

概要

言語表現の構造を，意味を持つ単位にパターン化した文型パターンを用いた，文型パターン翻訳方式が提案されている．この翻訳方式では，文型パターンを用いることで，統語構造と意味を一体化した処理が可能である．過去において，単語レベル文型パターン翻訳方式及び句レベル文型パターン翻訳方式が提案されている．しかし，単語レベル文型パターン翻訳方式及び句レベル文型パターン翻訳方式では，パターンマッチ文数が少ないという問題点が存在する．そこで，本研究では，パターンマッチ文数を増加するための翻訳方式として節レベル文型パターン翻訳方式を提案し試作する．また，従来の文型パターン翻訳方式とパターンマッチ文数及び，翻訳精度の観点で比較評価を行う．

節レベル文型パターン翻訳方式を用いた翻訳実験より，単語レベル文型パターン翻訳方式及び句レベル文型パターン翻訳方式と提案手法を比較して，パターンマッチ文数の増加及び翻訳精度の低下が確認できた．今後，別の節パターンの作成方法の考案，文型パターン選択方法の調査を行うことにより，パターンマッチ文数及び翻訳精度が向上すると思われる．

目次

1	はじめに	1
2	日英重文複文文型パターン辞書	2
2.1	概要	2
2.2	単語レベル文型パターン	5
2.3	句レベル文型パターン	6
2.4	節レベル文型パターン	7
2.5	文型パターン辞書の被覆率	7
2.5.1	実験条件	7
2.5.2	実験結果	8
3	単語レベル文型パターン翻訳方式	9
3.1	概要	9
3.2	翻訳手順	9
3.3	翻訳例	10
4	句レベル文型パターン翻訳方式	11
4.1	概要	11
4.2	句パターン辞書	11
4.2.1	句パターン辞書の概要	11
4.2.2	Varlist	12
4.2.3	句パターン辞書の作成方法	12
4.2.4	句パターンの例	12
4.3	翻訳手順	13
4.4	翻訳例	14
5	提案手法 (節レベル文型パターン翻訳方式)	15
5.1	概要	15
5.2	節パターン辞書	15
5.2.1	節パターン辞書の概要	15
5.2.2	節パターン辞書の作成手順	15
5.2.3	節パターンの例	16

5.3	翻訳手順	17
6	実験条件	19
6.1	実験データ	19
6.2	パターンマッチ文数	20
6.3	翻訳精度	21
6.4	実験結果	22
6.5	パターンマッチ文数の結果	22
6.6	翻訳精度の結果	23
6.6.1	自動評価の結果	23
6.6.2	人手評価結果	23
7	考察	35
7.1	パターンマッチ文数について	35
7.2	翻訳精度について	38
8	おわりに	39

目 次

1	原文の記述例	5
2	単語レベル文型パターンの例	6
3	句レベル文型パターンの例	6
4	節レベル文型パターンの例	7
5	統語的被覆率と意味的被覆率	8
6	単語レベル文型パターン翻訳方式の流れ	10
7	Varlist の記述例	12
8	句レベル文型パターン翻訳方式の流れ	14
9	節パターン辞書作成の流れ	16
10	提案手法の流れ	18
11	他の翻訳システムを利用	37

目 次

2.1	変数	4
2.2	関数の例	4
2.3	記号の例	5
3.1	単語レベル文型パターン翻訳方式の翻訳例	10
4.1	句パターン (VP) の例	13
4.2	句レベル文型パターン翻訳方式の翻訳例	14
5.1	節パターンの例	17
6.1	パターンマッチ文数 (5,000 文中)	22
6.2	パターンマッチ文数として数えられた文の例 (提案手法)	22
6.3	自動評価の結果	23
6.4	人手評価結果 (提案手法 VS. 単語レベル文型パターン翻訳方式)	23
6.5	提案手法 > 単語レベル文型パターン翻訳方式の例 1	24
6.6	提案手法 > 単語レベル文型パターン翻訳方式の例 2	24
6.7	提案手法 > 単語レベル文型パターン翻訳方式の例 3	25
6.8	提案手法 < 単語レベル文型パターン翻訳方式の例 1	26
6.9	提案手法 < 単語レベル文型パターン翻訳方式の例 2	27
6.10	提案手法 < 単語レベル文型パターン翻訳方式の例 3	27
6.11	提案手法 = 単語レベル文型パターン翻訳方式の例 1	28
6.12	提案手法 = 単語レベル文型パターン翻訳方式の例 2	28
6.13	提案手法 = 単語レベル文型パターン翻訳方式の例 3	29
6.14	人手評価結果 (提案手法 VS. 句レベル文型パターン翻訳方式)	30
6.15	提案手法 > 句レベル文型パターン翻訳方式の例 1	30
6.16	提案手法 > 句レベル文型パターン翻訳方式の例 2	31
6.17	提案手法 > 句レベル文型パターン翻訳方式の例 3	31
6.18	提案手法 < 句レベル文型パターン翻訳方式の例 1	32
6.19	提案手法 < 句レベル文型パターン翻訳方式の例 2	32
6.20	提案手法 < 句レベル文型パターン翻訳方式の例 3	33
6.21	提案手法 = 句レベル文型パターン翻訳方式の例 1	33
6.22	提案手法 = 句レベル文型パターン翻訳方式の例 2	34
6.23	提案手法 = 句レベル文型パターン翻訳方式の例 3	34

7.1	節パターン辞書の作成で句レベル文型パターンを用いることで良くなると 思われる例	36
7.2	単文文型パターン辞書が有効な例	36
7.3	不適切な節レベル文型パターンが照合された例	38
7.4	提案手法で翻訳精度が低かった例	38

1 はじめに

言語表現の構造を，意味を持つ単位にパターン化した文型パターン [1] を用いた，文型パターン翻訳方式 [2] が提案されている．この翻訳方式では，文型パターンを用いることで，統語構造と意味を一体化した処理が可能である．また，文型パターンは変数の範囲に応じて単語，句，節の3つのレベルごとにパターンが作成されている．節は句と単語から成り立ち，句は単語から成り立つという関係性がある．

過去において，単語レベル文型パターンを用いた翻訳方式として，単語レベル文型パターン翻訳方式 [2] が実装され，句レベル文型パターンを用いた翻訳方式として，句レベル文型パターン翻訳方式 [3] が実装された．しかし，単語レベル文型パターン翻訳方式及び句レベル文型パターン翻訳方式は，1つの入力文に対して1つ以上のパターンがマッチする，パターンマッチ文数が少ないという問題点が存在する．パターンマッチ文数を増加させるための1つの手法として，過去において利用されていない節レベル文型パターンを用いる翻訳方式が考えられる．

そこで，本研究では，パターンマッチ文数を増加するための翻訳方式として節レベル文型パターンを用いる，節レベル文型パターン翻訳方式を提案し，試作する．また，パターンマッチ文数及び翻訳精度の観点で，従来の文型パターン翻訳方式と比較評価を行う．

本論文の構成を以下に示す．2章で文型パターン翻訳方式で利用する日英重文複文文型パターン辞書について説明する．3章では単語レベル文型パターン翻訳方式について説明し，4章では句レベル文型パターン翻訳方式を説明する．5章では提案手法である節レベル文型パターン翻訳方式について説明する．6章では実験条件について説明し，最後に，7章で本研究の考察を述べる．

2 日英重文複文文型パターン辞書

2.1 概要

日英重文複文文型パターン辞書は、日本語と英語の重文複文を対象に非線形言語モデルに基づいた文型パターンが記述されている辞書であり、約13万文の対訳コーパスより作成された [1].

非線形言語モデルとは、言語表現に含まれる線形な要素と非線形な要素を見極め、線形表現要素を汎化することによって非線形な表現構造を取り出そうとするものである [4]. 線形要素とは、言語表現の構成要素 (単語, 句, 節, など) のうち, それと同等の意味 (概念) を持つ他の要素に置き換えても元の言語表現全体の意味 (概念) が変化しない要素を「その言語表現に対する線形要素」という. 逆に, 言語表現全体の意味 (概念) が変化するとき, その要素を「その言語表現に対する非線形要素」という. また, 線形要素は, それ自身1つの言語表現 (句や節) であり, 非線形要素と線形要素で構成されていると考えることもできる.

