

概要

機械翻訳において埋め込み文は、解析が困難であり翻訳精度を低下させる要因の一つとなっている。埋め込み文は、英語に翻訳する場合、関係節、同格節、前置詞句など多種多様な表現に訳し分けられる。機械翻訳においてそれら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。

本稿では、埋め込み箇所特定の用言（「行く」「見る」「思う」「生まれる」「生じる」）を持つ埋め込み文（CREST コーパスから 460 文抽出）を対象として、日本語側の構造及び名詞の意味属性に着目して分類を行った。分類した埋め込み文において日本語構造と英語構造の対応関係を翻訳規則にまとめた。

作成した翻訳規則の総数は、日本語パターンが 77 パターン、英語パターンが 251 パターンであった。日本語パターンの内訳としては、「行く」が 22 パターン、「見る」が 18 パターン、「思う」が 19 パターン、「生まれる」が 6 パターン、「生じる」が 12 パターンとなった。また、英語パターンの内訳としては、「行く」が 126 パターン、「見る」が 50 パターン、「思う」が 18 パターン、「生まれる」が 23 パターン、「生じる」が 34 パターンとなった。

さらに、作成した翻訳規則に対してクローズドテスト (50 文)、オープンテスト (100 文) による机上実験を行い、評価・検討した。その結果、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% の正解率を得た。

この結果より、埋め込み文の日本語構造と英語構造の対応関係をまとめた翻訳規則の有効性が示された。

目次

1	まえがき	1
2	埋め込み文の構造	3
2.1	研究対象	3
2.2	内と外の関係	4
3	従来研究及び本研究の目的	5
4	埋め込み文の日英パターン対応表	6
4.1	一般名詞意味属性体系	6
4.2	日英パターン対応表作成手順	7
4.3	日英パターンの作成	8
4.3.1	日本語例文からの日本語表現抽出	8
4.3.2	対訳からの英語表現抽出	9
4.3.3	一般名詞意味属性の付与	10
4.3.4	品詞の変数化	11
4.3.5	「内と外の関係」の付与	13
4.4	作成した日英パターンの数	15
4.5	翻訳規則の作成	16
5	評価実験	17
5.1	実験手順	17
5.2	評価実験	19
5.3	評価基準	19
5.4	実験結果	20
5.4.1	クローズドテストによる結果	20
5.4.2	オープンテストによる結果	21
6	正解と失敗の例	22
6.1	評価 の例	22
6.2	評価 の例	27
6.3	評価×の例	29

7	考察	31
7.1	文脈依存による問題	31
7.2	内と外の関係によるパターン絞り込み	32
7.3	意味属性の不足による失敗	34
7.4	意識による失敗	35
8	あとがき	36

目 次

1	「内の関係」と「外の関係」	4
2	一般名詞意味属性体系 (上位属性)	6

表目次

1	変数一覧	11
2	日英パターンの一部	15
3	翻訳規則の一部	16
4	格要素言い換え規則	17
5	評価基準	19
6	実験結果 (クローズドテスト)	20
7	パターン候補数調査 (クローズドテスト)	20
8	実験結果 (オープンテスト)	21
9	パターン候補数調査 (オープンテスト)	21
10	評価 におけるパターン絞り込み結果	32

1 まえがき

日英機械翻訳において、「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、解析が困難であり翻訳精度を低下させる要因の一つとなっている。埋め込み文は、英語に翻訳する場合、関係節、同格節、前置詞句など多種多様な表現に訳し分けられる。機械翻訳においてそれら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」の構造で考え、分類することで英語構造を関係節、同格節、前置詞句などへ訳し分ける効果があることが報告されている [1]。埋め込み文の分類方法としては、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] がある。何れの方法においても、計算機で判断することは困難となっていた。そこで、本研究室においても計算機による「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みに着目していたため、英語側との対応付けを行うまでには至っていない。そのため、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。研究対象には、埋め込み箇所特定の用言（「行く」「見る」「思う」「生まれる」「生じる」）を持つ埋め込み文 (CREST コーパスから 460 文抽出) を使用した。

作成した翻訳規則の総数は、日本語パターンが 77 パターン、英語パターンが 251 パターンであった。日本語パターンの内訳としては、「行く」が 22 パターン、「見る」が 18 パターン、「思う」が 19 パターン、「生まれる」が 6 パターン、「生じる」が 12 パターンとなった。また、英語パターンの内訳としては、「行く」が 126 パターン、「見る」が 50 パターン、「思う」が 18 パターン、「生まれる」が 23 パターン、「生じる」が 34 パターンとなった。

さらに、作成した翻訳規則に対してクローズドテスト (50 文)、オープンテスト (100 文) による机上実験を行い、評価・検討した。その結果、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% の正解率を得た。

以下、第 2 章では、埋め込み文の構造、第 3 章では、従来研究と本研究の目的、第 4 章

では，埋め込み文の日英パターン対応表，第5章では，評価実験，第6章では，正解と失敗の例，第7章では，考察，第7章では，今回の結論を述べる．

2 埋め込み文の構造

2.1 研究対象

本稿では、埋め込み文の翻訳規則を作成するにあたり、「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翻訳規則作成の対象とする。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。

以下に埋め込み文の例及び各々の例文における対訳を示す。

(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) 正しいと思う語の番号を丸で囲め。(内の関係)

(対訳 2) Enclose with a circle the number of the word you think right.

