

埋め込み文に対する日英翻訳規則の作成について

藤本 敬史 村上 仁一 徳久 雅人 池原 悟

鳥取大学工学部知能情報工学科

{fujimoto,murakami,tokuhisa,ikehara}@ike.tottori-u.ac.jp

1 はじめに

日英翻訳において、「太郎が学校に行く理由」や「東京に行く高速バス」のような埋め込み文は、関係節、同格節、前置詞句など多種多様な表現に訳し分ける必要がある。しかし、機械的に埋め込み文を正しく訳し分けることは困難であり、日英機械翻訳の重要な課題の一つとなっている。

従来の研究として、埋め込み文を「修飾部+底の名詞」の構造で分類することによって、英語構造に対応付ける方法が提案されている [1]。また、修飾部と底の名詞の間で生じる格関係の種類に応じて分類する方法 [2] 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 (内と外の関係) [3] もある。しかし、何れの方法においても、計算機で分類することは困難であった。

これに対して、計算機による「内と外の関係」の判定方法 [4, 5] が提案され、その有効性が示されている。しかし、[4, 5] は日本語側の構造の分類にのみ着目していたため、英語側との対応付けを行うまでには至っておらず、翻訳規則としては今だ不完全である。

そこで本稿では、埋め込み文における日英機械翻訳の精度向上を計るため、埋め込み文を日本語側の構造及び名詞の意味属性に着目して分類し、日本語構造と英語構造の対応関係をまとめた翻訳規則を作成する。さらに、作成した翻訳規則に対して人手による机上実験を行い、評価・検討する。

2 埋め込み文の構造

本稿では、「修飾部+底の名詞」で構成される箇所を含んだ埋め込み文を対象とする。ここで、「底の名詞」とは修飾部の係り先となる名詞を意味し、英語の先行詞等に対応する。さらに、埋め込み文を修飾部と底の名詞の間に格関係が有る場合を「内の関係」、格関係が無い場合を「外の関係」として分類し本研究で使用する。

以下に埋め込み文の例及び各々の対訳を 2 つ示す。例文中において埋め込み箇所 (修飾部+底の名詞) を波線で示す。

(例文 1) この窓から見る町の風景はすばらしかった。(内の関係)

(対訳 1) The townscape from this window is pretty dismal.

(例文 2) 普通は、学校に行く時間だ。(外の関係)

(対訳 2) Normally, it is the time when children go to school.

コーパスには、複数の電子辞書より重文複文を抽出した日英対訳の例文集 (CREST コーパス [6]: 約 12 万文) を使用する。

3 一般名詞意味属性体系

埋め込み文における日英機械翻訳の規則作成にあたり、底の名詞及び修飾部内の名詞の意味属性を用いる。意味属性には、日本語語彙大系 [7] に記載されている「一般名詞意味属性体系」を使用する。

一般名詞意味属性は、名詞の意味的用法に着目してシソーラスとして体系化されている。一般名詞意味属性体系は、登録単語数約 40 万語、最大 12 段の木構造であり、2,710 の意味属性に分類されている。また、各ノードにおいて上位の意味属性の性質を下位の意味属性が継承する (図 1 参照)。

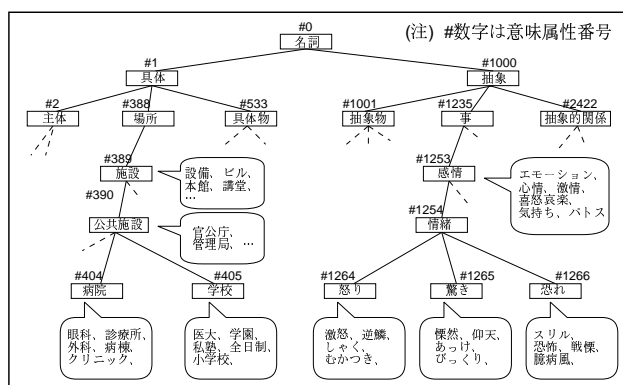


図 1: 一般名詞意味属性体系 (上位属性)