文型パターンは, 線形な要素を変数として記述しており, 非線形な要素を記号, 関数や字面で表している. 変数には単語変数, 句変数, 節変数が存在する. 単語変数は, 表現に含まれる名詞, 動詞, 形容詞, 副詞などの自立語を表現要素とし, そのうちの線形なものを変数化している. 句変数は, 名詞句, 形容詞句, 動詞句, 副詞句など句を要素とし, そのうちの線形なものを変数化している. 節変数は, 連体節, 連用節などを要素と考え, そのうちの線形なものを変数化している. 単語変数は10種類に分類されており, 句変数は5種類に分類されており, 節変数は1種類である. 表2.1に, 変数についてまとめた表を示す. 関数は, 表記上の揺らぎを吸収し, 適用範囲の広い文型パターンを記述するために用いられている. 関数は, 様相関数と同値型グループ関数の2種類に分けられる. 様相関数は39種類あり, 時制, 相, 態, 様相などの助動詞相当表現を指定している. 同値型グループ関数は2種類あり, 同種の意味の関数をグループ化して指定している. 表2.2に関数の例を示す. 記号は, 文型パターン全体の非線形な構造を保ちながら, 構造上の線形性を記述するために用いられている. 記号は, 離散記号, 文節境界記号, 要素選択記号, 任意要素記号, 順序任意記号, 位置変更可能記号の6種類である. 記号の例を表2.3に示す.

また, 文型パターンは言語表現の文法的な構成単位に着目して3レベルで記述されている. 単語変数のみを用いて単語レベル文型パターンが作成されており, 単語変数及び句変数を用いて句レベル文型パターンが作成されている. また, 単語変数, 句変数及び

節変数を用いて節レベル文型パターンが作成されている。単語レベル文型パターンについて 2.2 節で説明し，句レベル文型パターンについて 2.3 節で説明し，節レベル文型パターンについて 2.4 節で説明する。そして，2.5 節で文型パターン辞書の被覆率について説明する。

表 2.1: 変数

変数名	意味
<i>N</i>	名詞または名詞複合語
<i>V</i>	動詞
<i>AJ</i>	形容詞
<i>AJV</i>	形容動詞
<i>ADV</i>	副詞
<i>TIME</i>	時詞
<i>GEN</i>	限定詞
<i>NUM</i>	数詞
<i>REN</i>	連体詞
<i>ND</i>	用言性名詞
<i>NP</i>	名詞句
<i>VP</i>	動詞句
<i>AJP</i>	形容詞句
<i>AJVP</i>	形容動詞句
<i>ADVP</i>	副詞句
<i>CL</i>	節

表 2.2: 関数の例

記述子名	記述形式	意味
様相関数	.kako	日本語の過去形
様相関数	^past	英語の過去形
同値型グループ関数	.#da	体言述部“だ”, “で”などを表す

表 2.3: 記号の例

記述子名	記述形式	意味
離散記号	\y	連用節
文節境界記号	!	文節の境界を受理
要素選択記号	(... ...)	いずれかの要素列を受理
任意要素記号	[...]	文型選択上, 任意の要素
順序任意記号	{..., ...}	順序入れ替え可能な範囲
位置変更可能記号	\$n^..., \$n	指定位置に入れ替え可能

2.2 単語レベル文型パターン

単語レベル文型パターンについて説明する。単語レベル文型パターンとは、原文の表現構造を単語変数、関数、記号、字面で記述した文型パターンで、対訳例文に存在する線形な自立語を単語変数化している。単語レベル文型パターンは対訳例文約 13 万文より、121,904 件作成された。原文の記述例を図 1 に示す。また、図 2 に単語レベル文型パターンの例を示す。

AC000213-00

LJ000044-00:信号が青になってから道路を渡りなさい。

LE000042-00:Cross the street after the traffic light turns green.

図 1: 原文の記述例

図 1 において、AC000213-00 は文型パターン ID である。また、LJ000044-00 は日本語原文 ID であり、LE000042-00 は英語原文 ID である。

図 2 において、WJ000043-00 は日本語単語レベル文型パターン ID である。また、NI, NY 等は意味分類番号である。また、WE000042-00 は英語単語レベル文型パターン ID である。

WJ000043-00:/y#1{/tcfkN1(OR:信号,NI:1100,NI:926)が,/tcfkN2(OR:青,NI:2352)に}/cfなってから/ytckN3(OR:道路,NI:418)を/cf(V4(OR:渡り,NY:0200,NY:0506)^meirei|V4(NY:0200,NY:0506).meireigo)。
WE000042-00:V4(OR:Cross) N3(OR:the street) after N1(OR:the traffic light) turn N2(OR:green)。

図 2: 単語レベル文型パターンの例

2.3 句レベル文型パターン

句レベル文型パターンについて説明する。句レベル文型パターンとは、原文の表現構造を単語変数、句変数、関数、記号、字面で記述した文型パターンである。作成されたすべての単語レベル文型パターンを対象に句レベル文型パターンに汎化が試みられた。すでに単語レベルで汎化された要素(変数)については、それ自身が句レベルの変数に置換できるかどうか、また、それを含むより大きな要素が汎化出来ないかを調べられた。作成された句レベル文型パターンは 79,438 件である。図 3 に句レベル文型パターンの例を示す。

PJ000030-00:/ytcfkN1(OR:信号,NI:1100,NI:926)が!VP2(OR:青になっ,NY:0506,NY:0802)(てから|でから)/ytck(VP3(OR:道路を渡り,NY:0200,NY:0506)^meirei|VP3(NY:0200,NY:0506).meireigo)。
PE000029-00:VP3(OR:Cross the street) after N1(OR:the traffic light) VP2(OR:turns green)。

図 3: 句レベル文型パターンの例

PJ000030-00 は日本語句レベル文型パターン ID であり、PE000029 は英語句レベル文型パターン ID である。

2.4 節レベル文型パターン

節レベル文型パターンについて説明する。節レベル文型パターンとは、原文の表現構造を単語変数、句変数、関数、記号、字面で記述した文型パターンである。節レベル文型パターンの汎化は、線形な節(「事態命名のレベル」の単文(核文))を発見してそれを節変数に置換することで行われた。作成された節レベル文型パターンは25,475件である。図4に節レベル文型パターンの例を示す。

```
CJ000012-00:/ytcfkCL1(OR:信号が青になっ,NY:0506,NY:0802)(てから|でから)
/ytck(CL2(OR:道路を渡り,NY:0200,NY:0506)^meirei|CL2(NY:0200,NY:0506).meireigo)。
```

```
CE000012-00:CL2(OR:Cross the street) after CL1(OR:the traffic light
turns green)。
```

図 4: 節レベル文型パターンの例

CJ000012-00 は日本語節レベル文型パターン ID であり, CE000012-00 は英語節レベル文型パターン ID である。

2.5 文型パターン辞書の被覆率

本節は [4] を参照して記述している。過去において作成された文型パターン辞書の被覆率が調査された。評価方法は入力文とのパターン照合実験によって、パターン辞書の被覆率特性を評価した。実験条件を 2.5.1 に示す。

2.5.1 実験条件

(a) 変数の意味的な制約条件は無視する

(b) 実験はクロスヴァリデーションの方法

クロスヴァリデーションの方法は、標本データを「学習に使用するデータ」と「試験に使用するデータ」に分け、それらを順次入れ替えて実験する方法である。少ない標本データを使用して信頼性の高いオープンテストを行うための方法として使用されている。実験では、文型パターン作成用の例文の中からランダムに1万文を

抽出し、入力文として使用された。また、当該入力文から作成された文型パターン(自己パターン)へは必ず適合するため無視し、それ以外のパターンへの適合のみ評価された。

入力文に対して適合する文型パターンは通常複数存在し、それらがすべて意味的に正しいとは限らない。そこで、被覆率は以下のパラメータにより評価された。

- 統語的被覆率

入力した文のうち、1文型パターン以上に適合した入力文の割合

- 意味的被覆率

1入力文に適合した文型パターンの中に意味的に正しく、入力文の翻訳に使えるパターンが1つ以上存在する割合

2.5.2 実験結果

図5に実験結果を示す。

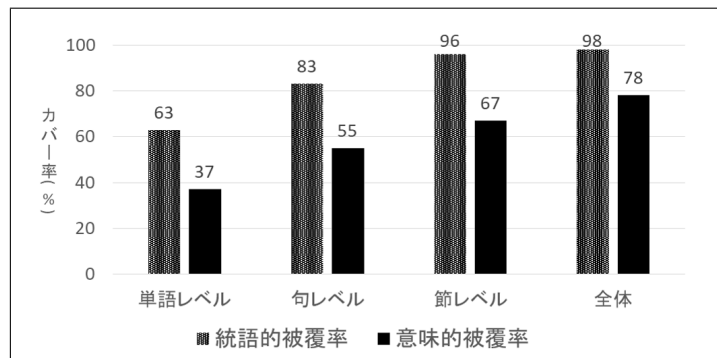


図 5: 統語的被覆率と意味的被覆率

図5より統語的被覆率、意味的被覆率ともに、単語レベル文型パターンが最も低く、節レベル文型パターンが最も高いことが確認できた。また、3レベルの文型パターンは単語レベル、句レベル、節レベルの順に表現の幅が広がるため、意味の曖昧さは増大する。

