

重文・複文文型パターン辞書を用いた 日英パターン翻訳システムにおける離散記号問題の解決手法

水田理夫

Michio Mizuta

鳥取大学 工学部 知能情報工学科

Department of Information and Knowledge Engineering, Tottori University

1 はじめに

日英機械翻訳の分野において、重文・複文の日英機械翻訳のために文型パターン辞書が構築されている [1]。文型パターン辞書は原言語と目的言語をパターンで結ぶことにより、構造と意味の関係を対応づけたデータベースである。

従来の要素合成法を用いた日英翻訳方式では、統語構造の単位で一律に直訳が行われるため、その過程において意味が失われる問題がある。本稿で用いる日英パターン翻訳システムでは、表現構造と意味の関係に着目し、要素合成法により翻訳できる所を区別している。そのため、容易に良訳を得ることができると期待されている。

現在、文型パターン辞書 [1] に基づく日英パターン翻訳システム ITM [2] が試作されている。そして、文型パターン辞書に付与されている離散記号が、ITM の翻訳失敗の主要原因となっている。本稿では、この問題を解決するため、離散記号処理の実装に向けて実験を行う。

2 文型パターン辞書

文型パターン辞書 [1] は、言語表現に使用される対象認識を、重文・複文の形式によって網羅的に収集したものである。この方法により、日英言語族間における意味の対応付けを実現している。本辞書は、単語、句、節の3つのレベルで分類された約 22.7 万件の日英文型パターン対で構成されている。

2.1 文型パターン辞書の例

文型パターン辞書 [1] は、日本語文、対訳英文、日本語パターン、英語パターンによって構築されている。以下に、単語、句、節レベルの文型パターンの例を示す。

例

日本語文	信号が青になってから道路を渡りなさい
対訳英文	Cross the street after the traffic light turns green.
(単) 日パ	/ytcfkN1 が/tefkN2 に/cf なってから/ytckN3 を/cfV4.
(単) 英パ	V4 N3 after N1 turn N2.
(句) 日パ	/ytcfkN1 が VP2(てから でから)/ytckVP3.
(句) 英パ	VP3 after N1 VP2.
(節) 日パ	/ytcfkCL1(てから でから)/ytckCL2.
(節) 英パ	CL2 after CL1.

文型パターンには線形要素記号が付与されている。線形要素記号は、変数に対応する日本語表現を局所的に翻訳した結果を、英語パターンの対応箇所に挿入可能な記号である。

N , V , VP , CL , $/ytcfk$ は、それぞれ名詞変数、動詞変数、動詞句変数、節変数、離散記号である。これらは、線形要素記号である。

2.2 離散記号について

離散記号は、日本語表現に対する文型パターン辞書の被覆率向上のために、日本語パターンのみに付与されている記号である。同辞書を翻訳に用いた場合、離散記号に適合した日本語表現が英語表現に反映されない問題がある。本稿では、この問題の解決を目指して実験を行う。

まず、各離散記号が適合する日本語表現を表 1 に示す。

表 1: 各離散記号と適合する日本語要素

/y	連用節
/t	連体節
/c	格要素
/f	連用修飾要素
/k	連体修飾要素

次に、日本語パターンにおいて各離散記号が存在する位置と、日本語表現が各離散記号に適合した例を示す。太字部分が離散記号が適合した部分である。

/y: 文型, 連用・連体節の直前, 連用節の直後

絶えず騒音にさらされていると聴力が損なわれる。
彼が演説すると聴衆が耳を澄まして聞く。

/t: 名詞句の直前

「アンクル・トムの小屋」は奴隷廃止制度をうたった最も有名な文学作品であった。
選別されて残った応募者に面接をした。

/c: 格要素の直前・直後

血管の中を血がどくどくと流れていくのを感じた。
首相は近く渡米して大統領と会談する予定だ。

/f: 文頭, 述部・形容詞・動詞の直前

彼女は彼の求婚を拒絶したが、それは大して驚くにはあたらない。
立派に成し遂げられた仕事への満足感という真の喜びを味わう人はほとんどいない。

/k: 名詞句・単独名詞の直前

この病気の特徴は高熱が出ることだ。
彼女は和夫の突然の死を知って気絶寸前だった。

3 日英パターン翻訳システム

3.1 概要

日英パターン翻訳システムは、パターン検索プログラムと、同プログラムによって得たパターンを用いて翻訳を行うプログラム ITM に分かれている。

パターン検索プログラムにより、日本語文と文型パターン辞書を照合し、適合する日本語パターン、対となっている英語パターン、線形要素記号の対応情報を得る。[3]。

ITM は、パターン検索プログラムによって得た英語パターンと線形要素記号の対応情報を用いて、英文を訳出する。なお、線形要素記号の対応情報をバインド値と呼ぶ。

以下に処理の流れを示す。

1. 日本語文を入力する。
2. パターン検索プログラムを実行し、入力文に適合するパターンを全て得る。
3. 人手で、翻訳に使用する日英パターン対を選択する。
4. ITM が英語パターンとバインド値を用いて自動的に英文を訳出する。具体的には以下の2つの処理を行う [2]。