(例文 3) 普通は、学校に行く時間だ。(外の関係)

(対訳 3) Normally, the time when children go to school.

(例文 4) 幻覚は極度のストレスから生じる現象である。

(対訳 4) Hallucination is a phenomenon produced by extreme stress.

例文 1 では、修飾部(この窓から見る)と底の名詞(町の風景)によって構成される「この窓から見る町の風景」が埋め込み箇所になっており、埋め込み箇所の対訳は「the townscape from this window」である。

同様に例文 2 では、埋め込み箇所は「学校に行く時間」であり、埋め込み箇所の対訳は「the time when children go to school」である。

さらに、本稿において埋め込み文をより明確に分類するため、寺村の「内と外の関係」法を用いて分類を行うこととする。ここで、「内と外の関係」とは底の名詞と修飾部の間に格関係が有る場合を指し、「外と内との関係」とは底の名詞と修飾部の間に格関係を持たない場合を指す。「内と外との関係」の詳細については 2.2 節で説明する。

2.2 内と外の関係

「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、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、複文を「内の関係」と「外の関係」に分類した。以下、寺村の「内と外の関係」法 [3] について説明する。

寺村は、修飾部と底の名詞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」等に対応する。また、格関係を持たないものを「外の関係」と呼び、これは、英語の「同格節」等に対応する。

「内の関係」と「外の関係」の例を図1に示す。

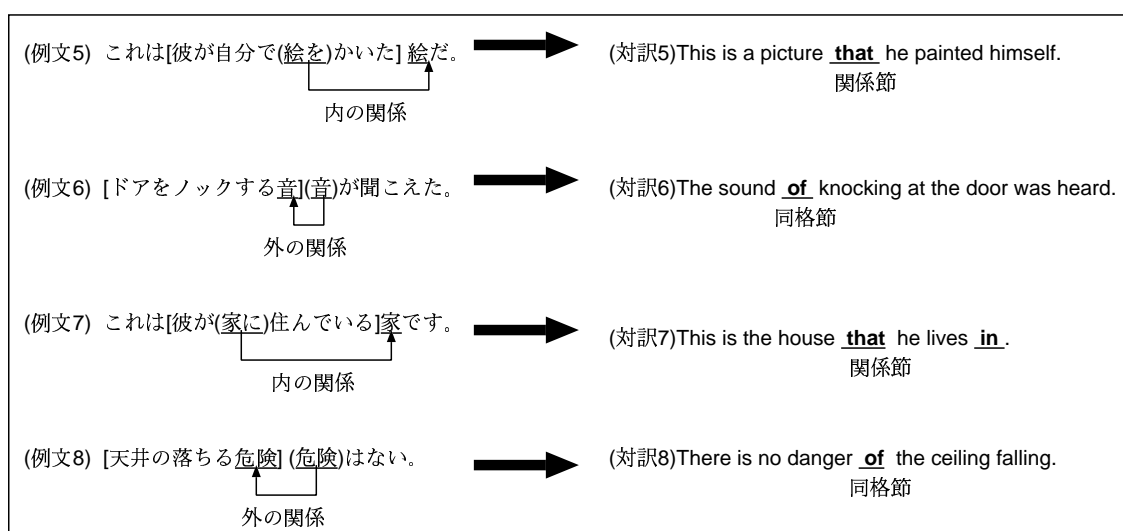


図 1: 「内の関係」と「外の関係」

例文5では、(修飾部)「彼が自分でかいた」と(底の名詞)「絵」の間に「彼が(絵を)かいた」という格関係『ヲ格』が存在するため、このような場合を「内の関係」と分類できる。

一方、例文6においては、(修飾部)「ドアをノックする」と(底の名詞)「音」の間に格関係は存在せず、(底の名詞)「音」の具体的な内容を(修飾部)「ドアをノックする」が説明する構造となっているため、このような場合を「外の関係」と分類できる。

本研究では、この「内の関係」と「外の関係」を埋め込み文の日英機械翻訳規則で使用する。

3 従来研究及び本研究の目的

埋め込み文に関する従来の研究としては、日本語埋め込み文がどのような英語表現に訳されているかを調べた研究 [1] がある。[1] では、埋め込み文を「内と外の関係」に分類し、修飾部と底の名詞の格関係に着目することで、「内関係」となる場合を 12 種類に下位分類した。また、「外関係」となる場合を底の名詞の抽象度と修飾部の陳述度に着目して、4 つの意味分類を行った。

しかし、[1] で扱う「内と外の関係」や「底の名詞の抽象度」、「修飾部の陳述度」を計算機で判断することが困難であるという問題があった。

これに対して、「内と外の関係」を計算機で判断する方法が提案された [4]。[4] では、埋め込み文の底の名詞の意味属性、修飾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た。判断規則の作成には、アンカー和英辞典より抽出した埋め込み文 200 文を使用した。作成された規則に対し、アンカー和英事典より抽出した 400 文を使用してオープンテストを行ったところ、「内関係」で 84.9%、「外関係」で 71.1% の正解率が得られ、その有効性が示されている。しかし、[4] は日本語側の構造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十分であった。