3 単語レベル文型パターン翻訳方式

本章では、従来手法の1つとして、単語レベル文型パターン翻訳方式について説明する。

3.1 概要

単語レベル文型パターン翻訳方式では、日英重文複文文型パターン辞書の単語レベル文型パターンを利用して翻訳を行う。ただし、実装が困難なため、単語レベル文型パターンに存在する離散記号は使用していない。また、文型パターンの選択の際には、意味分類により文型パターンの絞り込みを行っている。評価値の算出の際に、訳語候補の選択には言語モデル (tri-gram) 及び翻訳モデルを用いている。3.2 節に翻訳手順を示し、3.3 節に翻訳例を示す。

3.2 翻訳手順

以下に単語レベル文型パターン翻訳方式の翻訳手順を示す。文型パターン照合にはSPM[6]を使用している。また、図6に翻訳の流れを示す。

手順1 入力文を形態素解析

手順2 手順1の結果と単語レベル文型パターンを照合

手順3 意味分類により文型パターンの絞り込み

手順4 得られた文型パターンの各要素についてそれぞれ文頭から文末にかけて要素を
訳出

手順4.1 要素が字面ならば英語文型パターンの対応する記述を利用

手順4.2 要素が単語変数ならば単語辞書を用いて翻訳候補を訳出

手順5 翻訳候補文すべての評価値を算出 (言語モデル, 翻訳モデルより算出)

手順6 評価値の最も高い文を出力

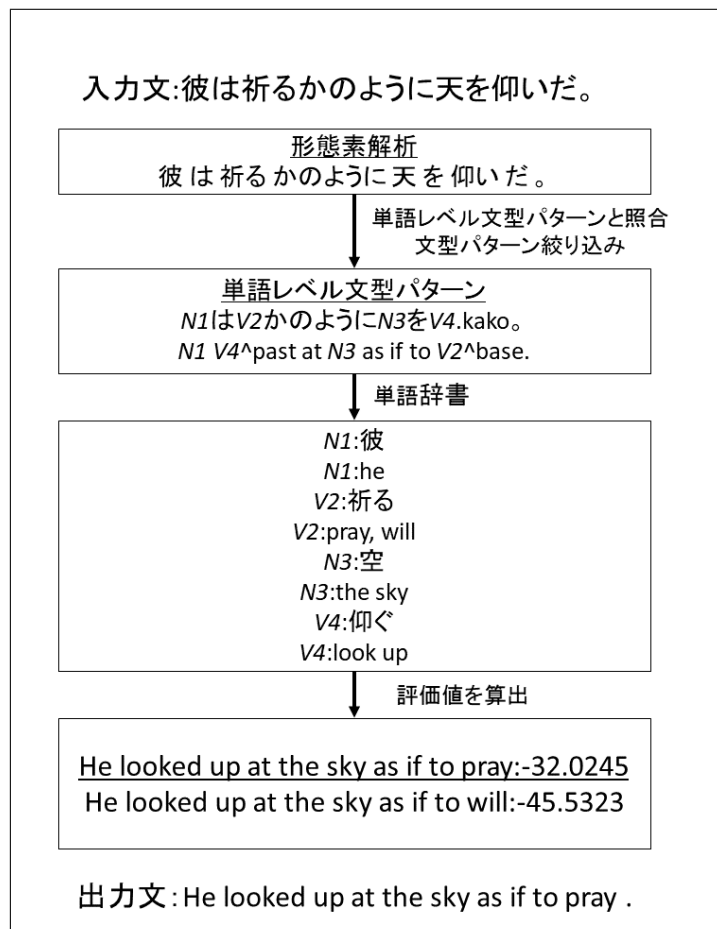


図 6: 単語レベル文型パターン翻訳方式の流れ

3.3 翻訳例

単語レベル文型パターン翻訳方式を用いた翻訳例を表 3.1 に示す。

表 3.1: 単語レベル文型パターン翻訳方式の翻訳例

入力文	君が辞任する必要はない。
参照文	You do not have to resign.
出力文	There is no need for you to resign .
文型パターン (日)	N1がV2^rentai N3はない。
文型パターン (英)	There is no N3 for N1 to V2^base .

4 句レベル文型パターン翻訳方式

本章では、従来手法の1つとして句レベル文型パターン翻訳方式について説明する。

4.1 概要

句レベル文型パターン翻訳方式では、日英重文複文文型パターン辞書の句レベル文型パターンを利用して翻訳を行う。ただし、実装が困難なため、句レベル文型パターンに存在する離散記号は使用していない。また、文型パターンの選択の際には、意味分類により文型パターンの絞り込みを行っている。そして、句レベル文型パターンには句変数が存在する。句変数に対応する記述の翻訳方法には、句パターン辞書を利用する。句パターン辞書は、句変数の構造を表すパターン辞書であり、単語レベル文型パターンを用いて作成された。4.2節に句パターン辞書について説明する。また、4.3節に翻訳手順を示し、4.4節に翻訳例を示す。

4.2 句パターン辞書

4.2.1 句パターン辞書の概要

句パターン辞書とは、句変数の構造を表すパターン辞書である。句変数(日)は5種類存在し、それぞれについてパターン辞書を作成する。句パターン辞書の作成には、単語レベル文型パターンの句変数に対応する記述を用いて、プログラムにより自動的に作成する。また、単語レベル文型パターンの句変数に対応する記述は、日英重文複文文型パターン辞書の作成途中に作成された、“Varlist”を利用する。Varlistについて4.2.2節で説明する。4.2.3節に句パターン辞書の作成手順を示し、4.2.4節で句パターンの例を示す。

4.2.2 Varlist

Varlist には、文型パターン ID、変数、日本語原文での出現位置、日本語原文での記述、日本語原文での変数の種類、英語原文での出現位置、英語原文での記述、英語原文での変数の種類が記述されている。図 7 に Varlist の記述例を示す。

AC000213-00	W	11:道路:N@7:the street:N
AC000213-00	C	11:道路を渡り:CL@1:Cross the street:CL
AC000213-00	P	11:道路を渡り:VP@1:Cross the street:VP
AC000213-00	W	14:渡り:V@1:Cross:V
AC000213-00	WP	1:信号:N@24:the traffic light:N
AC000213-00	C	1:信号が青になっ:CL@24:the traffic light turns green:CL
AC000213-00	W	4:青:N@48:green:N
AC000213-00	P	4:青になっ:VP@42:turns green:VP

図 7: Varlist の記述例

図 7 中の W は単語変数を示し、P は句変数を示し、C は節変数であることを示している。

4.2.3 句パターン辞書の作成方法

句パターン辞書の作成手順を以下に示す。

手順 1 Varlist より句変数に関する情報を抽出。

手順 2 手順 1 で得られた情報を用いて単語レベル文型パターンより句変数に対応する記述を抽出。

手順 3 手順 2 で得られた記述を日本語標準表記に修正。

手順 4 句変数の表記や句変数に対応する単語レベル文型パターンの記述に誤りがある句パターンを削除。

4.2.4 句パターンの例

作成された句パターンの例を表 4.1 に示す。

「@」は過去形など時制の変化や 3 人称単数形などに対応させるための記号である。

表 4.1: 句パターン (VP) の例

日本語原文 (一部)	青になっ
英語原文 (一部)	turns green
句パターン (日)	<i>N1</i> に “なる”
句パターン (英)	@turn <i>N1</i>

4.3 翻訳手順

句レベル文型パターン翻訳方式の翻訳手順を以下に示す。文型パターン照合にはSPM[6]を使用している。また、評価値の算出には、言語モデル (tri-gram) 及び翻訳モデルを用いている。翻訳の流れを図8に示す。