4 日英パターン対応表の作成

本研究では、使用頻度の高い用言 (「行く」「見る」「思う」「生じる」「生まれる」) を埋め込み箇所を含む文を対象とする。

日英パターン対応表の作成手順を以下に示す。

1. 日本語パターンの抽出

2. 英語パターンの抽出
3. 日本語パターンと英語パターンの対応付け

以下に各々の手順の詳細を述べる。

4.1 日本語パターンの抽出

埋め込み文の日本語パターンを以下の手順により抽出する。

(1) 日本語表現の抽出

CREST コーパスより日本語の埋め込み表現を抽出する。例文と抽出例を以下に示す。下線は「底の名詞」を意味する。

(例文 3) 普通は、学校に行く時間だ。

(抽出例) 学校に行く 時間

(2) 名詞及び形容詞の変数化

形態素解析結果より、名詞及び形容詞を変数化する。

本稿で使用する品詞の変数を、表 1 に示す。

表 1: 変数一覧

変数名	品詞名
<i>N</i>	名詞
<i>V</i>	動詞
<i>ADJ</i>	形容詞
<i>ADV</i>	副詞
<i>REL</i>	関係代名詞
<i>REL.ADV</i>	関係副詞
<i>CONJ</i>	接続詞
<i>PRON</i>	代名詞
<i>ZERO</i>	補完されたゼロ代名詞
<i>AUX.V</i>	助動詞

変数化の際、以下の 3 点に注意する。

1. 時制は現在形に修正
2. 日本語において副詞を除外
3. A の B 型名詞句や「形容詞+名詞」のように連体修飾部を持つ名詞句は、単一名詞に置換

以下に変数化の例を示す。

(形態素解析結果)

$N1$ [【学校】【公共機関】]/に/行く/ $N2$ [【時刻】【時点】【時間】【教授】]

【】は一般名詞意味属性名(以下意味属性と表記する)、/は形態素の区切りを意味する。

(3) 意味属性の絞り込み

通常、名詞には多義が存在するため、文意に合う意味属性を付与する必要がある。そこで、人手による意味属性の絞り込みを行う。上記の形態素解析結果においては、 $N1$ 【公共機関】、 $N2$ 【時間】と人手により一意に決定する。

(4) 「内と外の関係」の付与

パターンの絞り込みを行うために「内と外の関係」を付与する。作成した日本語パターンを以下に示す。(日本語パターン 1)

$N1$ [【公共機関】]に行く $N2$ [【時間】](外の関係)

上記の手順により、CREST コーパス 460 文から日本語パターン (77 パターン) を抽出した。内訳は、「行く」: 22 パターン、「見る」18 パターン:「思う」: 19 パターン、「生まれる」: 6 パターン、「生じる」: 12 パターンとなった。作成した日本語パターンの一部を表 2 に示す。

表 2 において(内)は「内の関係」、(外)は「外の関係」を意味する。

表 2: 日本語パターンの一部

日本語パターン
(ZERO が) 行く $N1$ [【地域(人間活動)】](内)
$N1$ [【散歩】]に行く $N2$ [用言性【計画】](外)
$N1$ [【出版物】]を見る $N2$ [時詞【時刻】](外)
$N1$ [【問題】]を $V1$ と思う $N2$ [【人間】](内)
$N1$ [【アイデア】]が生まれる $N2$ [【理由】](外)

4.2 英語パターンの抽出

機械翻訳を行うために、日本語パターンに対応する英語パターンが必要となる。以下の手順で日本語埋め込み表現に対応した英語パターンを対訳から抽出する。

(1) 英語表現の抽出

CREST コーパスから英語の埋め込み表現を抽出する。以下に(例文 3)の対訳と抽出例を示す。対訳中の波線は埋め込み箇所を示し、抽出例中の下線は「底の名詞」を示す。

(対訳 3)Normally, it is the time when he goes to school.

(抽出例)the time when he goes to school

(2) 英語パターンの作成

4.2 節で抽出した日本語パターンに対応させて品詞を変数化し、英語パターンを作成する。(対訳 3)より作成した英語パターンを以下に示す。ここで、4.2 節の(日本語パターン 1)と(英語パターン 1)は互いに対となるパターンである。

(英語パターン 1) $N2$ REL.ADV($N2$) ZERO go to $N1$

本稿では、意識や慣用表現といった特殊な場合はパターン化が困難なため、対象外とする。

上記の手順により、CREST 例文 460 文の対訳から英語パターン (251 パターン) を抽出した。内訳は、「行く」: 126 パターン、「見る」: 50 パターン、「思う」: 18 パターン、「生まれる」: 23 パターン、「生じる」: 34 パターンとなった。作成した英語パターンの一部を表 3 に示す。