そこで本稿では、[4] において考慮されていなかった英語構造との対応関係を翻訳規則にまとめる。日本語構造と英語構造の対応付けには、日本語パターン・英語パターン及び名詞の意味属性を使用する。4 章では、日本語パターンと英語パターンの作成について述べる。

4 埋め込み文の日英パターン対応表

4.1 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、本研究では日本語語彙大系 [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 2 参照)

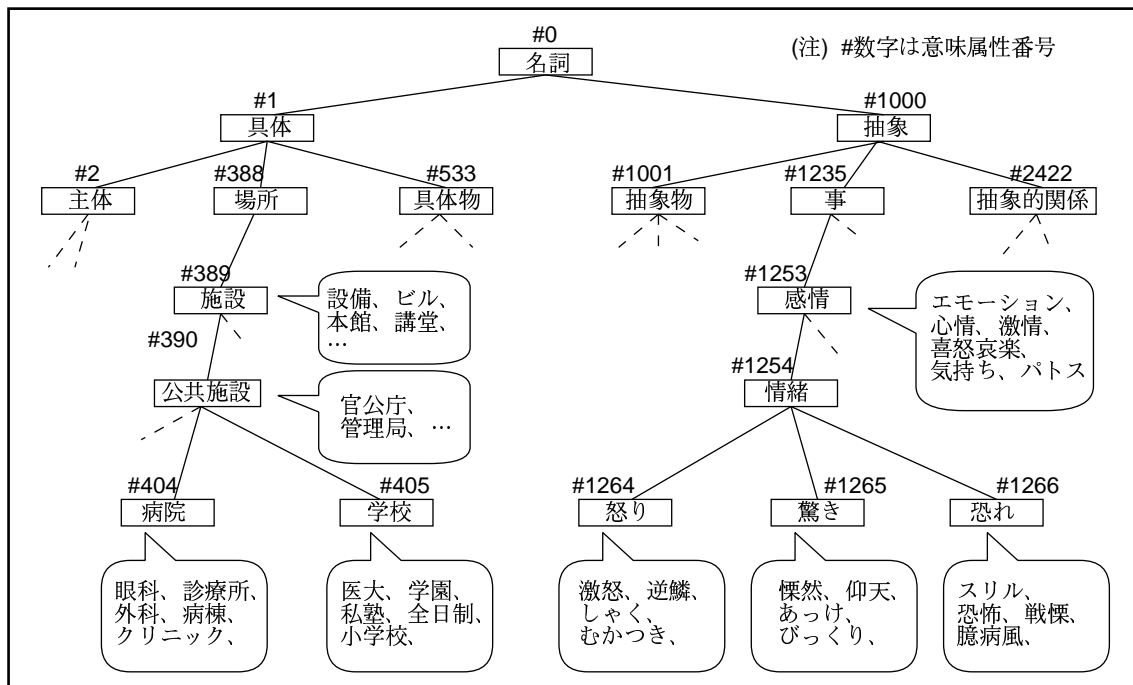


図 2: 一般名詞意味属性体系 (上位属性)

4.2 日英パターン対応表作成手順

本稿では，埋め込み文の日英機械翻訳を行うために，埋め込み文の翻訳規則を日英パターン対応表にまとめる．対応表には，日本語パターンと対になる英語パターンを CREST コーパスから各々人手で抽出して登録する．そして，登録した日本語パターンと英語パターンを一般名詞意味属性により制約をかける．一般名詞意味属性による制約により意味的にまとまりのあるパターンとして関係付けることができる．

対応表の作成手順は以下のとおりである．

1. 日本語例文及び対訳より埋め込み箇所を抽出する．
2. 日本語埋め込み箇所を形態素解析し，名詞及び形容詞を変数化する．変数化の際には，以下の 3 点に注意する．
 - (a) 時制は現在形に修正
 - (b) 日本語において副詞は除外
 - (c) 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 (例 1，例 2 参照)

(例 1) 私の本
(単一名詞置換後) 本

(例 2) 美しい景色
(単一名詞置換後) 景色
3. 名詞の意味属性を一意に決定し，日本語パターンを作成する．このとき，ラベルとして「内と外の関係」を付与する．
4. 対訳において，日本語に対応する品詞を変数化し英語パターンを作成する．

4.3 日英パターンの作成

4.3.1 日本語例文からの日本語表現抽出

対象とするコーパスには、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：約 12 万文) を使用する。今回、埋め込み箇所特定の用言を含む日英パターン対応表を作成する。対象とする用言には、CREST コーパスで使用される用言のうち「ある」「する」といった汎用性の高い用言を除き、頻度が上位の用言を対象とする。本研究では、用言「行く」「見る」「思う」「生じる」「生まれる」の 5 つを対象とする。