(a) 単語を翻訳し、複数の英単語を得る。

例えば、日本語文「彼女はりんごを持っている」、英語パターン「<She|N1> V3 (a|an) N2.」, 変数と単語の対応関係「N1=彼女, N2=りんご, V3=持つ」の場合、候補に [[she], [have, has, keep,...],[a, an],[apple,apples]] を得る。

(b) 文型パターン辞書 [1] に基づいた英語言語モデルを用いて、単語の連鎖確率を求める。そして、尤もらしい候補を選択する。

以下に、上例で得た候補 [[she], [have, has, keep,...],[a, an],[apple,apples]] を選択する様子を示す。

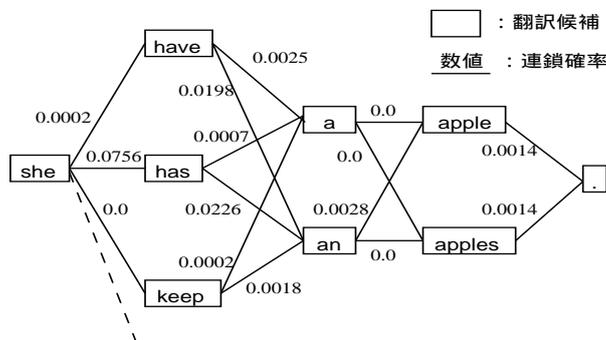


図 1: 単語の連鎖確率による候補選択の例

最も確率が高い「she has an apple .」が選択される。

この過程で得られた英文を訳出文とする。

3.2 動作例

日英パターン翻訳システムの動作例を以下に示す。

文入力 「人は年をとると大人になる。」を入力。

パターン検索 入力文と日本語パターンを照合し、適合した日本語パターンを全て得る。

1. 日パ /ytcfkN1 は/tcfk 年を/cf(取る | とる) と/tck(大人 | おとな) に/cf(成る | なる)。
 英パ As N1 grow older , N1^pron will grow in moderation .
2. 日パ /ytcfkN1 は/cfV2 と/tckN3 に/cf(成る | なる)。
 英パ N1 V2 to form N3 .
3. 日パ /ytcfkN1 を/cfV2 と/tckN3 に/cf なる。
 英パ It is N3 to V2^base N1 .

パターン選択 翻訳に用いる日英パターンを選択する。(1 を選択)

日パ /ytcfkN1 は/tcfk 年を/cf(取る | とる) と/tck(大人 | おとな) に/cf(成る | なる)。

英パ As N1 grow older , N1^pron will grow in moderation .

変数翻訳 「N1=人」を翻訳し、「you , person , man , ...」を得る。

候補選択 図 1 のように、単語の連鎖確率により候補を選択する。

英文訳出 As you grow older, it will grow in moderation.

3.3 評価実験

3.3.1 実験条件

実験に用いる入力文は、[1] の辞書からランダムに抽出した意識文 135 文とする。意識文とは、日本語文の節数より英文の節数が少ない文章を指す。以下に意識と直訳の例を示す。

例

意識文		
日本語文	欠点はあるがそれでも彼が好きだ。	2 節
英文	I like him in spite of his faults.	1 節
直訳文		
日本語文	瓶の栓が甘いので、中の物がこぼれそうです。	2 節
英文	The bottle top is a bit loose, so the contents might leak out.	2 節

実験はクロードテストとオープンテストの 2 種類を行う。なお、オープンテストはクロスバリデーションとする。クロードテストとは、入力文から作られた日英パターンを用いて、英文を訳出するテストである。クロスバリデーションテストとは、入力文とは別の文から作成された日英パターンを用いて、英文を訳出するテストである。

基本的な性能を評価するために、使用する文型パターン辞書は単語レベルとする。

3.3.2 評価方法

入力文の対訳英文を正解英文とし、訳出文を手動で評価する。評価基準は、日英文の意味的な対応関係の正否と、英文法上の正誤によって判断する。具体的には、以下の 4 段階とする。