表 3: 英語パターンの一部

英語パターン
REL.ADV($N1$) to go $N1$
$N2$ for $N1$
REL.ADV($N2$) reading $N1$
$N2$ REL($N2$) want to $V1$ $N1$
ADV($N2$) $N1$ develop

表 4: 翻訳規則の一部

	日本語パターン 英語パターン	意味属性 (N1)	意味属性 (N2)
①	N1 へ行く N2 N2 to N1	【行政機関】【土地】【月 (天体)】【公共機関】【駅】 【行政区画】	【道路】【接辞 (人間/ 単数)】
②	N1 へ行く N2 N2 to go N1	【領土】	【決心】 [用言性]
③	N1 に行く N2 N2 REL(N2) be going to N1	【旅行・遊覧】 [用言性]	【人間】
④	(ZERO が) N1 に行く N2 CONJ(N2) ZERO go N1	【買い】 [用言性] 【領土】	【以前】 [時詞]
⑤	N1 を見る N2 N2 for N1	【像・書画】【商品】【観 (かん)】【道具】	【目 (器官 (本体))】
⑥	(ZERO が) N1 を見る N2 CONJ(N2) ZERO take a look at N1	【帳】	【常時】 [時詞]
⑦	N1 が生まれる N2 ADV(N2) N1 develop	【アイデア】	【理由】

4.3 翻訳規則の作成

抽出した日本語パターンと英語パターンから翻訳規則を作成する。作成したルール数は 251 個となった。翻訳規則の一部を表 4 に示す。表 4 において、各々の変数は表 1 で定義した変数である。

また、意味属性において [] 内に名詞の種類を付与し、用言性名詞や時詞等、5 種類に場合分けしている。省略の場合は、普通名詞を意味する。

作成した埋め込み文日英翻訳規則は、日本語パターンの用言の字面及び名詞の意味属性により、英語パターンを訳出する。

5 実験

作成した翻訳規則の精度を調べるため、規則製作者による人手の評価実験を行う。

5.1 評価実験

評価実験は、クローズドテスト及びオープンテストを行う。クローズドテストにおいては、翻訳規則の作成に用いた例文の中からランダムに抽出した 50 文 (「行く」: 25 文、「見る」: 8 文、「思う」: 7 文、「生まれる」: 5 文、「生じる」: 5 文) を対象とする。オープンテストにおいては、翻訳規則作成に使用されていないコーパス (旺文社: マルチ辞書) から、本稿で用いた用言を含む文をランダムに 100 文 (「行く」: 62 文、「見る」: 13 文、「思う」: 19 文、「生まれる」: 3 文、「生じる」: 3 文) を抽出して行う。

5.2 評価基準

評価は、翻訳規則製作者本人による 3 段階の評価を付ける。表 5 に評価基準の詳細を示す。

表 5: 評価基準

評価	評価基準
	英語パターンを一意に絞りがめ完全一致した場合
	複数候補中に正解となる英語パターンを含んだ場合
x	正解となる英語パターンを出力できなかった場合

5.3 実験手順

(1) 埋め込み箇所への入力

実験に使用する日本語例文から人手で埋め込み箇所を抽出し、入力データとする。ここで、格要素の言い換えを補うことを許す。

(日本語例文) 彼の助言で私は外国へ行く決心を固めた。
(入力データ) 外国へ (に) 行く決心

(2) 名詞の変数化

入力データを形態素解析し、名詞を変数化することにより、日本語パターンを作成する。

(作成した日本語パターン)

N1[【国家】【領土】]/へ/行く/N2 [用言性【決心】【事柄】]

(3) 英語パターンの訳出

作成した日本語パターンを翻訳規則に適応する。この例の場合、表 4 の②に適合する。

(訳出した英語パターン) N2 to go N1

(4) 対訳の英語パターンによる出力英語パターンの評価

(3) で訳出したパターンを対訳から作成した英語パターンと比較し、評価する。以下に対訳と対訳より抽出した英語パターンを示す。

(対訳) His advice confirmed my decision to go
abroad.