CREST コーパスの例文と人手により抽出した日本語表現の例を以下に示す。波線は埋め込み箇所、下線は底の名詞を意味する。

(例文 9) 普通は、学校に行く時間だ。

(抽出した日本語表現) 学校に行く 時間

(例文 10) 彼は私が生まれる前に死んだ。

(抽出した日本語表現) 私が生まれる 前

(例文 11) 広範にわたる飢饉が生じる危険がある。

(抽出した日本語表現) 飢饉が生じる 危険

(例文 12) 新幹線の旅は、美しい日本の田園風景を見る良いチャンスとなりました。

(抽出した日本語表現) 日本の田園風景を見る 良いチャンス

4.3.2 対訳からの英語表現抽出

CREST コーパスの対訳から埋め込み箇所の英語表現を抽出する。以下に CREST コーパスの対訳と対訳から抽出した英語表現の例を示す。ここで、波線は埋め込み箇所を示し、下線は底の名詞を意味する。

以下は、各々(例文 9)(例文 10)(例文 11)(例文 12)の対訳である。

(対訳 9)Normally, it is the time when he goes to school.

(抽出した英語表現)the time when he goes to school

(対訳 10)He died before I was born.

(抽出した英語表現)before I was born

(対訳 11)There is a danger of widespread famine.

(抽出した英語表現)a danger of widespread famine

(対訳 12)The trip by Shinkansen gave us a good chance to see the beautiful Japanese countryside.

(抽出した英語表現)a good chance to see the beautiful Japanese countryside

4.3.3 一般名詞意味属性の付与

4.3.1 節で抽出した日本語表現を形態素解析プログラムにかけ、品詞の解析及び一般名詞意味属性を付与する。以下に形態素解析結果の例を示す。ここで、下線は底の名詞、/は形態素の区切り【】は一般名詞意味属性を意味する。[]内の代名詞や時詞は名詞の種類を意味し、全部で5種類(普通名詞、用言性名詞、転生名詞、時詞、代名詞)に分類している。省略の場合は普通名詞を意味する。

(例3)

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

(例4)

私 [代名詞【私 し】【自称(単数)]]/が/生まれる/前[時詞【過去】【先(時間)】【以前】]

(例5)

飢饉 [【作況】【害】]/が/生じる/危険 [【危急】【危険】]

通常、上記の例のように名詞には多義が存在するため、複数の意味属性が付与される。翻訳規則を作成する上で、文意に合う意味属性を付与する必要があるため、以下の例のように人手による意味属性の絞り込みを行う。

(例7)

学校 [【公共機関】]/に/行く/時間[【時間】]

(例8)

私 [代名詞【自称(単数)]]/が/生まれる/前[時詞【以前】]

(例9)

飢饉 [【害】]/が/生じる/危険 [【危険】]

4.3.4 品詞の変数化

4.3.3 節の形態素解析結果から，品詞部分を変数化して日本語パターン及び英語パターンを作成する．ただし，特定の用言（「行く」「見る」「思う」「生まれる」「生じる」）は字面で残す．

本研究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

表 1 の変数に従って作成した日本語パターン及び英語パターンの例を以下に示す．

(日本語パターン 1) (*ZERO* が) *N1* [【公共機関】] へ行く *N2* [【時間】]

(英語パターン 1) *N2* *REL*(*N2*) (*ZERO*) go to *N1*

上記の(日本語パターン 1)と(英語パターン 1)は各々対となるパターンであり，日本語・英語において各々の変数(*N1*, *N2*, ...)が対応づけられている．また，*ZERO*は補完されたゼロ代名詞を意味し，日本語パターンには存在しない格要素である．英語パターンを作成した際にゼロ代名詞が存在する場合，日本語パターンには「(*ZERO*が)」と追加する．

(日本語パターン 2) $N1$ 【単数】で $N2$ 【代名詞【地域(人間活動)】】に(へ)行く $N3$ 【勇氣】]

(英語パターン 2) $AUX.V(N3)$ go $N2$ $N1$

(英語パターン 2) に存在する「 $AUX.V(N3)$ 」は、英訳する際に(日本語パターン 2) の「 $N3$ 」に対応する助動詞を当てはめることを意味する。

例えば、 $N3$ に意味属性【勇氣】を持つ名詞が当てはまった場合、「～する勇氣」を意味する助動詞「dare」が $AUX.V$ に対応する。

4.3.5 「内と外の関係」の付与

3.3.4 節で変数化したパターンに対して、日本語パターンに「内と外の関係」を付与する。「内と外の関係」は、複数候補パターンが出現した場合のパターン絞り込みに使用する。

「内と外の関係」付与後のパターン例を以下に示す。

(例文 13) あのお宮にはお参りに行く人が多い。

(対訳 13) There are lot of people who go to worship at that shrine.

(日本語パターン 3) $N1$ に行く $N2$ (内の関係)

(英語パターン 3) $N2$ REL($N2$) go to $N1$

(例文 14) 外国に行く前に予防接種を受けねばならない。

(対訳 14) You'll have to get some jabs before you go abroad.

(日本語パターン 4) $N1$ に行く $N2$ (外の関係)

(英語パターン 4) CONJ($N2$) ZERO go $N1$

(例文 15) テレビの深夜番組を見る人々にはすっかりおなじみの顔になった。

(対訳 15) She has become quite familiar to viewers of late-night programs on television.

(日本語パターン 5) $N1$ を見る $N2$ (内の関係)

(英語パターン 5) $N2$ of $N1$

(例文 16) 新聞は世界を見る窓である。

(対訳 16) Newspapers provide a window on the world.