手順 1 入力文を形態素解析

手順 2 手順 1 の結果と句レベル文型パターンを照合

手順 3 意味分類により文型パターンの絞り込み

手順 4 得られた文型パターンの各要素についてそれぞれ文頭から文末にかけて要素を
訳出

手順 4.1 要素が字面ならば英語文型パターンの対応する記述を訳出

手順 4.2 要素が単語変数ならば、単語辞書を用いて翻訳候補を訳出

手順 4.3 要素が句変数ならば、句パターン辞書及び単語辞書を利用して翻訳候補を訳出

手順 5 翻訳候補文すべての評価値を算出

手順 6 評価値の最も高い文を出力

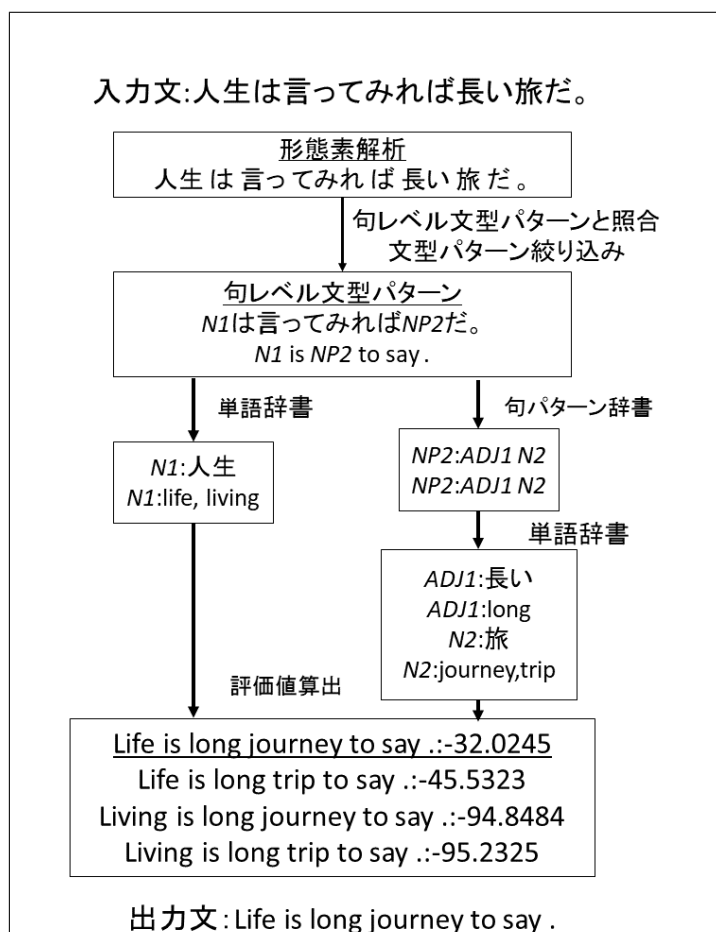


図 8: 句レベル文型パターン翻訳方式の流れ

4.4 翻訳例

句レベル文型パターン翻訳方式を用いた翻訳例を表 4.2 に示す。

表 4.2: 句レベル文型パターン翻訳方式の翻訳例

入力文	いつか原作者に会いたいと思う。
参照文	I would like to see the author some time.
出力文	I would like to see the author some day .
文型パターン(日)	VP2.tai と'思う'.genzai。
文型パターン(英)	<I N1> @will#1(^present ^past) like to VP2^base .

5 提案手法 (節レベル文型パターン翻訳方式)

5.1 概要

従来の文型パターン翻訳方式は、単語レベル文型パターンや句レベル文型パターンを用いている。しかし、2.5節より、単語レベル文型パターンや句レベル文型パターンは節レベル文型パターンと比較して統語的被覆率、意味的被覆率が低いことが明らかとなっており、1つの文に対して1つ以上のパターンがマッチするパターンマッチ文数が少ないという問題点がある。そこで本研究では、パターンマッチ文数を増加するため、節レベル文型パターン¹を利用する“節レベル文型パターン翻訳方式”を提案する。具体的には、入力文と節レベル文型パターンを照合し、得られた節レベル文型パターンの各要素を翻訳する手法である。また、節レベル文型パターンには、節変数が存在するが、節変数に対応する記述を翻訳する方法は未だ提案されていない。そこで本研究では、節変数の構造を単語レベル文型パターンを用いて表現した“節パターン辞書”を作成する。単語レベル文型パターンを利用して節パターン辞書を作成することにより、節変数の構造を統語的意味的に解析することができると期待したためである。5.2節で節パターン辞書について説明し、5.3節で翻訳手順について説明する。

5.2 節パターン辞書

本節では、節パターン辞書について説明する。節パターン辞書の概要について5.2.1節で説明し、5.2.2節で節パターン辞書の作成手順を説明する。

5.2.1 節パターン辞書の概要

節パターン辞書とは、節変数の構造を表現するための辞書である。本研究では、節パターン辞書の作成に単語レベル文型パターンの記述を利用する。単語レベル文型パターンを利用する理由として、曖昧性が少なく、節変数の構造を統語的意味的に解析することができると考えためである。

5.2.2 節パターン辞書の作成手順

節パターン辞書の作成手順を以下に示す。

¹実装が困難なため、離散記号の情報は利用していない

手順1 Varlist より節変数に関する情報を抽出。

手順2 手順1 で得られた情報を用いて単語レベル文型パターンより節変数に対応する記述を抽出。

手順3 手順2 で得られた日本語記述を標準表記に修正。

手順4 節変数の表記や節変数に対応する単語レベル文型パターンの記述に誤りがある句パターンを削除。

作成手順を表した図を図9に示す。

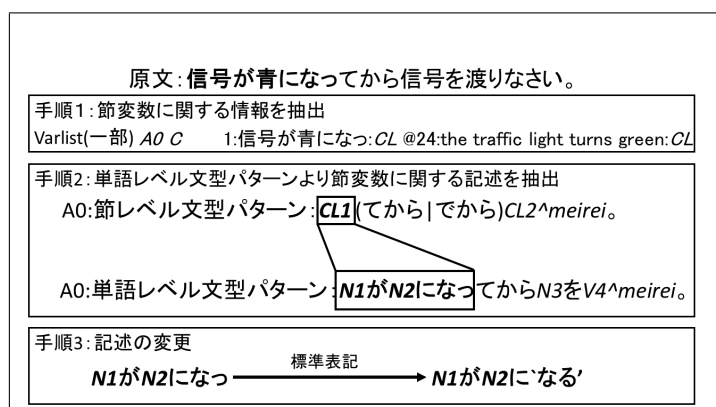


図9: 節パターン辞書作成の流れ

図9について説明する。手順1で Varlist より節変数に対応する記述である文型パターンID(A0), 変数名(C), 日本語記述, 英語記述などを抽出している。次に手順2で, 手順1で得られた情報を利用し, 同文型パターンIDの単語レベル文型パターンより節変数に関する記述(N1がN2になっ)を抽出している。次に手順3で, 抽出した日本語記述を標準表記(N1にN2に‘なる’)に変更している。また, 図9では, 日本語記述の抽出方法を示しているが, 英語記述の情報は手順2の際に単語レベル文型パターンの英語記述より抽出している。

5.2.3 節パターンの例

表5.1に作成した節パターンの例を示す。

表 5.1: 節パターンの例

日本語原文 (一部)	信号が青になっ
英語原文 (一部)	the traffic light turns green
節パターン(日)	$N1$ が $N2$ に'なる'
節パターン(英)	$N1$ @turn $N2$

5.3 翻訳手順

提案手法の翻訳手順は以下である。文型パターン照合にはSPM[6]を使用している。また、評価値の算出には言語モデル (tri-gram) 及び翻訳モデルを用いている。提案手法の流れ図を図 10 に示す。

