6Y-03 埋め込み文の日英翻訳方式について
—内と外の関係の判断方法—

藤本敬史　表克次　池原悟　村上仁一
鳥取大学工学部知能情報工学科

1　はじめに

日本語複文 (埋め込み文) は、英語に翻訳したとき、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。これ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 1 つとなっている。その課題の 1 つが、「内と外の関係」である。

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」を考慮することにより内と外の関係を計算機で判断する方法を提案する。

2　内と外の関係について

寺村 [1] 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。図 1 に示すように、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、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。

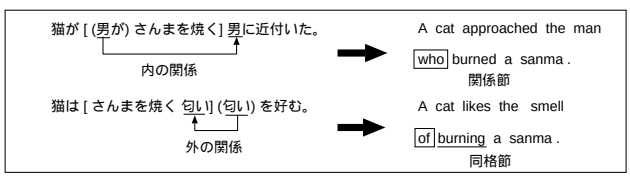

図 1: 「内の関係」と「外の関係」

3　日本語語彙体系

「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、日本語語彙体系 [2] を用いて格関係の有無を判断する。池原らによって作成された日本語語彙体系は「意味体系」「単語体系」「構文体系」により構成されている。本研究ではその中で「意味体系」と「構文体系」を用いる。「意味体系」の中で名詞はその意味属性ごとに 2710 種類、12 段の木構造に分類されており、上位の属性は下位の属性を内含するという性質を持つ (図 2 参照)。また、「構文体系」は用言の文型を結合価パターンにまとめている。

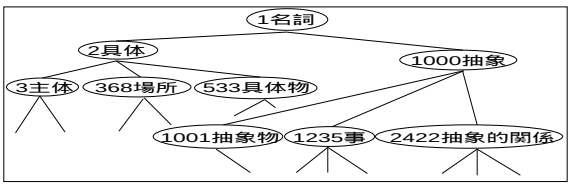

図 2: 一般意味属性体系 (上位属性)

4　結合価パターンと格

「構文体系」において、結合価パターン (図 3 参照) として登録されている格関係は『必須格』とされ、動詞にとって必須な格関係である。また、登録されていない格関係は『任意格』とされる。これは、「時」「場」などのようになくてもよいものであるが格とみなすことができる。従って、日本語語彙体系の「構文体系」に登録されている結合価パターンを用いることで格関係の有無を決定できる。

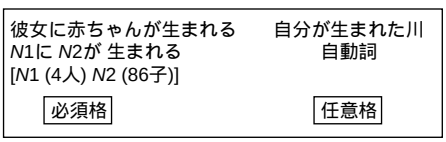

図 3: 「必須格」と「任意格」

5　内と外の関係の判断方法

5.1　底の名詞の意味属性

「アンカー和英辞典」(約 46000 文) 中よりランダムに抽出した複文 348 文を対象に底の名詞の意味属性を日本語語彙体系によって調べた。その結果を表 1 に示す。ここで、表 1 中の「... 」はその属性の配下を表す。

表 1: 「内と外の関係」と意味属性

	内の関係	外の関係	合計
/1名詞/2具体/...	118	1	119
/1名詞/1000抽象/...	42	83	125
合計	160	84	244

表 1 より以下のことが分かる。

1. 底の名詞の意味属性が「**/1** 名詞/**2** 具体/... 」となる場合、ほぼ「内の関係」となる。

2. 底の名詞の意味属性が「**/1** 名詞/**1000** 抽象/... 」となる場合、半分以上「外の関係」となる。

English Translations of Japanese Embedded Clauses
Takashi FUJIMOTO,Katsuji OMOTE,Satoru IKEHARA,Jin'ichi MURAKAMI
Tottori University

5.2 述部の結合価パターン

5.1 節より「具体」配下の名詞は「内の関係」とみなすことができる。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について考える。

修飾部に「という」がある場合は明らかに「外の関係」となる。意味属性が「抽象」配下の場合、動詞によっては「内の関係」にも「外の関係」にもなりうる (図 4)。具体的には、「聞く」という動詞は「$N1$ が $N2$ を 聞く $[N1\ (3\text{ 主体})\ N2\ (1009\text{ 意見})]$」という結合価パターンを持ち「.../1009 意見/」の意味属性を持つ「意見」という名詞との間に格関係を持つので「内の関係」と言える。また、「行く」という動詞の結合価パターンでは「意見」という名詞の意味属性は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。