評価基準

評価 4	動作主、動作、対象の意味的關係が正しく、統語的誤り*がない
評価 3	上記の意味的關係が正しいが、統語的誤り*がある
評価 2	上記の意味的關係が部分的に正しい
評価 1	上記の意味的關係が正しくない

* 冠詞と複数形については誤りとしない

また、評価値 3, 4 の訳出文を「翻訳成功」、評価値 1, 2 の訳出文を「翻訳失敗」、パターンに適合しなかった文を「適合無し」とする。以下に評価例を示す。

評価例

評価 4
 入力文 夕方になってやっと暑さが和らいだ。
 正解文 The heat finally let up in the evening.
 訳出文 Heat finally let up in evening.

評価 3
 入力文 うまくいったので意気揚々としている。
 正解文 He is flushed with success.
 訳出文 He be flushed with success.

評価 2
 入力文 とても優遇されて英文学の講座を担当した。
 正解文 The heat finally let up in the evening.
 訳出文 I covered chair of english literature .

評価 1
 入力文 彼は頭もよく勤勉でもある。
 正解文 He is as diligent as bright.
 訳出文 We and is <n5>.

3.3.3 結果

評価結果を表 2 に示す。

表 2: ITM 翻訳性能：離散記号処理を行わない場合

実験条件	翻訳成功 (評価 4, 3)	翻訳失敗 (評価 2, 1)	適合無し
ITM (closed test)	97% (131/135)	2% (3/135)	1% (1/135)
ITM (cross validation)	24% (33/135)	41% (55/135)	35% (47/135)

以下に、表 2 のクロスバリデーションテストにおける翻訳失敗文の例を示す。

翻訳失敗例

評価 2
 入力文 そのような行為は私たちの彼に対する尊敬を傷つけるものだ。
 正解文 Such conduct derogates from our respect for him.
 訳出文 Such behavior will hurt holy minds.
 原因 「私たちの彼に対する」が離散記号に適合している。

評価 1
 入力文 この市場を予測するには、かなり不確実なところがあった。
 正解文 There was considerable uncertainty in predicting the market.
 訳出文 She was the market .
 原因 不適切なパターンを選択している。

翻訳失敗文の多くは、日本語表現が離散記号に適合したものであった。表 3 に各離散記号の適合頻度を示す。

表 3: 離散記号の適合頻度

離散記号	/y	/t	/c	/f	/k
割合	54% (37/55)	14% (8/55)	27% (12/55)	11% (9/55)	20% (11/55)

なお、先行研究 [4] において、離散記号を理想的に処理できれば、翻訳成功割合が 40% まで向上することが確認されている。

4 離散記号処理

本稿では、離散記号に適合した日本語部分の翻訳を行い、離散記号に適合した部分の英文を ITM 訳出文に挿入する。この 2 つの工程を以下に示す。

1. 離散記号に適合した日本語部分の翻訳

離散記号に適合した日本語部分の翻訳には ALT-J/E を使用する。ALT-J/E に入力する際に、離散記号に適合した日本語部分を単文、または句に変換する。離散記号に適合した日本語部分を単文、または句に変換するため、語・節を追加する。変換のための規則は、表 1 と 2.2 節で示した離散記号の記述位置から決定する。

2. ITM 訳出文への離散記号に適合した日本語部分翻訳結果の挿入

ALT-J/E で翻訳した英文から、手作業で、離散記号に適合した日本語部分に該当する英文を ITM 訳出文に挿入する。翻訳した英文を ITM 訳出文に挿入する作業における規則を、表 1 と 2.2 節で示した離散記号が適合する日本語表現と離散記号の記述位置から決定する。

次節において、ALT-J/E の入力文、及び、翻訳した英文を ITM 訳出文に挿入する規則を示す。

4.1 /y について

/y は入力文の連用節に適合する離散記号である。また、日本語パターンにおいて /y は、文型の直前、連用・連体節の直前、連用節の直後に記述されている。特に、/y は文型・連用節の直前に存在する連用節に適合する場合が多い。

4.1.1 /y に適合した日本語部分の翻訳規則

表 4 に、/y に適合した日本語部分を、単文に変換する規則を示す。

表 4: /y : ALT-J/E の入力文を作成する規則

適合部分の語尾	ALT-J/E の入力文
動詞終止形+との	離散記号に適合した日本語部分 + 「ことだ。」
動詞+ための	" + 「ことだ。」
動詞+ように	" + 「する。」
動詞+のは	" + 「重要だ。」
名詞+を	" + 「知る。」
'デフォルト規則'	" + 「私が本を読む。」