手順 1 入力文を形態素解析

手順 2 手順 1 の結果と節レベル文型パターンを照合

手順 3 意味分類により文型パターンの絞り込み

手順 4 得られた文型パターンの各要素についてそれぞれ文頭から文末にかけて要素を
訳出

手順 4.1 要素が字面ならば英語文型パターンの対応する記述を訳出

手順 4.2 要素が単語変数ならば、単語辞書を用いて翻訳候補を訳出

手順 4.3 要素が句変数ならば、句パターン辞書及び単語辞書を利用して翻訳候補を訳出

手順 4.4 要素が節変数ならば、節パターン辞書及び単語辞書を利用して翻訳候補を訳出

手順 5 翻訳候補文すべての評価値を算出

手順 6 評価値の最も高い文を出力

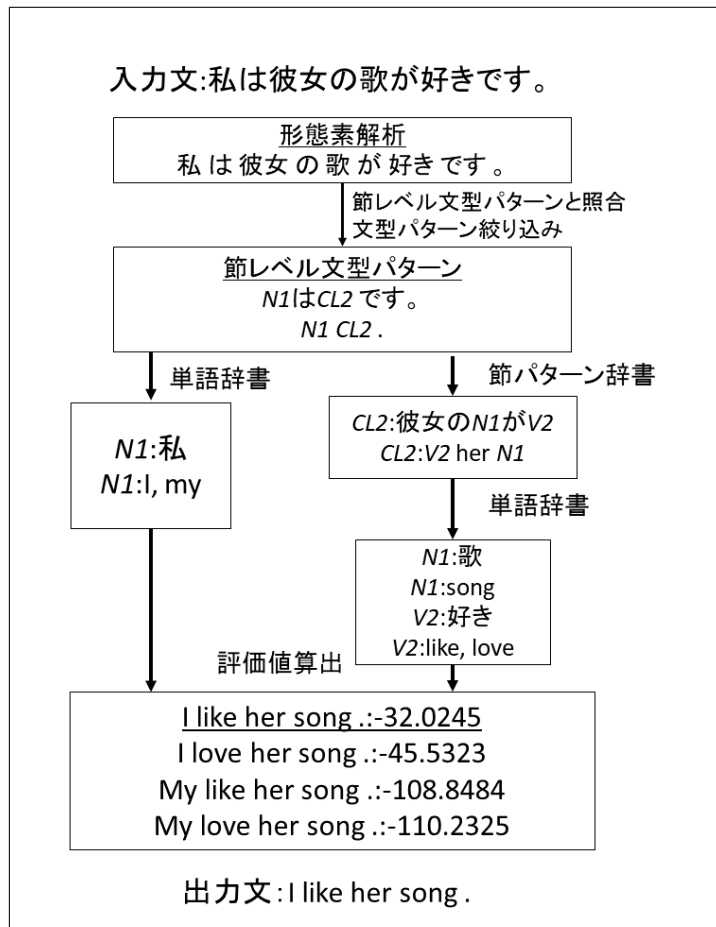


図 10: 提案手法の流れ

6 実験条件

本研究では、単語レベル文型パターン翻訳方式と提案手法及び句レベル文型パターン翻訳方式と提案手法でパターンマッチ文数及び翻訳精度の観点で比較評価を行う。6.1節に実験データについて示し、6.2節でパターンマッチ文数について説明し、6.3節で翻訳精度について説明する。

6.1 実験データ

本研究では、日英重文複文文型パターン辞書の原文対約12万から10万文対を抽出し以下に使用する。

- 翻訳モデルの学習・言語モデルの学習
- 単語辞書の作成
- 句パターン辞書の作成
- 節パターン辞書の作成

残り2万文対から5,000文対を抽出し、テスト文に用いる。提案手法で照合する節レベル文型パターンは20,851件である。また、句レベル文型パターン翻訳方式で照合する文型パターンは65,171件であり、単語レベル文型パターン翻訳方式で照合する文型パターンは100,000件である。また、文型パターン照合の際、得られた文型パターンの数が10件以上である場合は翻訳を行うのに大量の時間が必要になるため、10件に制限して実験を行う。

6.2 パターンマッチ文数

パターンマッチ文数について説明する。

単語レベル文型パターン翻訳方式においてパターンマッチ文数とは，入力文と単語レベル文型パターンの照合に成功し，得られた単語レベル文型パターンに存在する単語変数を訳出できた文の数である。

句レベル文型パターン翻訳方式においてパターンマッチ文数とは，入力文と句レベル文型パターンの照合に成功し，得られた句レベル文型パターンに存在する，句変数の構造を句パターン辞書を用いて表現できた文の数である。

提案手法においてパターンマッチ文数とは，入力文と節レベル文型パターンの照合に成功し，得られた節レベル文型パターンに存在する，節変数の構造を節パターン辞書を用いて表現できた文の数である。

6.3 翻訳精度

本研究では、自動評価 (BLEU, METEOR, RIBES, TER) 及び対比較人手評価を用いて翻訳精度を評価する。自動評価にはパターンマッチ文数で得られた文を用いる。また、対比較人手評価には、単語レベル文型パターン翻訳方式及び提案手法での比較として両手法で未知語のない文 100 文を用い、句レベル文型パターン翻訳方式及び提案手法での比較として両手法で未知語のない文 100 文を用いる。

6.4 実験結果

6.5 節にパターンマッチ文数の結果について示し，6.6 節に翻訳精度の結果を示す。

6.5 パターンマッチ文数の結果

表 6.1 にパターンマッチ文数の結果を示す。

表 6.1: パターンマッチ文数 (5,000 文中)

提案手法	単語レベル	句レベル
1,842 文 (37%)	483 文 (10%)	1,180 文 (24%)

表 6.1 より提案手法は，従来の翻訳方式よりもパターンマッチ文数が増加したことが確認できた。提案手法でパターンマッチ文数として数えられた文の例を表 6.2 に示す。

表 6.2: パターンマッチ文数として数えられた文の例 (提案手法)

入力文	その知らせを聞いて彼女の顔は明るくなった。
出力文	Her face lightened when she heard that the news .
文型パターン (日)	$CL1^{\wedge}renyou$ て $CL2.kako$ 。
文型パターン (英)	$CL2$ when $CL1$.

6.6 翻訳精度の結果

自動評価の結果を 6.6.1 節に示し，人手評価の結果を 6.6.2 節に示す。

6.6.1 自動評価の結果

自動評価の結果を表 6.3 に示す。評価文数は，パターンマッチ文数として数えられた出力文の数である。提案手法では，1,842 文であり，単語レベル文型パターン翻訳方式では，483 文であり，句レベル文型パターン翻訳方式では，1,180 文である。

表 6.3: 自動評価の結果

	BLEU	METEOR	RIBES	TER ²
提案手法	0.0798	0.2992	0.3168	1.0473
従来手法 (単語)	0.3379	0.6032	0.6150	0.7558
従来手法 (句)	0.1591	0.4148	0.4710	0.9153

自動評価の結果より提案手法の翻訳精度の低下が確認できた。

6.6.2 人手評価結果

提案手法及び単語レベル文型パターン翻訳方式の対比較人手評価の結果を表 6.4 に示す。

表 6.4: 人手評価結果 (提案手法 VS. 単語レベル文型パターン翻訳方式)

提案手法○	単語レベル文型パターン翻訳方式○	差なし
3 文	76 文	21 文

表 6.4 より，提案手法よりも単語レベル文型パターン翻訳方式のほうが翻訳精度が高いことがわかる。提案手法の方が単語レベル文型パターン翻訳方式よりも良かった例を表 6.5，表 6.6，表 6.7 に示す。また，単語レベル文型パターン翻訳方式の方が提案手法よりも良かった例を表 6.8，表 6.9，表 6.10 に示し，差なしの例を表 6.11，表 6.12，表 6.13 に示す。

²TER は翻訳のエラー率を示す評価指標であり，数値が 0 に近づくほどよい。

表 6.5: 提案手法 > 単語レベル文型パターン翻訳方式の例 1

入力文	彼の服装は彼が貧乏であることを示している。
参照文	His clothes argue him to be poor .
提案手法	his clothes argue that he is poor .
単語レベル文型パターン翻訳方式	his clothes argue poverty .
節レベル文型パターン (日)	NP1 は CL2.tearu.genzai^rentai ことを V3.teiru.genzai。
節レベル文型パターン (英)	NP1 V3#2(^present ^past) that CL2 .
節パターン CL2	(日)N1 が AJV2 (英)N1 @be AJ2
単語レベル文型パターン (日)	彼の N2 は N3 が AJV4.tearu^rentai ことを V5.teiru。
単語レベル文型パターン (英)	#1[N3^poss] N2 V5 N4 .

表 6.5 は、単語レベル文型パターン翻訳方式の出力では主語が “his clothes” となり誤っているため、提案手法のほうが単語レベル文型パターン翻訳方式よりも良い例とした。

表 6.6: 提案手法 > 単語レベル文型パターン翻訳方式の例 2

入力文	塀から飛び降りて足首をくじいた。
参照文	He wrenched his ankle in the jump down from the wall .
提案手法	you plunged down from the fence and he sprained his ankle .
単語レベル文型パターン翻訳方式	he plunged down from his fence and twisted ankle .
節レベル文型パターン (日)	CL1^renyou て CL2。
節レベル文型パターン (英)	CL1 and CL2 .
節パターン CL1	(日)N1 から V2 (英)you V2 from N1
節パターン CL2	(日)N1 を V2.kako (英)<He N3> V2 <his N100^poss> N1
単語レベル文型パターン (日)	N3 から V4 て N5 を V6.kako。
単語レベル文型パターン (英)	<He N1> V4^past from <his N2^poss> N3 and V6^past N5 .

表 6.6 は、単語レベル文型パターン翻訳方式の出力では “フェンス” に該当する訳が “his fence” となっているため、提案手法のほうが単語レベル文型パターンよりも良い例とした。

表 6.7: 提案手法 > 単語レベル文型パターン翻訳方式の例 3

入力文	彼は彼女が困っているのを見た。
参照文	He saw that she was troubled .
提案手法	he observed that she was troubled .
単語レベル文型パターン翻訳方式	he observed that she was embarrassed .
節レベル文型パターン (日)	$N1$ は $CL2$ \hat{r} entai のを $V3$.kako。
節レベル文型パターン (英)	$N1$ $V3$ \hat{p} ast that $CL2$.
節パターン $CL2$	(日) $N1$ が $V2$.teiru \hat{r} entai (英) $N1$ $V2$
単語レベル文型パターン (日)	$N1$ は $N2$ が $V3$.teiru \hat{r} entai のを $V4$.kako。
単語レベル文型パターン (英)	$N1$ $V4$ \hat{p} ast that $N2$ $V3$ \hat{p} ast .

