複文における底の名詞と修飾部の内と外の関係の判断規則

藤本 敬史　池原 悟　村上 仁一　表 克次
鳥取大学工学部知能情報工学科
{fujimoto,ikehara,murakami}@ike.tottori-u.ac.jp

1. はじめに

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．

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．

2. 内と外の関係について

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1参照)

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1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，例文1のように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，例文2のように格関係を含まず，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．

3. 結合価パターンと意味属性

「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 [3] に記載されている「意味体系」と「構文体系」を用いて格関係の有無を判断する．「意味体系」では，約40万語の一般名詞が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2参照)

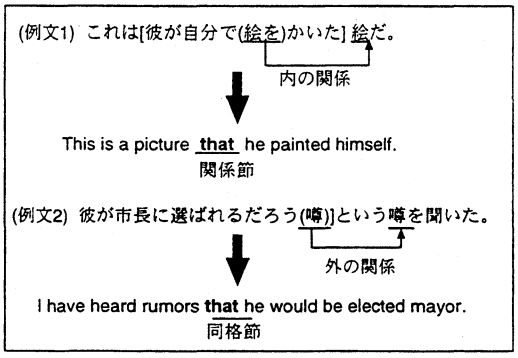

図 1: 「内の関係」と「外の関係」

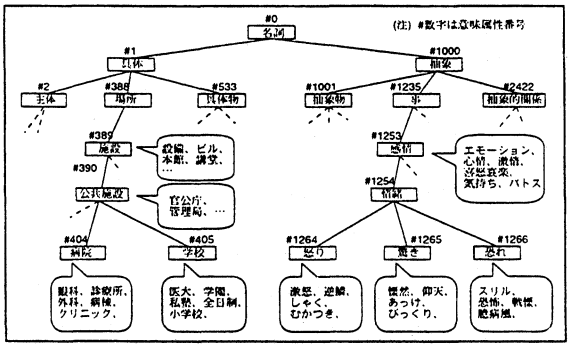

図 2: 一般名詞意味属性体系 (上位属性)

また，「構文体系」の結合価パターン (図参照) 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．

4. 内と外の関係の判断方法

4.1 底の名詞の一般名詞意味属性

従来の研究 [4] より，底の名詞の意味属性が「名詞/具体/…」となる場合，約99%が「内の関係」であり，底

彼女に赤ちゃんが生まれる	自分が生まれた場所
$N1$ に $N2$ が 生まれる	自動詞
[$N1$(4 人) $N2$(86 子)]	
必須格	任意格

図 3: 結合価パターンの例

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．ここで「…」はその属性の配下を示す．

4.2 固有名詞の意味属性

4.1 節より底の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 1 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．

表 1: 一般名詞と固有名詞の意味属性対応表

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

4.3 述部の結合価パターン

4.1 節より底の名詞が「具体」配下の場合,「内の関係」とみなすことができるが，底の名詞が「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を考える．修飾部に「という」を含む場合は，明らかに「外の関係」となる．底の名詞が「抽象」配下の場合，動詞により「内の関係」と「外の関係」を判断できると考えられる (図 4 参照)．具体的には，動詞「実施する」は，「$N1$ が $N2$ を実施する」の結合価パターンを持ち，底の名詞「試験」の意味属性 […/抽象/事/人間活動/…] が当てはまるため，格関係を持ち「内の関係」となる．また，動詞「飲む」の結合価パターンには，底の名詞「時間」の意味属性が当てはまらないので格関係がないとし,「外の関係」とする．

国が実施する試験 (内の関係)
コーヒーを飲む時間 (外の関係)

図 4: 動詞と「内と外の関係」

4.4 補助動詞

4.3 節より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」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 2 に示す．

表 2: 補助動詞と「内と外の関係」

	抽象物	事	抽象的関係
内の関係	1	1	6
外の関係	3	3	3

表 2 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．

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.「結合価パターン」を適用するために，次のように変更する．

(例文 3) 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 4) 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.