以下に例を示す。太字部分が /y が入力文に適合した部分である。

例

< 動詞終止形+との >

日本語文： 彼はその計画を援助するとの力強い言葉をもらった。

/y に適合した部分+(語)： 彼はその計画を援助するとの(ことだ。)

< デフォルト規則 >

日本語文： 絶えず騒音にさらされていると聴力が損なわれる。

/y に適合した部分+(節)： 絶えず騒音にさらされていると(私が本を読む。)

4.1.2 /y に適合した部分の英文の挿入規則

/y 適合部分の翻訳結果の挿入規則 /y 適合部分の翻訳結果を、ITM 訳出文の文頭、または、文末に、追加した語・節の翻訳結果を削除して、手作業で挿入する。

以下に例を示す。

例

入力文：	絶えず騒音にさらされていると聴力が損なわれる。
ITM 訳出文：	Hearing was injured .
/y に適合した部分+(節)：	絶えず騒音にさらされていると (私が本を読む。)
ALT-J/E 翻訳結果：	(I read a book) when it is constantly exposed to noise.
ITM 訳出文に挿入：	Hearing was injured when it is constantly exposed to noise.

4.1.3 /y に適合した部分の翻訳と挿入の流れ

以下に、表 4 と上記の挿入規則を用いた処理の流れの例を示す。なお、太字部分が /y が入力文に適合した部分である。

例

< 日英パターン翻訳システム >	
1: 入力文：	絶えず騒音にさらされていると聴力が損なわれる。
2: 日本語パターンを検索 (パターン検索プログラム)：	(1) /ytcfkN1 が/cfV2。 (2) /yfn1 が/cfV2。
3.1: 日本語パターンを手動選択 (2 を選択)：	/yfn1 が/cfV2。
3.2: 対応する英語パターン：	N1 be V2 .
3.3: バインド値：	「聴力 = N1」, 「損なわ = V2」
4: ITM により変数翻訳	「聴力 hearing」, 「損なわ injure」
5: ITM 訳出文：	Hearing was injured .
< 離散記号処理 >	
6: /y に適合した部分+(節)：	絶えず騒音にさらされていると (私が本を読む。)
7: ALT-J/E 翻訳結果：	(I read a book) when it is constantly exposed to noise.
8: ITM 訳出文に挿入：	Hearing was injured when it is constantly exposed to noise.

4.2 /t について

日本語パターンにおいて /t は、名詞句の直前に記述されている。離散記号が付与されている変数は体言、または、用言である。表 3 に示した適合件数においては名詞に付与されている場合が多い。よって、追加する語には離散記号が付与されている変数の日本語表現を用いる。規則を以下に示す。

4.2.1 /t に適合した日本語部分の翻訳規則

/t: ALT-J/E の入力文を作成する規則 /t に適合した部分 + '付与変数の日本語'

以下に例を示す。

例

日本語文：	『アンクル・トム的小屋』は奴隷廃止をうたった最も有名な文学作品であった。
/t に適合した部分+(語)：	奴隷廃止制度をうたった (文学作品。)

4.2.2 /t に適合した部分の英文の挿入規則

/t 適合部分の翻訳結果の挿入規則 /t 適合部分の翻訳結果を、/t が付与されている変数の位置と追加した語の位置を合わせ、挿入する。

4.2.3 /t に適合した部分の翻訳と挿入の流れ

以下に大まかな例を示す．具体的な流れは /y における例と同様である．なお，下線は付与変数の対応関係を示している．

例

日本語文：	『アンクル・トム的小屋』は奴隷廃止をうたった最も有名な文学作品であった。
/t に適合した部分+(語)：	奴隷廃止制度をうたった (文学作品。)
ALT-J/E 翻訳結果：	The <u>literary work</u> as which an abolition slave institution was extolled.
ITM 訳出文：	Hut of <u>Uncle tom</u> was most well-known <u>literary work</u> .
挿入結果：	Hut of Uncle tom was most well-known <u>literary work</u> as which an abolition slave institution was extolled.