表 6.7 は、単語レベル文型パターン翻訳方式の出力では“困っている”の表現を“embarrassed”となっており誤っていると判断したため、提案手法のほうが単語レベル文型パターン翻訳方式よりも良い例とした。

表 6.8: 提案手法 < 単語レベル文型パターン翻訳方式の例 1

入力文	この辞書は説明が詳しくて使いやすい。
参照文	This dictionary is very useful , giving detailed explanations .
提案手法	it is a clever , because it has detailed explanations .
単語レベル文型パターン翻訳方式	this dictionary is easier to use because of detailed explanations .
節レベル文型パターン (日)	CL1^renyou て CL2。
節レベル文型パターン (英)	CL2 , because CL1 .
節パターン CL1	(日)GEN1N2は N3 が AJ4 (英)N2^pron @have^present AJ4 N3
節パターン CL2	(日) AJ1 (英)It is AJ1
単語レベル文型パターン (日)	GEN2N3は N4 が AJ5 て V6AJ7。
単語レベル文型パターン (英)	#1[AJ2] N3 @be AJ7 to V6^base because of AJ5 N4 .

表 6.8 は、提案手法の出力には“辞書”という意味の単語が存在しないため、提案手法よりも単語レベル文型パターン翻訳方式のほうが良い例とした。

表 6.9 は、提案手法の出力では“i did a chest”となっており誤った訳となっているため、提案手法よりも単語レベル文型パターン翻訳方式のほうが良い例とした。

表 6.10 は、提案手法の出力では“he shot the policeman”となり誤っているため、提案手法よりも単語レベル文型パターン翻訳方式のほうが良い例とした。

表 6.9: 提案手法 < 単語レベル文型パターン翻訳方式の例 2

入力文	彼女はふくよかな胸をしている。
参照文	She has full breasts .
提案手法	she rouns , but i did a chest .
単語レベル文型パターン翻訳方式	she has a round chest .
節レベル文型パターン (日)	CL1CL2。
節レベル文型パターン (英)	CL1 , but CL2 .
節パターン CL1	(日)N1は AJV2 (英)N1 N2
節パターン CL2	(日)N2を V3.teiru (英)<I N3> #1[still] V3 #2[AJ1] N2
単語レベル文型パターン (日)	N1は AJV4 [^] rentaiN5をしている。
単語レベル文型パターン (英)	<She N1> @have [^] present (a an) #1[ADV3] N4 N5 .

表 6.10: 提案手法 < 単語レベル文型パターン翻訳方式の例 3

入力文	警官が犯人を狙って撃った。
参照文	The policeman took aim at and shot the criminal .
提案手法	he shot the policeman aimed at a criminal .
単語レベル文型パターン翻訳方式	the policeman aimed at and shot the criminal .
節レベル文型パターン (日)	CL1 [^] renyou て CL2.kako。
節レベル文型パターン (英)	CL2 CL1 [^] past .
節パターン CL1	(日)V1 (英)he V1
節パターン CL2	(日)N1が N2を V3 (英)N1 V3 at N2
単語レベル文型パターン (日)	N1が N2を V3 て V4.kako。
単語レベル文型パターン (英)	N1 V3 [^] past and V4 [^] past N2 .

表 6.11: 提案手法 = 単語レベル文型パターン翻訳方式の例 1

入力文	君が帰るか帰らないかに彼が来た。
参照文	No sooner had you left than he came .
提案手法	he came just when you left .
単語レベル文型パターン翻訳方式	he came just when you left .
節レベル文型パターン (日)	<i>N1</i> が帰るか帰らないかに <i>CL2</i> . <i>kako</i> 。
節レベル文型パターン (英)	<i>CL2</i> ^past just when <i>N1</i> left .
節パターン <i>CL2</i>	(日) <i>N1</i> が' 来る' (英) <i>N1</i> @come
単語レベル文型パターン (日)	<i>N1</i> が帰るか帰らないかに <i>N2</i> が <i>V3</i> . <i>kako</i> 。
単語レベル文型パターン (英)	<i>N2</i> <i>V3</i> ^past just when <i>N1</i> left .

表 6.11 は、両手法で出力文が同一であるため差なしの例とした。

表 6.12: 提案手法 = 単語レベル文型パターン翻訳方式の例 2

入力文	声を合わせて歌った。
参照文	We sang in chorus .
提案手法	To voice and sang .
単語レベル文型パターン翻訳方式	he sang in checking the voice .
節レベル文型パターン (日)	<i>CL1</i> ^renyou て <i>CL</i> 。
節レベル文型パターン (英)	<i>CL1</i> and <i>CL2</i> .
節パターン <i>CL1</i>	(日) <i>N1</i> を <i>V2</i> (英) <i>ADV2</i> <i>N1</i>
節パターン <i>CL2</i>	(日) <i>V1</i> . <i>kako</i> (英) <i>V1</i>
単語レベル文型パターン (日)	<i>N2</i> を <i>V3</i> て <i>V4</i> . <i>kako</i> 。
単語レベル文型パターン (英)	<He <i>N1</i> > <i>V4</i> ^past in <i>AJ3</i> <i>N2</i> .

表 6.12 は、両手法とも入力文の意味をほとんど読み取れないため差なしとした。

表 6.13: 提案手法 = 単語レベル文型パターン翻訳方式の例 3

入力文	彼の作品が入選したとはめでたい。
参照文	I am pleased that his work was accepted .
提案手法	i am delighted that his work was accepted .
単語レベル文型パターン翻訳方式	i am delighted that his work was accepted .
節レベル文型パターン (日)	CL1.kako とは'めでたい'.genzai。
節レベル文型パターン (英)	I @be#1(^present ^past) delighted that CL1 .
節パターン CL1	(日) 彼の作品が'入選する' (英)his @work was accepted
単語レベル文型パターン (日)	彼の作品が入選したとは'めでたい'.genzai,。
単語レベル文型パターン (英)	I @be#1(^present ^past) delighted that his work was accepted .

表 6.13 は, 両手法で出力文が同一であるため差なしの例とした。

提案手法及び句レベル文型パターン翻訳方式の対比較人手評価の結果を表 6.14 に示す。

表 6.14: 人手評価結果 (提案手法 VS. 句レベル文型パターン翻訳方式)

提案手法○	句レベル文型パターン翻訳方式○	差なし
27 文	51 文	21 文

表 6.14 より，提案手法よりも句レベル文型パターン翻訳方式のほうが翻訳精度が高いことがわかる。提案手法の方が句レベル文型パターン翻訳方式よりも良かった例を表 6.15，表 6.16，表 6.17 に示す。また，句レベル文型パターン翻訳方式の方が提案手法よりも良かった例を表 6.18，表 6.19，表 6.20 に示し，差なしの例を表 6.21，表 6.22，表 6.23 に示す。

表 6.15: 提案手法 > 句レベル文型パターン翻訳方式の例 1

入力文	酒がまわるとしゃべり出す。
参照文	Wine loosens people's tongues .
提案手法	he fell to talking when the wine took hold .
句レベル文型パターン翻訳方式	a tour of the bottle fell to talking .
節レベル文型パターン (日)	CL1^syushi と CL2.genzai.
節レベル文型パターン (英)	CL2 when CL1 .
節パターン CL1	(日)N3が'回る'。(英)N3 took hold
節パターン CL2	(日)V4 (英)he V4
句レベル文型パターン (日)	VP1 と VP2.
句レベル文型パターン (英)	NP1 VP2 .
句パターン VP1	(日)N1 が V2 (英)N2 of N1
句パターン VP2	単語変数として処理

表 6.15 は，句レベル文型パターン翻訳方式の出力が，“酒がまわる”という表現に対し，“a tour of the bottle” となり，誤った表現になっているため，提案手法のほうが良い例とした。