4.6 本手法の判定方法

本手法では,「内と外の関係」を「名詞の意味属性」と「結合価パターン」をもとに決定する．図 5 に判断規則を示す．

5. 評価

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 文を用いてクローズドテストを行った．次に「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．結果を表 3 に示す．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，評価が○の場合を，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．

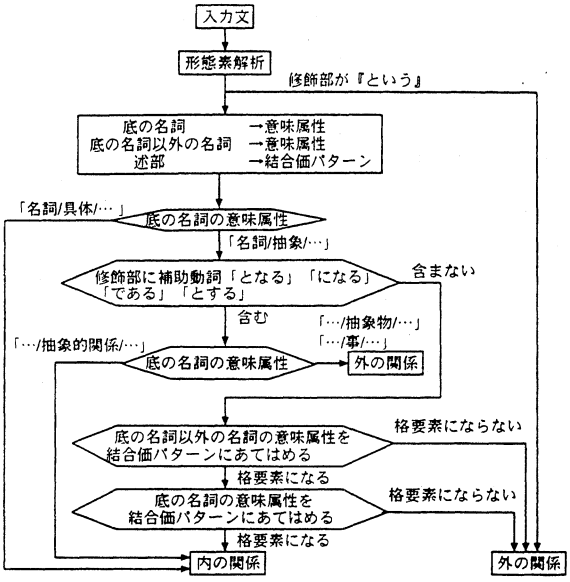

図 5: 「内と外の関係」の決定規則の流れ

なお，データベースには，著者が判断して予め正解を設定しておいた．

表 3: 評価結果

	クローズドテスト (200 文)			オープンテスト (400 文)		
◎	70.5%			68.9%		
○	21.5%	*	36.7%	16.6%	*	54.7%
		△	63.7%		△	45.3%
×	8.0%			14.5%		

その結果，クローズドテストにおいては 78.1%，オープンテストにおいては 78.0%の正解率を得た．

次に，クローズドテストとオープンテストにおける「内の関係」と「外の関係」の正解率を各々表 4・表 5 に示す．

表 4: 評価結果 (クローズドテスト 200 文)

	内の関係 (140 文)		
◎	69.1%(97 文)		
○	24.5%(35 文)	*	53.5%
		△	46.5%
×	6.4%(9 文)		
	外の関係 (60 文)		
◎	73.4%(44 文)		
○	13.3%(8 文)	*	75.0%
		△	25.0%
×	13.3%(8 文)		

表 5: 評価結果 (オープンテスト)

	オープンテスト (400 文)					
	内の関係			外の関係		
◎	72.6%			65.0%		
○	20.7%	*	59.2%	10.0%	*	61.2%
		△	40.8%		△	38.8%
×	6.7%			25.0%		

6. 成功と失敗の例

6.1 成功例

(例文 5) 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 6: 形態素解析の結果 (例文 5)

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 7: 意味属性 (例文 5)

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 8: 結合価パターン (例文 5)

○預かる (結合価パターン)
*N*1 が *N*2 を *N*3 から 預かる
[*N*1(3 主体) *N*2(760 人工物 1168 制度 (経済)) *N*3(3 主体)]
[*N*1(3 主体) *N*2(3 主体 534 生物 389 施設 458 地域 2537 立場) *N*3(3 主体)]

*N*1 が *N*2 を 預かる
[*N*1(3 主体) *N*2(“家計”)]
[*N*1(3 主体) *N*2(“喧嘩/ケンカ” 1754 紛争)]

例文 5 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 6・表 7・表 8 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「*N*1 が *N*2 を *N*3 から 預かる」の *N*3 に当てはまる．次に，底の名詞「金」の意味属性が *N*2 に当てはまるため,「内の関係」となり正解となる．

6.2 失敗例

(例文 6) 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 9: 形態素解析の結果 (例文 6)

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 10: 意味属性 (例文 6)

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 11: 結合価パターン (例文 6)

○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))]

例文 6 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 9・表 10・表 11 のようになる. 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. 考察

7.1 失敗となる原因

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 2 つがあげられる.

7.1.2 結合価パターンに全て (*) を含む

(例文 7) 彼は原因も分からない 難病 にかかっている. (オープンテスト)

例文 7 では, 述部「分かる」の結合価パターン「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, 本来「外の関係」と判断すべきところを,「内の関係」と判断した. 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.

7.1.2 結合価パターンが日本語語彙大系に未登録

(例文 8) その寺は一見する 値打ち がある. (オープンテスト)

「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.

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.

8. おわりに

本研究では, 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に着目して「内と外の関係」の判断規則を作成した.「内と外の関係」の判断規則を評価した結果, クローズドテストでは,「内の関係」で 82.2%,「外の関係」で 83.4%の正解率が得られた. オープンテストでは,「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. 今後は, 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.

参考文献

[1] 山田 孝雄 : 日本文法論, 宝文館 (1898).

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

[3] 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).

[4] 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).