4.3 /c について

日本語パターンにおいて /c は，格要素の直前・直後に記述されている．直前に記述されている場合に適合した場合，離散記号が付与されている変数は名詞である．直後に記述されている場合に適合した場合，離散記号が付与されている変数は用言である割合が高い．よって，どちらの場合においても，追加する語には離散記号が付与されている変数の日本語表現を用いる．規則を以下に示す．

4.3.1 /c に適合した日本語部分の翻訳規則

/c：ALT-J/E の入力文を作成する規則 /c に適合した部分 + '付与変数の日本語'

以下に例を示す．

例

日本語文：	首相は近く渡米して大統領と会談する予定だ。
/c に適合した部分+(語)：	大統領と (会談する。)

4.3.2 /c に適合した部分の英文の挿入規則

/c 適合部分の翻訳結果の挿入規則 /c 適合部分の翻訳結果を，ITM 訳出文の文末に挿入

4.3.3 /c に適合した部分の翻訳と挿入の流れ

以下に大まかな例を示す．具体的な流れは /y における例と同様である．

例

日本語文：	首相は近く渡米して大統領と会談する予定だ。
/c に適合した部分+(語)：	大統領と (会談する。)
ALT-J/E 翻訳結果：	(It confers) with presidents .
ITM 訳出文：	He intended in the art of conversation .
挿入結果：	He is intended in the art of conversation with presidents .

4.4 /f について

日本語パターンにおいて /f は，文頭，述部・形容詞・動詞の直前に記述されている．文頭に付与されている場合に適合した例は無く，主に動詞の直前に付与されている場合に適合した．よって，追加する語には付与されている変数の日本語表現を用いる．規則を以下に示す．

4.4.1 /f に適合した日本語部分の翻訳規則

/f: ALT-J/E の入力文を作成する規則 /f に適合した部分 + '付与変数の日本語'

以下に例を示す。

例

日本語文： 自分の行動の罪を免れようと口先だけの調子のよい発言をいくつか行なった。
/f に適合した部分+(語)： いくつか (行う。)

4.4.2 /f に適合した部分の英文の挿入規則

/f 適合部分の翻訳結果の挿入規則 /f 適合部分の翻訳結果を、/f が付与されている変数の位置と追加した語の位置を合わせ、挿入する。

4.4.3 /f に適合した部分の翻訳と挿入の流れ

以下に大まかな例を示す。具体的な流れは /y における例と同様である。

例

日本語文： 自分の行動の罪を免れようと口先だけの調子のよい発言をいくつか行なった。
/f に適合した部分+(語)： いくつか (行う。)
ALT-J/E 翻訳結果： (It is held) variously.
ITM 訳出文： He did remarks to get rid of guilt.
挿入結果： He did **variously** remarks to get rid of guilt.

4.5 /k について

日本語パターンにおいて /k は、名詞句・単独名詞の直前に記述されている。名詞句の直前に付与されている場合に適合した例は無く、主に単独名詞の直前に付与されている場合に適合した。よって、追加する語には付与されている変数の日本語表現を用いる。規則を以下に示す。

4.5.1 /k に適合した日本語部分の翻訳規則

/k: ALT-J/E の入力文を作成する規則 /k に適合した部分 + '付与変数の日本語'

以下に例を示す。

例

日本語文： 自分の行動の罪を免れようと口先だけの調子のよい発言をいくつか行なった。
/k に適合した部分+(語)： 自分の行動の (罪。)

4.5.2 /k に適合した部分の英文の挿入規則

/k 適合部分の翻訳結果の挿入規則 /k 適合部分の翻訳結果を、/k が付与されている変数の位置と追加した語の位置を合わせ、挿入する。

4.5.3 /k に適合した部分の翻訳と挿入の流れ

以下に大まかな例を示す．具体的な流れは /y における例と同様である．

例

日本語文：	自分の行動の罪を免れようと口先だけの調子のよい発言をいくつか行なった。
/k に適合した部分+(語)：	自分の行動の(罪。)
ALT-J/E 翻訳結果：	(The <u>crime</u>) of its own action.
ITM 訳出文：	He did remarks to get rid of <u>guilt</u> .
挿入結果：	He did remarks to get rid of <u>guilt of its own action</u> .