表 6.16 は，句レベル文型パターン翻訳方式の出力では，“has the truth” となり，誤った表現になっているため，提案手法のほうが良い例とした。

表 6.17 は，句レベル文型パターン翻訳方式の出力では，“to the grave” となり，誤った表現になっているため，提案手法のほうが良い例とした。

表 6.16: 提案手法 > 句レベル文型パターン翻訳方式の例 2

入力文	彼はそれがほんとうであることを認めた。
参照文	He allowed that it was true .
提案手法	he admitted it was the truth .
句レベル文型パターン翻訳方式	he admitted that it has the truth .
節レベル文型パターン (日)	$N1$ は $CL2^{\wedge}$ rentai ことを $VP3$.kako。
節レベル文型パターン (英)	$N1$ $VP3^{\wedge}$ past $CL2^{\wedge}$ past .
節パターン $CL2$	(日) $N1$ が $N2$.#da' ある' (英) $N1$ @be $N2$
句レベル文型パターン (日)	$N1$ は $NP2$ が $NP3$.#da あることを V_4 .kako。
句レベル文型パターン (英)	$N1$ V_4^{\wedge} past that $NP2$ @have [^] present $NP3$.
句パターン $NP2$	単語変数として処理
句パターン $NP3$	単語変数として処理

表 6.17: 提案手法 > 句レベル文型パターン翻訳方式の例 3

入力文	子どもが井戸へ入って死んだ。
参照文	A child drowned in the well .
提案手法	the child entered well and die .
句レベル文型パターン翻訳方式	the child entered well to the grave .
節レベル文型パターン (日)	$CL1^{\wedge}$ renyou て $CL2$.kako。
節レベル文型パターン (英)	$CL1$ and $CL2$.
節パターン $CL1$	(日) $N1$ が $N2$ へ' 入る' (英) $N1$ @enter $N2$
節パターン $CL2$	(日)'死ぬ' (英)@die
句レベル文型パターン (日)	$N1$ が $VP2$ て死んだ。
句レベル文型パターン (英)	$N1$ $VP2^{\wedge}$ past to the grave .
句パターン $VP2$	(日) $N1$ へ' 入る' (英)@enter $N1$

表 6.18: 提案手法 < 句レベル文型パターン翻訳方式の例 1

入力文	彼は非凡な学生だ。
参照文	He is no ordinary student .
提案手法	he is great , but you are a student .
句レベル文型パターン翻訳方式	he is a great student .
節レベル文型パターン (日)	CL1CL2。
節レベル文型パターン (英)	CL1 , but CL2 .
節パターン CL1	(日)N1は AJV2 :(英)N1 @be AJ2
節パターン CL2	(日)N1.#da (英)you are N1
句レベル文型パターン (日)	N1は NP2.#da。
句レベル文型パターン (英)	N1 @be NP2 .
句パターン NP1	(日)AJV1^rentaiN2 (英)AJ1 N2

表 6.18 は，提案手法の出力の “you are a student” という表現が主語が誤りであるため，句レベル文型パターン翻訳方式のほうが良い例とした。

表 6.19: 提案手法 < 句レベル文型パターン翻訳方式の例 2

入力文	彼女は幼いとき視力を失った。
参照文	She lost her eyesight in childhood .
提案手法	when she is small , he lost to sight .
句レベル文型パターン翻訳方式	when she was young , she lost sight .
節レベル文型パターン (日)	CL1.genzai^rentai とき CL2.kako。
節レベル文型パターン (英)	When CL1 , CL2 .
節パターン CL1	(日)N1は AJ2^rentai (英)N1 @be^past AJ2
節パターン CL2	(日)N1を V2 (英)he V2 to N1
句レベル文型パターン (日)	N1は AJ2^rentai とき VP3.kako。
句レベル文型パターン (英)	When N1 @be^past AJ2 , N1 VP3^past .
句パターン VP1	(日)N1を'失う'(英)@lose N1

表 6.19 は，提案手法の出力の “he lost” という表現が誤りであるため，句レベル文型パターン翻訳方式のほうが良い例とした。

表 6.20 は，提案手法の出力では “it will rain” という表現が誤りであるため，句レベル文型パターン翻訳方式のほうが良い例とした。

表 6.20: 提案手法 < 句レベル文型パターン翻訳方式の例 3

入力文	10 に 5 を加えると 15 になる。
参照文	Five added to 10 makes 15 .
提案手法	when ten add five , it will rain .
句レベル文型パターン翻訳方式	ten add five is fifteen .
節レベル文型パターン (日)	CL1 [^] syushi と CL2。
節レベル文型パターン (英)	When CL1 , CL2 .
節パターン CL1	(日)N1 に N2 を V3 (英)N1 V3 N2
節パターン CL2	(日) N2 に '成る' (英)#1[ADV1] it will @rain
句レベル文型パターン (日)	VP1 と VP2。
句レベル文型パターン (英)	NP1 VP2 .
句パターン VP1	(日)N1 に N2 を V3 (英)N1 V3 N2
句パターン VP2	(日)N1 に '成る' (英)@be N1

表 6.21: 提案手法 = 句レベル文型パターン翻訳方式の例 1

入力文	それが困難であることは認めます。
参照文	I admit that it is difficult.
提案手法	i admitted it was embarrassing .
句レベル文型パターン翻訳方式	i admitted it was embarrassing .
節レベル文型パターン (日)	CL1.tearu.genzai [^] rentai ことは V3.genzai。
節レベル文型パターン (英)	<I N2> V3#2([^] present [^] past) CL1 .
節パターン CL1	(日)N2 が AJV3 (英)N2 @be #1[ADV1] V3 N1 が AJVP2.tearu, .genzai, [^] rentai
句レベル文型パターン (日)	ことは V4.genzai。
句レベル文型パターン (英)	<I N3> V4#2([^] present [^] past) N1 @be#1([^] present [^] past) VP2 .
句パターン AJVP2	単語変数として処理

表 6.21 では、同一の出力であるため差なしの例とした。

表 6.22 では、同一の出力であるため差なしの例とした。

表 6.23 は、提案手法では、主語が “he” となっているのが誤りであり、句レベル文型パターン翻訳方式では、“he had committed” となっているため差なしとした。

表 6.22: 提案手法 = 句レベル文型パターン翻訳方式の例 2

入力文	彼は伝説を尋ねてその地方を訪れた。
参照文	He visited that area in search of legends .
提案手法	he asked legend and visited the local .
句レベル文型パターン翻訳方式	he asked legend and visited the local .
節レベル文型パターン (日)	CL1 ^{renyou} て CL2.kako。
節レベル文型パターン (英)	CL1 and CL2 .
節パターン CL1	(日)N1は N2を V3 (英)N1 V3 N2
節パターン CL2	(日)GEN1N2を V3 (英)V3 DEF1 N2
句レベル文型パターン (日)	N1は VP2 ^{te} VP3.kako。
句レベル文型パターン (英)	<He N1> VP2 ^{past} and VP3 ^{past} .
句パターン VP2	(日)N1を'尋ねる' (英)@ask N1
句パターン VP3	(日)GEN1N2を'訪れる' (英)@visit #1[DEF1] N2

表 6.23: 提案手法 = 句レベル文型パターン翻訳方式の例 3

入力文	彼女は罪を犯したと自供した。
参照文	She confessed herself guilty .
提案手法	he confessed that she committed the crime .
句レベル文型パターン翻訳方式	she confessed that he had committed the crime .
節レベル文型パターン (日)	CL1と VP3.kako。
節レベル文型パターン (英)	<He N2> VP3 ^{past} that CL1 ^{past} .
節パターン CL1	(日)N1は N2を V3.kako (英)N1 V3 N2
句レベル文型パターン (日)	N1は VP2.kakoと V3.kako。
句レベル文型パターン (英)	N1 V3 ^{past} that he had VP2 ^{ed} .
句パターン VP2	(日)N1を'犯す' (英)@commit N1

7 考察

7.1 パターンマッチ文数について

表 6.1 より提案手法は単語レベル文型パターン翻訳方式及び句レベル文型パターン翻訳方式よりもパターンマッチ文数が増加したのがわかる。しかし、まだパターンマッチ文数が少ないため更なるパターンマッチ文数の増加が期待される。パターンマッチ文数が少ない原因として以下のことが挙げられる。

- 原因 1
節パターンの数が少ない
- 原因 2
節変数の構造を単語レベル文型パターンで表すことが困難
- 原因 3
不適切な節レベル文型パターンと照合

これらの原因を解決する方法として以下の方法が考えられる。

- 方法 1
単語辞書、句パターン及び節パターンを人手で追加
- 方法 2
節パターン辞書の作成に句レベル文型パターンを利用
- 方法 3
単文文型パターン辞書 [7] を利用
- 方法 4
他の翻訳システムを利用
- 方法 5
文型パターン照合システムの改良

方法1は人手で作成するため、コストがかかる。

方法2は妥当性があり、より多くの節変数を表現することができ、パターンマッチ文数が増加するのではないかと考える。句変数は単語変数よりも多くの表現構造を表すことができるため、単語レベル文型パターンよりも句レベル文型パターンの方がより多くの表現を表すことができる。したがって、節パターン辞書を作成する際に単語レベル文型パターンを用いるのではなく、句レベル文型パターンを用いることで多くの節変数の表現構造を表すことができるのではないかと考える。表7.1に句レベル文型パターンを用いた節パターンの例を示す。表7.1の日英原文の一部を単語レベル文型パターンで表現すると、“ $AJ1N2$ の $N3$ が $N4$ に $V5$ ”となる。本実験ではこのような節パターンは作成されていないため節変数に対応する記述の翻訳に失敗した。