(日本語パターン 6) $N1$ を見る $N2$ (外の関係)

(英語パターン 6) $N2$ on $N1$

上記の日本語パターン 3・日本語パターン 4 のように、日本語パターンが同一でも「内

の関係」「外の関係」により英語パターンを訳しわけることが可能となる。また同様に、日本語パターン5・日本語パターン6においても「内の関係」「外の関係」により英語パターンを訳しわけることが可能となる。

4.4 作成した日英パターンの数

4.3 節の日英パターン作成手順により, CREST コーパス 460 文から日本語パターン (77 パターン), 英語パターン (251 パターン) を抽出した.

内訳は, 日本語パターンが「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. また, 英語パターンは「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.

作成した日本語パターンと英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.

また, 日英パターン作成に使用した原文と作成した日本語パターン及び英語パターンを付録 1 に載せる.

表 2: 日英パターンの一部

日本語パターン	英語パターン
(ZERO が) 行く N1[【地域 (人間活動)】](内)	<i>REL.ADV(N1) to go N1</i>
N1[【散歩】]に行く N2[用言性【計画】](外)	<i>N2 for N1</i>
N1[【出版物】]を見る N2[時詞【時刻】](外)	<i>REL.ADV(N2) reading N1</i>
N1[【問題】]を V1 と思う N2[【人間】](内)	<i>N2 REL(N2) want to V1 N1</i>
N1[【アイデア】]が生まれる N2[【理由】](外)	<i>ADV(N2) N1 develop</i>
N1[【動物 (部分)]]に生じる N2[用言性【震動】](内)	<i>N2 REL(N2) be generated N1</i>
N1[【苦しみ】]に対して生じる N2[用言性【反応】](外)	<i>N2 to N1</i>

4.5 翻訳規則の作成

3.3 節で抽出した日本語パターン及び英語パターンから埋め込み文の翻訳規則を作成する。作成した翻訳規則の一部を表3に示す。意味属性において [] 内には名詞の種類により、ラベルとして「用言性名詞」「時詞」「代名詞」「転生名詞」「形式名詞」を付与している。省略の場合は、普通名詞を意味する。

また、用言ごとに作成した翻訳規則は付録2に載せてある。

表 3: 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 REL.ADV(N2) N1 develop	【アイデア】	【理由】
⑧	N1 が生まれる N2 be going to have N1	【子】	用言性【計画】

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、対応する英語パターンを訳出する。

翻訳規則は、CREST コーパスより対象とする用言(「行く」「見る」「思う」「生じる」「生まれる」)を含む文(460文)から作成した。

作成した総パターン数は、日本語パターン(77パターン)、英語パターン(251パターン)となった。

5 評価実験

4章で作成した埋め込み文日英翻訳規則の性能を調査するため、翻訳規則製作者による机上での評価実験を行う。評価実験は、翻訳規則により訳出した英語パターンと対訳から作成した英語パターンとを比較して行う。

5.1 実験手順

評価実験は以下の手順に沿って行う。

(1) 埋め込み箇所を入力

日本語例文から人手により埋め込み箇所を抽出し、入力データとする。入力データの例を以下に示す。

(例文 15) 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

ここで、同一の意味を持つ格要素に対して格要素の言い換え補うことを許す。格要素の言い換え規則を表 4 に示す。

表 4: 格要素言い換え規則

格要素	言い換え可能な格要素
ニ	へ
デ	ヨリ
へ	ニ
ノ	ガ

例えば、上記の入力データにおいては、「外国へ行く決心」と「外国に行く決心」の2通りを考慮する。

(2) 名詞の変数化

入力データを形態素解析プログラムにかけ、意味属性の付与及び品詞の変数化を施す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の例を以下に示す。

(日本語パターン 5)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

得られた日本語パターンを 3 章で作成した埋め込み文翻訳規則 (表 3 参照) へ適用し、対応する英語パターンを訳出する。

上記の日本語パターン 5 の場合、表 3 の②に適合し、以下の英語パターンを訳出する。

(英語パターン 5)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) で訳出した英語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。上記の例文 15 に対する対訳とその対訳より抽出した英語パターンを以下に示す。

(対訳 15)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、英語パターン 5 と対訳より作成した英語パターンが一致するので正解となる。

5.2 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により行う．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 文を対象とする．

オープンテストにおいては，翻訳規則の作成に使用していないコーパス (旺文社：マルチ辞書) から，本研究で用いた用言を含む文をランダムに 100 文抽出して行う．以下に試験文の例を示す．

(クローズドテスト用例文) 彼は外国に行くチャンスを逸した．

(クローズドテスト用対訳) He lost the opportunity to go abroad.

(オープンテスト用例文) こんどの日曜日静岡のおじさんに会いに行く予定です．

(オープンテスト用対訳) I am to visit my uncle in Shizuoka next Sunday.