4.6 実験

4.6.1 (1) 実験条件

ALT-J/E により，離散記号に適合した日本語部分の翻訳を行った．対象は，表 2 に示した翻訳失敗の 55 文において離散記号に適合した日本語部分である．

(2) ALT-J/E による離散記号に適合した日本語部分の翻訳結果

表 5 に，表 2 の翻訳失敗文 55 文の，各離散記号に適合した日本語部分を ALT-J/E で翻訳した結果を示す．

表 5: ALT-J/E による離散記号に適合した日本語部分の翻訳結果

離散記号	翻訳成功	翻訳失敗
/y に適合した部分	59% (22/37)	41% (15/37)
/t に適合した部分	62% (5/8)	37% (3/8)
/c に適合した部分	83% (10/12)	16% (2/12)
/f に適合した部分	33% (3/9)	66% (6/9)
/k に適合した部分	91% (10/11)	9%(1/11)

4.6.2 (1) 実験条件

ITM 訳出文への挿入処理の対象は，表 2 に示した翻訳失敗の 55 文において離散記号に適合した部分を ALT-J/E で翻訳した英文である．

(2) ITM 訳出文に挿入後の結果

離散記号処理により，表 2 における翻訳失敗文の 18 文が翻訳成功文となった．結果を表 6 に示す．

表 6: ITM 翻訳性能：離散記号処理を行った場合

実験条件	翻訳成功 (評価 4, 3)	翻訳失敗 (評価 2, 1)	適合無し
ITM (cross validation)	37% (51/135)	27% (37/135)	35% (47/135)

翻訳失敗文は，大きく以下の 2 種類に分けることができた．

1. ALT-J/E の翻訳文が誤っていた文
2. ITM の翻訳においてパターンの選択が不適切だった文

後者の原因は、入力文が膨大な数のパターンに適合した場合に、人手で最適なパターンを選択することは困難であるためである。

以下に、離散記号処理による翻訳成功文、及び、翻訳失敗文を示す。太字部分が離散記号に適合した日本語部分である。

(3)

例： 翻訳成功

/y

< 日英パターン翻訳システム >

1： 入力文：

常にトップを守るというのは大変なことだ。

2： 日本語パターンを検索 (パターン検索プログラム)：

(1) /ytck トップを/cf 守る (という | と言う) のは/cfAJV1 ことだ。

(2) /yfAJV1 ことだ。

3.1： 日本語パターンを手動選択 (2 を選択)：

/yfAJV1 ことだ。

3.2： 対応する英語パターン：

It is AJ1 .

3.3： バインド値：

「大変な = AJV1」

4： ITM により変数翻訳

「大変な great」

5： ITM 訳出文：

It is great .

6： 正解英文：

It is really hard to stay on top all the time.

< 離散記号処理 >

7： /y に適合した部分+(語)：

常にトップを守るというのは (重要だ。)

8： ALT-J/E 翻訳結果：

(It is important) **to always protect the top.**

9： ITM 訳出文に挿入：

It is great **to always protect the top** .

/t

< 日英パターン翻訳システム >

1： 入力文：

『アンクル・トム的小屋』は奴隷廃止をうたった最も有名な文学作品であった。

2： 日本語パターンを検索 (パターン検索プログラム)：

(1) /ytcfkN1 は/tcfkADV2/fAJV3/fN4 あった。

(2) /ytcfkN1/ytfkAJV3N4。

3.1： 日本語パターンを手動選択 (1 を選択)：

/ytcfkN1 は/tcfkADV2/fAJV3/fN4 あった。

3.2： 対応する英語パターン：

N1 be ADV2 AJ3 N4 .

3.3： バインド値：

「アンクル・トム的小屋 = N1」, 「最も = AVD2」, 「有名な = AJV3」, 「文学作品 = N4」

4： ITM により変数翻訳

「アンクル・トム的小屋 Hut of Uncle tom」, 「最も = most」, 「有名な = well-known」, 「文学作品 = literary work」

5： ITM 訳出文：

Hut of Uncle tom was most well-known literary work .

6： 正解英文：

Uncle Tom's Cabin was the most famous piece of antislavery literature.

< 離散記号処理 >

7： /t に適合した部分+(語)：

奴隷廃止制度をうたった (文学作品。)

8： ALT-J/E 翻訳結果：

(The literary work) **as which an abolition slave institution was extolled.**

9： ITM 訳出文に挿入：

Hut of Uncle tom was most well-known literary work **as which an abolition slave institution was extolled.**

/c, /k

< 日英パターン翻訳システム >

1: 入力文:

彼女の動機を適切な脈絡の中で理解するように心がけなさい。

2: 日本語パターンを検索 (パターン検索プログラム):

(1) /ytckN1 を/cfV2/yfV3。

(2) /ytcfkN1 の/kN2 を/cfAJV3/fN4 の/k(中 | なか) で/cfV5/yf(心がけ | 心掛け | 心懸け) なさい。

3.1: 日本語パターンを手動選択 (1 を選択):

/ytckN1 を/cfV2/yfV3。

3.2: 対応する英語パターン:

V3 to V2 AJ1 .