(対訳の埋め込み箇所) my decision to go abroad

(対訳より作成した英語パターン) N2 to go N1

この場合、対訳より作成した英語パターンと翻訳規則より訳出した英語パターンが一致するので正解となる。

5.4 実験結果

クローズドテストとオープンテストにおける評価結果をそれぞれ表 6・表 7 に示す。

表 6: 実験結果 (クローズドテスト)

	クローズドテスト (50 文)	割合 (%)
	34 文	68%
	16 文	32%
×	0 文	0%

表 7: 実験結果 (オープンテスト)

	オープンテスト (100 文)	割合 (%)
	4 文	4%
	53 文	53%
×	43 文	43%

机上実験の結果、クローズドテストにおいて 100%、オープンテストにおいて 57% の正解率が得られた。

6 考察

5 節の評価実験において評価が 及び × となった埋め込み文について考察する。複数候補パターンに関する考察及び誤った英語表現を訳出する原因を以下に示す。

6.1 文脈依存による問題

文脈により複数に訳し別け可能な埋め込み文においては、パターンを一意に決定することが困難である。以下に例を示す。

(例文 3) 彼は老人に駅へ行く道を地図で教えた。

(対訳 3) He showed the old man the way to the station with the map.

例文 3 において「～へ行く道」の英訳には、使われる状況により「how to get to city hall」「the way to the station」「on my way to school」「in my way to school」が存在する。この場合においては、意味属性で訳し分けるのは困難であると考えられる。

6.2 内と外の関係によるパターン絞り込み

複数の候補が出現する評価 について、予め付与した「内と外の関係」ラベルに基づきパターンの絞り込みを行なった。その結果を表 8 に示す。

表 8: 評価 におけるパターン絞り込み結果

	「内と外の関係」ラベル適用前	適用後
平均パターン数	3.5 パターン	2.7 パターン

表 8 に示すように、「内と外の関係」ラベルを付与することで、複数候補のパターンが出現した場合への効果が期待できる。

6.3 意味属性の不足による失敗

作成した翻訳規則は、コーパスから抽出して作成したため、名詞がコーパス中に存在せず意味属性が不足している場合、誤った英語パターンを訳出してしまう。以下に例を示す。

(例文 4) 雨のせいでピクニックに行く望みは絶たれた。

(対訳 4) The rain dashed our hopes for a picnic.

(日本語パターン)

N1【散歩】/に/行く/N2[転生名詞【期待】【予想】【希望】【遠望】]

(英語パターン (正解)) N2 for N1

例文 4 において、「N2 for N1」を訳出するための意味属性が作成した翻訳規則では不足しており、誤りとなる。この場合においては、意味属性を汎化することや、意味属性の追加により精度向上が見込めると考えられる。

6.4 意識による失敗

今回作成した翻訳規則では、意識の表現を規則として考慮していなかったため、翻訳規則に新たなルールを追加する必要がある。意識の例を以下に示す。

(例文 5) その村へ行く交通機関は何もない。

(対訳 5) No transportation is available to the village.

7 おわりに

本研究では、埋め込み文の翻訳規則を頻度の高い用言を基に作成した。人手による机上実験を行ったところ、翻訳精度はクローズドテストにおいて 100%、オープンテストにおいては、57% の正解率となった。オープンテストにおいて評価 × の例を見ると、意味属性不足で翻訳規則に適應できなかった場合が多く、意味属性の汎化次第で精度向上が見込める。

今後は、他の用言についての検証や、意味属性の汎化によるカバー率の向上が必要である。

参考文献

- [1] 竹内: 日本語複文の意味的構造の解析, 鳥取大学卒業論文, (1999)
- [2] 山田 孝雄: 日本文法論, 宝文館 (1898)
- [3] 寺村 秀夫: 日本語シンタクスと意味 ~ , くるしお出版 (1982~1991)
- [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 (Sep. 2001).
- [5] 藤本, 池原, 村上, 表: 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, 言語処理学会第 8 回年次大会, pp.679-682 (Mar. 2002).
- [6] 村上, 池原, 徳久: 日本語英語の文対応の対訳データベースの作成, 「言語・認識・表現」第 7 回年次研究会, (Dec. 2002)
- [7] 池原, 宮崎, 白井 他: 日本語語彙大系, 岩波書店 (1997)