5.3 評価基準

評価は，埋め込み文翻訳規則作成者本人による 3 段階評価で行う．表 5 に評価基準の詳細を示す．

表 5: 評価基準

評価の種類	評価基準
×	英語パターンを一意に絞りこめ完全一致した場合 候補が複数でた場合に正解となる英語パターンを含んだ場合 正解となる英語パターンを出力できなかった場合

5.4 実験結果

5.4.1 クローズドテストによる結果

CREST コーパスからランダムに 50 文抽出してクローズドテストを行った結果を表 6 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%
x	0 文	0%

クローズドテストの結果 , 正解率は 100% であり , そのうち 68% がパターンを一意に決定することができた .

また , 評価 の場合について最大パターン候補数及び平均パターン候補数を調査した . 表 7 に結果を示す .

表 7: パターン候補数調査 (クローズドテスト)

	評価 (16 文)
最大	4 パターン
平均	2.6 パターン

5.4.2 オープンテストによる結果

旺文社「マルチ辞書」よりランダムに 100 文抽出してオープンテストを行った結果を表 8 に示す。

表 8: 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%

オープンテストの結果、正解率は 57% であり、そのうち 4% がパターンを一意に決定することができた。

また、評価 の場合について最大パターン候補数及び平均パターン候補数を調査した。表 9 に結果を示す。

表 9: パターン候補数調査 (オープンテスト)

	評価 (53 文)
最大	6 パターン
平均	3.6 パターン

6 正解と失敗の例

6.1 評価 の例

この節では，クローズドテスト及びオープンテストの実験結果におけるそれぞれの評価の例及び実行の様子を示す．

(例文 16) 第一段階は，新商品のアイデアが通常生まれる理由と場所を理解することである．

(対訳 16) The first step is to understand why and where new-product ideas usually develop.

例文 16 の場合，入力データは「アイデアが生まれる理由」であり，入力データに形態素解析プログラムと品詞の変数化を施すと以下の日本語パターン 6 を得る．

(日本語パターン 6) $N1$ [【アイデア】] が (の) 生まれる $N2$ [【原因】【理由】【証拠】]

このパターンを作成した翻訳規則に適用すると，以下の英語パターンを訳出する．

(英語パターン 6) $REL.ADV(N2) N1$ develop

訳出された英語パターン 6 において， $ADV(N2)$ は日本語パターン 6 中の名詞 $N2$ [【理由】] に対応する副詞 (関係副詞) を意味する．

次に，訳出したパターン (英語パターン 6) と対訳 16 の英語パターンを比較するため，対訳 16 より英語パターンを作成する．作成したパターンは以下のとおりである．

(対訳 16 より作成した英語パターン) $REL.ADV(N2) N1$ develop

英語パターン 6 と対訳 16 より作成したパターンを比較すると同一パターンであるため，評価は となる．

(例文 17) 買い物に行く前に , クレジットカードを手に入れなさい . (クローズドテスト)

(対訳 17) Get your credit card before you go shopping .

同様に例文 17 の場合入力データは「買い物に行く前」であり , 入力データに形態素解析プログラムと品詞の変数化を施すと以下の日本語パターン 7 を得る .

(日本語パターン 7) $N1$ [用言性]【買い】【商品】/へ/行く/ $N2$ [時詞]【過去】【先(時間)】【以前】

この日本語パターン 7 を作成した翻訳規則に適用すると , 以下の英語パターンを得る .

(英語パターン 7) $CONJ(N2)$ ZERO go $N1$

訳出された英語パターン 7 において , $CONJ(N2)$ は日本語パターン 7 中の名詞 $N2$ [時詞【以前】] に対応する接続詞を意味する .

次に訳出した英語パターン (英語パターン 7) を評価するため , 対訳 17 より英語パターンを作成し比較する . 作成した英語パターンは以下のとおりである .

(対訳 17 より作成した英語パターン) $CONJ(N2)$ ZERO go $N1$

英語パターン 7 と対訳 17 より作成した英語パターンを比較すると同一パターンであるため , 評価は となる .

(例文 18) あの顔はテレビでときどき見る顔だ。(クローズドテスト)

(対訳 18) That's a face which appears on TV every now and then.

同様に例文 18 の場合入力データは「テレビで見る顔」であり、入力データに形態素解析プログラムと品詞の変数化を施すと以下の日本語パターン 8 を得る。

(日本語パターン 8) $N1$ [【通信機器】] / で / 見る / $N2$ [【顔つき】【観〔かん〕】【顔】【顔面】]

この日本語パターン 8 を作成した翻訳規則に適用すると、以下の英語パターンを得る。

(英語パターン 8) $N2$ $REL(N2)$ appear on $N1$

訳出された英語パターン 8 において、 $REL(N2)$ は日本語パターン 8 中の名詞 $REL(N2)$ [【】] に対応する接続詞を意味する。

(例文 19) 我々は行く先々で歓迎された (オープンテスト) .

(対訳 19) We were welcomed wherever we went.

次に, 例文 19 の場合入力データは「行く先々」であり, 入力データに形態素解析プログラムと品詞の変数化を施すと以下の日本語パターン 7 を得る .

(日本語パターン 9) 行く / N1 [時詞【未来】【方面】]

このパターンを作成した翻訳規則に適用すると, 以下の英語パターンを訳出する .