3.3: バインド値:

「動機 = N1」, 「理解する = V2」, 「心がけ = V3」

4: ITM により変数翻訳

「動機 motive」, 「理解する get」, 「心がけ be careful」

5: ITM 訳出文:

Be careful to get motive .

6: 正解英文:

Try to understand her motives in their proper context.

< 離散記号処理 >

7: /y に適合した部分+(語):

/c: 適切な脈絡の中で (理解する。)

/k: 彼女の (動機。)

8: ALT-J/E 翻訳結果:

/c: It is understood **in an appropriate logical connection**.

/k: **Her** motive.

9: ITM 訳出文に挿入:

Be careful to get **her** motive **in an appropriate logical connection**.

(4)

例: 翻訳失敗

/y

入力文:

両親から旅行に行っていていいという承認をもらった。

正解文:

I got an O.K. from my parents on my travel plans.

ITM 訳出文:

He has received acknowledgment .

/y に適合した部分+(節):

両親から旅行に行っていていいと (私が本を読む。)

ALT-J/E 翻訳結果:

(I read a book) **to receive from parents on a trip and to be good.**

離散記号処理後:

He has received acknowledgment **to receive from parents on a trip and to be good.**

原因

ALT-J/E が /y に適合した部分の翻訳に失敗

/y

入力文:

とても優遇されて英文学の講座を担当した。

正解文:

He filled the chair of English literature with great distinction.

ITM 訳出文:

I covered chair of english literature .

/y に適合した部分+(節):

とても優遇されて (私が本を読む。)

ALT-J/E 翻訳結果:

Very being treated well (I read a book) .

離散記号処理後:

Very being treated well I covered chair of english literature .

原因

ALT-J/E が /y に適合した部分の翻訳に失敗

/y

入力文:

君は疲れたのだから二三日休めば治る。

正解文:

You are only tired. A few days' rest will set you up.

ITM 訳出文:

If you will rest in , I will cure you .

/y に適合した部分+(節):

君は疲れたのだから (私が本を読む。)

ALT-J/E 翻訳結果:

Because you got tired, (I read a book).

離散記号処理後:

Because you got tired, if you will rest in , I will cure you .

原因

不適切なパターン選択が原因

/t	
入力文：	それはあまりにはっきりと立証されたので否認する余地がなかった。
正解文：	It was too clearly proved to admit of denial.
ITM 訳出文：	I wanted his room .
/t に適合した部分+(語)：	否認する (余地。)
ALT-J/E 翻訳結果：	(The room) denied .
離散記号処理後：	I wanted his room denied .
原因	ALT-J/E が /t に適合した部分の翻訳に失敗 不適切なパターン選択が原因

/c	
入力文：	そのような行為は私たちの彼に対する尊敬を傷つけるものだ。
正解文：	Such conduct derogates from our respect for him.
ITM 訳出文：	Such behavior will hurt holy minds.
/c に適合した部分+(語)：	私たちの彼に対する (尊敬。)
ALT-J/E 翻訳結果：	(Respect) for our him.
離散記号処理後：	Such behavior will hurt holy minds for our him .
原因	ALT-J/E が /c に適合した部分の翻訳に失敗

/f	
入力文：	すっかり遅れたが、今こそ事業の完成めざして改めて取り組むべきときだ。
正解文：	After all this delay it's time to make a new attack on the project.
ITM 訳出文：	This is time for all work to grapple .
/c に適合した部分+(語)：	完成めざして改めて (取り組む)
ALT-J/E 翻訳結果：	The completion look て (is engaged) anew .
離散記号処理後：	This is time for all work to the completion look て grapple anew .
原因	ALT-J/E が /c に適合した部分の翻訳に失敗 不適切なパターン選択が原因

4.6.3 考察

表 6 より、離散記号処理は、不完全ながらも良い結果と判断している。表 6 における翻訳失敗文の多くは、最適なパターンを選択できていないことが原因である。文型パターン辞書 [1] の単語・句・節レベルパターンを併用すれば、入力文の 8 割に対して、意味的に適切なパターンが使用できることが知られている。よって解決方法として、句、節レベルパターンを用いた翻訳が考えられる。また、多変量解析による最適パターンの選択といった手法が考えられる。

今後、ALT-J/E の入力文に対する規則の変更と追加によって、翻訳失敗文を翻訳成功文に導くのは困難であると考えられる。しかし、ALT-J/E における翻訳失敗文において前置詞や接続詞が誤っている場合には、解決の見込みがある。解決方法として、文型パターン辞書における日本語表現と英語表現の共起要素を翻訳知識として利用する方法が考えられる。