彼から聞いた 意見 (内の関係)
山に行きたいという 意見 (外の関係)

図 4: 動詞と格関係

5.3 任意格

5.2 節より「内と外の関係」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」に依存していると言える。しかし、「任意格」は無くてもよい格関係のため、結合価パターンが登録されていない。そこで、「任意格」における格関係の決定方法を考える。

「任意格」をとる動詞は主に自動詞である。自動詞の大きな意味としては「目的格をとらない」と定義されている。そのため、結合価パターンが登録されていない。そこで、ガ格・ヲ格・ニ格・デ格がどのような意味属性の名詞をとるか、「内の関係」となる文を対象に調べた。その結果を表 2 に示す。

表 2: 意味属性による格関係の決定

	.../1001抽象物/...	.../1235事/...	.../2422抽象的関係/...
ガ格	(8)	6	2
ヲ格	(11)	6	0
ニ格	1	1	(6)
デ格	1	4	2

表 2 から、「底の名詞の意味属性」により底の名詞がとる格に規則性があることが確かめられた。そのため、「任意格」において「底の名詞の意味属性」を用いることで格関係が決定できる。表 2 の○は本手法において「底の名詞の意味属性」ごとに、どの格をとるかを示している。「1001 抽象物」については、ガ格またはヲ格のどちらかをとる。「1235 事」はデータにばらつきがあるため決定できない。

5.4 本手法の判定方法

本手法では、「格関係の有無」と「格関係」を「底の名詞の意味属性」と「必須格」をもとに決定し、決定できないものに関しては「任意格」をもとに決定する。図 5 に「内と外の関係」の判断規則の流れを示す。

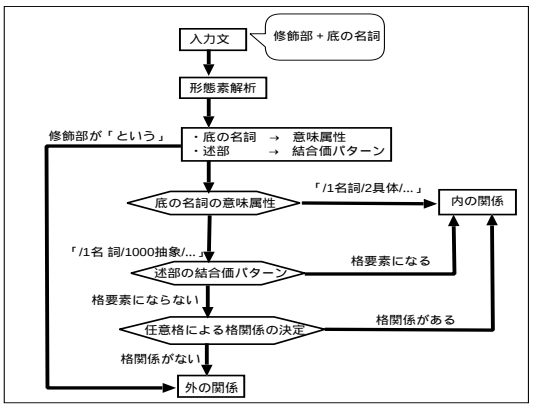

図 5: 「内と外の関係」の決定規則の流れ

6 評価

本手法で提案した「内と外の関係」の判断規則の有効性を、「アンカー和英辞典」(約 46000 文) よりランダムに抽出した複文 204 文を用いて検証した。本実験は規則作成に用いたデータと同一のデータを用いたためクローズドテストである。結果は表 3 に示すように、内の関係については 94.9%、外の関係については 63.8%の正解率を得た。成功例と失敗例を図 6 に示す。

表 3: 「内と外の関係」の正解率

		本手法の結果			
		内の関係	外の関係	小計	合計
人手による結果	内の関係	149(94.9%)	8(5.1%)	157	204
	外の関係	17(36.2%)	30(63.8%)	47	

・成功例
(例文1)暮らしに困る人
本手法による結果...内の関係
人手による結果 ...内の関係

・失敗例
(例文2)救助された登山者
本手法による結果...外の関係
人手による結果 ...内の関係
登山者...日本語彙体系に無い

図 6: 失敗と成功の例

例文 2 のように、「行商人、登山者、...」など日本語語彙大系 [2] に登録されていない名詞については、日本語語彙大系 [2] の精度を上げる必要がある。または、接尾辞「人、者」によって意味属性を決定することで解消されると思われる。

7 おわりに

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、「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、「内の関係」で 94.9%、「外の関係」で 63.8%の正解率が得られ、判断規則の有効性が示された。

今後は、「内と外の関係」の判断規則の精度を向上するため、複数の意味属性を持つ底の名詞の意味属性の決定や、修飾部内の名詞の意味属性による格関係の絞り込みを考慮する必要がある。

参考文献

[1] 寺村 秀夫 : 日本語シンタクスと意味I ~山、くろしお出版 (1982 ~ 1991)

[2] 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)