(英語パターン 9) *ADV(N1) ZERO go*

ここで, *ZERO* はゼロ代名詞の補完を意味する . ゼロ代名詞とは文中で省略された格要素である . 日本語パターン 7 では, 主語「～が」が省略されているため, 英語パターン 9 において変数 *ZERO* によりゼロ代名詞が補完されている .

次に, 訳出したパターン (英語パターン 7) と対訳 19 の英語パターンを比較するため, 対訳 19 より英語パターンを作成する . 作成したパターンは以下のとおりである .

(対訳 19 より作成した英語パターン) *REL.ADV(N1) (ZERO go)*

英語パターン 9 と対訳 19 より作成したパターンを比較すると同一パターンであるため, 評価は となる .

(例文 20) 彼の助言で私は外国へ行く決心を固めた。(オープンテスト)

(対訳 20) His advice confirmed my decision to go abroad.

次に、例文 20 の場合入力データは「外国へ(に)行く決心」であり、入力データに形態素解析プログラムと品詞の変数化を施すと以下の日本語パターン 10 を得る。

(日本語パターン 10) $N1$ [【国家】【領土】]/へ(に)/行く/ $N2$ [用言性【事柄】【決心】]

このパターンを作成した翻訳規則に適用すると、以下の英語パターンを訳出する。

(英語パターン 10) $N2$ to go $N1$

次に、訳出したパターン(英語パターン 10)と対訳 20 の英語パターンを比較するため、対訳 20 より英語パターンを作成する。作成したパターンは以下のとおりである。

(対訳 20 より作成した英語パターン) $N2$ to go $N1$

英語パターン 10 と対訳 20 より作成したパターンを比較すると同一パターンであるため、評価は となる。

6.2 評価 の例

(例文 18) それから何らかの益が生じる見込みは全くない。(クローズドテスト)

(対訳 18) There is no hope of good coming from that.

例文 18 の場合，入力データは「益が生じる見込み」であり，入力データに形態素解析プログラムと品詞の変数化を施すと，以下の日本語パターン 8 を得る．

(日本語パターン 8)

$N1$ [【利益】【恩】【恩恵】【利得】【効果】]/が(の)/生じる/ $N2$ [転生名詞【偽】【期待】【計算】【予想】【時期】]

日本語パターン 8 を翻訳規則に適用すると，以下の 2 つの英語パターン (英語パターン 8, 9) を訳出する．

(英語パターン 8) $N2$ of $N1$ arising

(英語パターン 9) $N2$ of $N1$ coming

上記の 2 つの英語パターンと対訳 18 の英語パターンを比較するために，対訳 18 より英語パターンを作成する．作成した英語パターンは以下のとおりである．

(対訳 18 より作成した英語パターン) $N2$ of $N1$ coming

従って，対訳 18 より作成した英語パターンと翻訳規則が訳出した英語パターン (英語パターン 9) が一致するので評価は となる．

(例文 19) 明日は大阪へ行く予定だ。(オープンテスト)

(対訳 19) We are going to Osaka tomorrow.

例文 19 の場合，入力データは「大阪へ行く予定」であり，入力データに形態素解析プログラムと品詞の変数化を施すと，以下の日本語パターン 9 を得る．

(日本語パターン 9)

$N1$ [【行政区画】【行政機関】【公共機関】【学校】【乗降場】]/へ(に)/行く/ $N2$ [用言性【日程】【準備】【決定】【計画】]

日本語パターン 8 を翻訳規則に適用すると，以下の 6 つの英語パターン (英語パターン 10, 11, 12, 13, 14, 15) を訳出する．

(英語パターン 10) be planning to go to $N1$

(英語パターン 11) will go to $N1$

(英語パターン 12) $N2$ for $N1$

(英語パターン 13) be going to $N1$

(英語パターン 14) be going to take $N1$

(英語パターン 15) be planning on going to $N1$

上記の 6 つの英語パターンと対訳 19 の英語パターンを比較するために，対訳 19 より英語パターンを作成する．作成した英語パターンは以下のとおりである．

(対訳 19 より作成した英語パターン) be going to $N1$

従って，対訳 19 より作成した英語パターンと翻訳規則が訳出した英語パターン (英語パターン 14) が一致するので評価は となる．

6.3 評価×の例

(例文 20) 私は買い物に行く途中です。(オープンテスト)

(対訳 20) I'm on my way to go shopping.

例文 20 の場合，入力データは「買い物に行く途中」であり，入力データに形態素解析プログラムと品詞の変数化を施すと，以下の日本語パターン 10 を得る．

(日本語パターン 10)

$N1$ [用言性【買い】【商品】]/に/行く/ $N2$ [時詞【最中】]

日本語パターン 10 を翻訳規則に適応すると，以下の 2 つの英語パターンを得る．

(英語パターン 16) on $N2$ to $N1$

(英語パターン 17) on *ZERO* $N2$ to $N1$

上記の 2 つの英語パターンと対訳 20 の英語パターンを比較するために，対訳 20 より英語パターンを作成する．作成した英語パターンは以下のとおりである．

(対訳 20 より作成した英語パターン) on *ZERO* $N2$ to go $N1$

従って，翻訳規則より訳出した英語パターン (英語パターン 16・英語パターン 17) は対訳より作成した英語パターンと一致しないので，評価は×となる．

(例文 21) これは私が今までに見た映画の中で最もよい。(オープンテスト)

(対訳 21) This is the best film I've ever seen.