5 今後の課題

離散記号処理による ITM 訳出文の品質向上の可能性が確認できた。今後の課題を以下に示す。

- 離散記号処理の全自動化

現在、離散記号処理には手動の部分が含まれている。離散記号処理の全自動化のため、以下を課題とする。

1. ALT-J/E への入力文作成のため、各離散記号に適合した日本語部分に語・節を追加するプログラムを作成する。
2. ALT-J/E の翻訳結果を ITM 訳出文に挿入するプログラムを作成する。

- 文型パターン辞書を用いた離散記号処理

/y 適合部分を翻訳し、ITM 訳出文に挿入した後、ALT-J/E によって出力された接続詞に誤りがある場合に、以下の方法を検討している。

1. ALT-J/E が出力した接続詞を削除する .
2. 文型パターン辞書 [1] における日本語表現と英語表現の共起要素を翻訳知識として用い , 正しい接続詞を追加する .

共起要素の具体例を以下に示す . なお , 太字部分が共起要素 , CL は節を示している .

例 : 日本語表現と英語表現の共起要素

日本語文 :	わたしの書斎は南向きなので日がよく当たる
英文 :	Because my study faces south, it gets plenty of sunshine.
(節) 日本語パターン :	<i>CL1</i> ので <i>CL3</i> .
(節) 英語パターン :	Because <i>CL1</i> , <i>CL3</i> .

/y 適合部分の語尾別の翻訳知識の例を以下に示す .

例 : /y 適合部分の語尾別の翻訳知識の一部

日本語表現	英語表現
<i>CL1</i> て <i>CL2</i> 。	<i>CL2</i> when <i>CL1</i> .
<i>CL1</i> 前に <i>CL2</i> 。	<i>CL2</i> before <i>CL1</i> .
<i>CL1</i> と <i>CL2</i> 。	<i>CL2</i> when <i>CL1</i> . As soon as <i>CL1</i> , <i>CL2</i> .
<i>CL1</i> であれば <i>CL2</i> 。	If <i>CL1</i> , <i>CL2</i> .
...	...

例 : 翻訳知識の適用

/y	
入力文 :	風邪を引いてはなが出て困る。
正解文 :	I have caught cold, and my nose is always running.
/y に適合した部分+(節) :	はなが出て (私は本を読む)
ALT-J/E 翻訳結果 :	My nose flows and (I read a book).
ITM 訳出文 :	He has his cold and have a hard time .
離散記号処理後 :	My nose flows and he has his cold and have a hard time .
ALT-J/E 翻訳結果から接続詞を削除 :	My nose flows
/y の翻訳知識を選択 (手動) :	「 <i>CL1</i> て <i>CL2</i> 。」 「 <i>CL2</i> when <i>CL1</i> . 」
翻訳知識を用いた結果 :	He has his cold and have a hard time <u>when</u> my nose flows .

● ITM に使用するパターンの選択

現在 , ITM で使用する文型パターンは入力文と日本語文型パターンの照合の後に , 手動で選択している . この自動化のため , 多変量解析の手法が検討されている [5, 6] . この手法によれば , 単語レベルパターンにおいて 83% の割合で最適なパターン選択が可能である .

参考文献

- [1] 池原, 阿部, 徳久, 村上: 非線形な表現構造に着目した重文と複文の日英文型パターン化, 自然言語処理, 11(3), pp.149-164, 2004.
- [2] 石上, 水田, 徳久, 村上, 池原: 関数・記号付き文型パターンを用いた機械翻訳の試作と評価, 言語処理学会第 13 回年次大会発表論文集, pp.67-70, 2007.
- [3] 徳久, 村上, 池原: 重文・複文文型パターン辞書からの構造照合型パターン検索, 情報処理学会研究報告, 自然言語処理, 2006-NL-176, pp.9-16, 2006
- [4] 水田, 徳久, 村上, 池原: 重文・複文文型パターン辞書による意識の可能性, 電子情報通信学会ソサイエティ大会発表, 基礎・境界, 言語の意味と思考過程, AS-5-4, 2007.
- [5] 岡田, 村上, 徳久, 池原: 多変量解析による最適文型パターンの選択方式, 言語処理学会第 11 回年次大会発表論文集, pp.25-28, 2005.
- [6] 原, 村上, 徳久, 池原: 日英機械翻訳における多変量解析を用いた最適パターンの選択, 言語処理学会第 12 回年次大会発表論文集, pp.268-271, 2006.