 具体/3 主体/4 人/5 人間/6 人間〈人称〉/31 自他/33 他人]

表 24: 結合価パターン (例文 22)

○ 働く
*N*1 が 働く
[*N*1(4 人)]

*N*1 が *N*2 を 働く
[*N*1(4 人) *N*2(2045 犯罪・騒ぎ 1214 害悪)]

*N*1 に *N*2 が 働く
[*N*1(527 天体) *N*2(“ 引力/重力 ”)]

例文 22 の場合，底の名詞「人」の意味属性が②・③のとき，述部「働く」の結合価パターンに当てはまるので，「内の関係」と判断し正解となる．しかし，底の名詞の意味属性が①のとき，「働く」の結合価パターンに当てはまらず「外の関係」と判断し不正解となる．

8　考察

8.1　失敗となる原因 (オープンテスト)

本手法で提案した「内と外の関係」の判断規則での失敗の原因として，主に以下の 4 つがあげられる．

1. 結合価パターンに全て (*) を含む (**15** 文/**46** 文)

(例文 23) 彼は原因も治療法もわからない難病にかかっている．(オープンテスト)

表 25: 形態素解析の結果 (例文 23)

1. 彼 (1710,[23,48])/は (7530)
2. 原因 (1210,[2450])/も (7530)
3. 治療 (1220,[1985,2006,1035])/法 (6280,[1035])/も (7530)
4. わから (2182, 分かる)/ない (7187)
5. 難病 (1100,[2419])/に (7430)
6. かかっ(2184, 掛かる)(2184, 罹る)/ている (2816)/。([P]0110)

表 26: 意味属性 (例文 23)

難病
[/1 名詞/1000 抽象/1235 事/2304 自然現象/2385 生命現象/2413 健康・不全/2415 不全/2416 病気/2419 病気類]

表 27: 結合価パターン (例文 23)

○ 分かる $N1$ が $N2$ と 分かる [$N1$(*) $N2$(*)]
$N1$ が $N2$ に 分かる [$N1$(1210 業績 2451 結果) $N2$(3 主体)] [$N1$(1037 創作物) $N2$(4 人)] [$N1$(2701 未来) $N2$(3 主体)]
$N1$ が $N2$ に $N3$ で 分かる [$N1$(1002 抽象物 (精神) 1236 人間活動) $N2$(3 主体) $N3$(1064 言葉)]
$N1$ が $N2$ に $N3$ と 分かる [$N1$(*) $N2$(3 主体) $N3$(*)]
$N1$ に $N2$ が 分かる [$N1$(3 主体) $N2$(1030 要点)]

例文 23 では，述部「分かる」の結合価パターン「$N1$(*) が $N2$ (*) と分かる」のように，結合価パターンに全て (*) を含むために失敗する．

ここで，(*) はどの意味属性でも当てはまることを意味するので，「外の関係」と判断すべきところを，「内の関係」と判断してしまう．このような例については，結合価パターンの精度向上に依存するか，新たに規則を作成する必要があると考えられる．

2. 結合価パターンが日本語語彙大系に未登録 (5 文/46 文)

「並外れる , 知り合う , 一見する」のように日本語語彙大系に登録されていないもの . その内 ,「一見する」のように日本語語彙大系に未登録のサ変型動詞については ,「する」の結合価パターンを用いて「一見をする」と変換して用いることで解決できると考えられる .

(例文 24) その寺の庭は一見する値打ちがある。(オープンテスト)

表 28: 形態素解析の結果 (例文 24)

1. その (4200, 其の)
2. 寺 (1100,[454,377])/の (7410)
3. 庭 (1100,[889,423])/は (7530)
4. 一見する ([*]2637,[1])
5. 値打ち (1100,[2532,2590,2441])/が (7410)
6. ある (2186)/。 ([P]0110)

表 29: 意味属性 (例文 24)

値打ち
[/1 名詞/1000 抽象/2422 抽象的関係/2507 状態/2508 様相/2532 是非]
[/1 名詞/1000 抽象/2422 抽象的関係/2585 数量/2587 量/2590 値・額]
[/1 名詞/1000 抽象/2422 抽象的関係/2432 類・系/2441 等級]

表 30: 結合価パターン (例文 24)

一見する ... パターン未登録

また ,「知り合う」のような複合動詞については ,「知る」の結合価パターンを用いることで解決できると考えられる .

3. 形態素解析の失敗 (**19** 文/**46** 文)

(例文 25) 電車の乗客は座席に寝そべっている 酔っぱらい を見て見ないふりをした。
(オープンテスト)

表 31: 形態素解析の結果 (例文 25)

1. 電車 (1100,[988])/の (7410)
2. 乗客 (1100,[160])/は (7530)
3. 座席 (1100,[2612,460])/に (7430)
4. 寝そべっ(2184, 寝そべる)/ている (2817)
5. 酔っぱらい (2193, 酔っ払う)
6. を ([*]1100,[1])
7. 見 (2413,[1464,1819], 見る)/て (7630)
8. 見 (2411,[1464,1819], 見る)/ない (7187)
9. ふり (1420,[2563,2595,2588,2111,1289], 振り)/を (7430)
10. し (2433, する)/た (7216)/。 ([P]0110)

例文 25 の場合，底の名詞は「酔っぱらい」であるが，形態素解析の結果では「酔っ払う」という動詞で判断したため，本手法が適用できなかった．このような例については，形態素解析の向上，または，漢字入力により解決できると考えられる．

酔っぱらい (2193, 酔っ払う) ⇒ 酔っ払い (1410)

(注) 数字は品詞を示す番号であり，(2***) は動詞，(1***) は名詞を示す．

9　あとがき

本研究では，「底の名詞の意味属性」及び，「修飾部の述部の結合価パターン」に着目して，日本語複文 (埋め込み文) を「内の関係」と「外の関係」に分類する判断規則を作成した．「内と外の関係」の判断規則をアンカー和英辞典 (約 46000 文) を用いて評価した結果，クローズドテスト (200 文) においては，「内の関係」で 82.2%，「外の関係」で 83.4%の正解率が得られた．オープンテスト (400 文) においては，「内の関係」で 84.9%，「外の関係」で 71.1%の正解率が得られ，本研究で提案した判断規則の有効性が示された．

しかし，意味属性に複数の候補が存在する場合が多く，精度向上のためには意味属性を絞り込む必要があると考えられる．

今後の課題としては，結合価パターンに (*) を含む場合の規則作成と，底の名詞の意味属性が複数存在する場合の一意決定手法を考える必要がある．

謝辞

最後に本研究において一年間御指導頂いた鳥取大学工学部知能情報工学科計算機 C 研究室の池原教授と村上助教授にお礼申し上げます．また，本研究で alt-jaws を使用させて下さったコミュニケーション科学基礎研究所の方々，および，本研究に使用させて頂いた論文，本の著者の方々にお礼申し上げます．

参考文献

[1] 山田 孝雄：日本文法論，宝文館 (1898)．

[2] 寺村 秀夫：日本語のシンタクスと意味I ～山, くろしお出版 (1982～1991)．

[3] 藤本，表，池原，村上：埋め込み文の日英翻訳方式について，情報処理学会第 63 回全国大会，pp.2-264-265(2001).

[4] 池原, 宮崎, 白井, 横尾, 中岩, 小倉, 大山, 林: 日本語語彙大系 1. 意味体系,5. 構文体系，岩波書店 (1997)．