例文 21 の場合，入力データは副詞(今まで)を省略した「私が見た映画」であり，入力データに形態素解析プログラムと品詞の変数化を施すと，以下の日本語パターン 11 を得る．

(日本語パターン 11)

$N1$ [代名詞【私〔し〕】【自称(単数)]]が見る $N2$ [【映画・映像】]

日本語パターン 11 を翻訳規則に適用すると，以下の 4 つの英語パターンを得る．

(英語パターン 18) in $PRON(N1)$ apprehension

(英語パターン 19) in $PRON(N1)$ book

(英語パターン 20) in $PRON(N1)$ opinion

(英語パターン 21) $PRON(N1)$ $N2$

上記の 4 つの英語パターンと対訳 20 の英語パターンを比較するために，対訳 20 より英語パターンを作成する．作成した英語パターンは以下のとおりである．

(対訳 21 より作成した英語パターン) $N2$ $PRON(N1)$ see

従って，翻訳規則より訳出した英語パターン(英語パターン 18・英語パターン 19・英語パターン 20・英語パターン 21)は対訳より作成した英語パターンと一致しないので，評価は×となる．

7 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．

7.1 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 22) 彼は老人に駅へ行く道を地図で教えた．

(対訳 22) He showed the old man the way to the station with the map.

(訳出された英語パターン)

(英語パターン 22) *ADV(N2) to get to N1*

(英語パターン 23) *N2 to N1*

(英語パターン 24) *on ZERO N2 to N1*

(英語パターン 25) *in ZERO N2 to N1*

例文 21 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．

7.2 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」に基づきパターンの絞り込みを行なった。以下にパターンの絞り込みが可能な例を示す。

(例文 23) ほら穴へ行く道は深い雪で閉ざされていた。(内の関係)

(対訳 23) The approach to the cave was blocked by deep snow.

例文 23 の埋め込み箇所の日本語パターンを以下に示す。

(日本語パターン 11)

$N1$ [【洞穴】【穴】]へ(に)行く $N2$ [【方法】【道路】【道徳】【人道】]

日本語パターン 11 を翻訳規則に適応させると以下の英語パターンを訳出する。

(英語パターン 26) *CONJ(N2) to get to N1*(外の関係)

(英語パターン 27) $N2$ to $N1$ (内の関係)

(英語パターン 28) $N2$ to go $N1$ (内の関係)

(英語パターン 29) on $N2$ to $N1$ (外の関係)

(英語パターン 30) on *ZERO* $N2$ to $N1$ (外の関係)

(英語パターン 31) in *ZERO* $N2$ to $N1$ (外の関係)

ここで、訳出した英語パターンにおいて「内と外の関係」に着目すると、(英語パターン 27) と (英語パターン 28) が「内の関係」であり、この場合 6 パターンから 2 パターンへのパターン絞り込みができる。

「内と外の関係」付与によるパターンの絞り込み結果を表 10 に示す。

表 10: 評価 におけるパターン絞り込み結果

	「内と外の関係」 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 10 に示すように、「内と外の関係」を付与することで、複数候補のパターンが出現した場合のパターンの絞り込み効果が期待できると考えられる。

7.3 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，英語パターンが訳出できず失敗となる．以下に例を示す．

(例文 24) 雨のせいでピクニックに行く望みは絶たれた．

(対訳 24) 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 24 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．

(例文 25) 我々は山へ行く計画を作り上げた。

(対訳 25) We worked out a plan to go to the mountains.

(日本語パターン) $N1$ [【山(本体)】【最中】]へ行く $N2$ [[用言性]【案】[用言性]【計画】]

(英語パターン (正解)) $N2$ to go to $N1$

上の例文では、本来「 $N2$ to go to $N1$ 」という英語パターンを出力すべきだが、日本語パターン中の名詞の属性が翻訳規則に当てはまらず誤りとなる．

いずれの場合においても，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．

7.4 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 26) その村へ行く交通機関は何もない。

(対訳 26) No transportation is available to the village.

例文 26 の対訳を見ると、「どの交通機関でもその村へ行く可能性がない」と意識されている。このように意識される文は本稿で作成した翻訳規則を使用しても解決することは困難である。

8 あとがき

本研究では、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、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した。研究対象として使用頻度の高い用言（「行く」「見る」「思う」「生まれる」）を埋め込み箇所を含む文を使用し、翻訳規則を作成した。翻訳規則は、CREST コーパスから 460 文を抽出して作成した。CREST コーパスから作成したパターン数は、日本語パターン (77 パターン)、英語パターン (251 パターン) であった。内訳は、日本語パターンが「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。また、英語パターンは「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。

作成した翻訳規則は、日本語パターン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応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。

謝辞

最後に，本研究において御指導頂きまし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，村上助教授，徳久助手に厚くお礼申し上げます．

また，本研究で alt-jaws を使用させて頂いたコミュニケーション科学基礎研究所の方々及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] 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 ~ , くらしお出版 (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,(Nov. 2002)
- [7] 池原, 宮崎, 白井 他:日本語語彙大系 , 岩波書店 (1997)