表 7.1: 節パターン辞書の作成で句レベル文型パターンを用いることで良くなると思われる例

日本語原文 (一部)	美しい古城のイメージが心に浮かぶ
英語原文 (一部)	an image of a beautiful old castle come into my mind
節パターン (日)	$NP1$ が $VP2$
節パターン (英)	$NP1$ $VP2$

方法3は節変数に対応する記述を単文文型パターンを用いることにより、多くの節変数の表現構造を表すのではないかと考える。日英重文複文文型パターン辞書において節変数の範囲は、「事態命名」の単文を変数化して作成されている。したがって、単文文型パターン辞書を利用することにより節変数に対応する単文の表記を翻訳できるのではないかと考える。表7.2に単文文型パターン辞書を用いることにより、パターンマッチ文数が増加すると思われる例を示す。

表 7.2: 単文文型パターン辞書が有効な例

入力文	出力文
歯並びを直すのに 矯正器を使っています。	<cl4:矯正器を使っています> though fix teeth
このぼんこつテレビを直そうと してもむだ骨を折るだけだ	<cl4:むだ骨を折るだけだ> even when I fix this jalopy boxes .
道路で遊ぶのはたいへん危険だ。	It is very dangerous that <cl1:道路で遊ぶ> .

方法4は他の翻訳システムを利用することにより、節変数に対応する記述を翻訳する方法である。具体的には、提案手法の出力文において節変数に対応する記述が翻訳され

ていない場合に他の翻訳システムに節変数に対応する記述を入力文として入力することにより節変数に対応する記述を翻訳する。図 11 に翻訳の流れを示す。

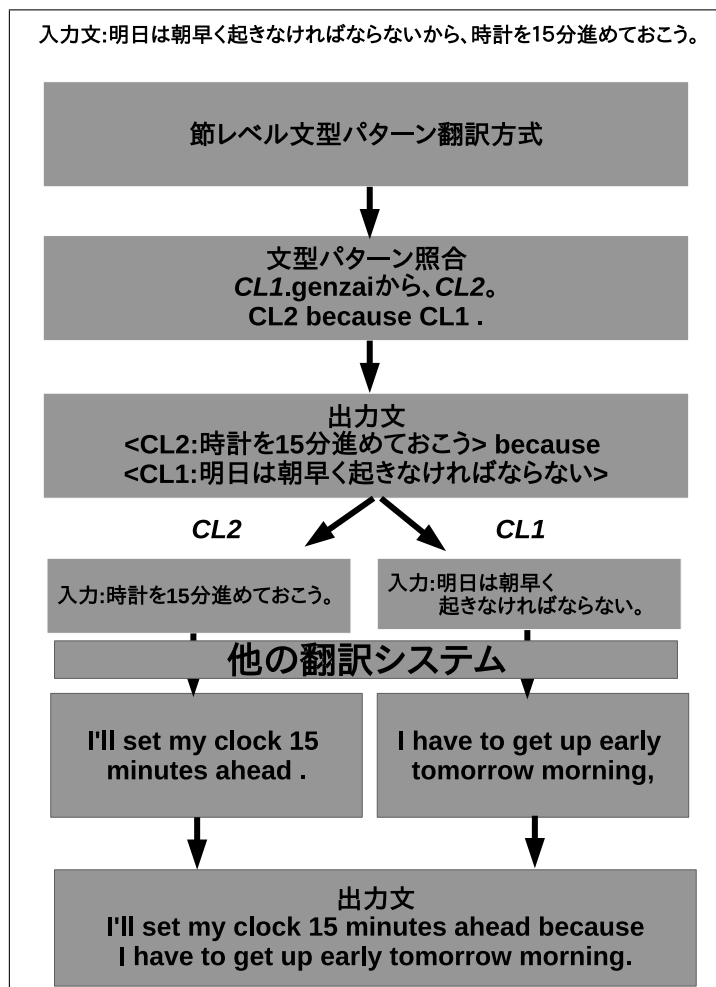


図 11: 他の翻訳システムを利用

方法5は文型パターン照合システムを改良することにより、適切な節レベル文型パターンを得ることができ、適切な対応の節変数を得ることができる。その結果、節変数の構造を節パターン辞書を用いて表現することができるようになると思う。表 7.3 に不適切な節レベル文型パターンが照合された例を示す。

表 7.3 では、すべての要素を1つの節変数として表現しているが、理想としては、“CL1で、CL2”のような節レベル文型パターンが照合に成功することが理想である。

表 7.3: 不適切な節レベル文型パターンが照合された例

入力文	彼女は大変な暑がり 夏は毎年軽井沢で避暑をする。
出力文 (提案手法)	, <cl4:彼女は大変な暑がり 夏は毎年軽井沢で避暑をする > .
参照文	Since she is very sensitive to the heat, she spends the summer at Karuizawa every year.
節レベル文型パターン (日)	CL4。
節レベル文型パターン (英)	#1[Needless to say] , CL4 .

7.2 翻訳精度について

表 6.3, 表 6.4 及び表 6.14 より自動評価及び人手評価において提案手法は単語レベル文型パターン翻訳方式及び句レベル文型パターン翻訳方式よりも性能が低いことが分かる。その理由の一つとして、文型パターン選択の誤りが挙げられる。表 7.4 に文型パターン選択を誤った例を示す。

表 7.4: 提案手法で翻訳精度が低かった例

入力文	炉がうなりシューっと音を立てた。
出力文 (提案手法)	furnaces is groan as patter whoosh to thump .
参照文	The furnace roared and hissed .
節レベル文型パターン (日)	CL1CL2.kako。
節レベル文型パターン (英)	CL1 as CL2 .
節パターン (CL1)	(日)N1が V2 (英)N1 @be V2
節パターン (CL2)	(日)ADV1 N2 を V3 (英)V3 #1[ADV1] to N2
参考情報 (句レベル)	
出力文 (句レベル)	furnaces growled and fizzled .
句レベル文型パターン (日)	NP1が VP2VP3.kako。
句レベル文型パターン (英)	NP1 VP2^past and VP3^past .

表 7.4 では、英語の節レベル文型パターン内に “as” が存在する。入力文では、動作の連続を表しているが “as” の用法として動作の連続を表す表現は存在しない。そのため、表 7.4 の英語の節レベル文型パターンを用いると翻訳精度が低下する。このように、文型パターンの選択を誤った例が他にも多く見られた。この問題点を解決するには、適切な文型パターンを用いる必要がある。現在、適切な文型パターンを用いるためにどのような情報が必要か未調査のため、今後調査する必要がある。

8 おわりに

本研究では、パターンマッチ文数の増加を目的として、節レベル文型パターン翻訳方式を提案し、実装した。また、節変数の構造を表現するための節パターン辞書の作成も行った。最後に、提案手法と単語レベル文型パターン翻訳方式及び、提案手法と句レベル文型パターン翻訳方式をパターンマッチ文数及び翻訳精度について翻訳実験を行った。

翻訳実験の結果より、提案手法は単語レベル文型パターン翻訳方式及び句レベル文型パターン翻訳方式に比べて、パターンマッチ文数が増加したことが確認できた。しかし、翻訳精度の低下も確認できた。

今後の課題として以下が挙げられる。

- 文型パターンを選択する精度についての詳しい調査

また、パターンマッチ文数の更なる増加の方法として以下の方法が挙げられる。

- 単語辞書、句パターン及び節パターンを人手で追加
- 節パターン辞書の作成に句レベル文型パターンを利用
- 単文文型パターン辞書 [7] を利用
- 他の翻訳システムを利用
- 文型パターン照合システムの改良

これらの調査などを行うことにより、提案手法のパターンマッチ文数の更なる増加及び翻訳精度の向上が期待できる。

謝辞

最後に、一年間に渡り、本研究のご指導をいただきました鳥取大学工学部知能情報工学科自然言語処理研究室の村上仁一准教授、村田真樹教授に深く感謝すると共に、厚く御礼申し上げます。そして、日常の議論を通じて多くの知識や示唆を頂いた同研究室の皆様に深謝いたします。また、参考にさせていただいた論文の著者の方々に対して、深く感謝申し上げます。

参考文献

- [1] 池原悟, 阿部さつき, 徳久雅人, 村上仁一. “非線形な表現構造に着目した重文と復文の日英文型パターン化”, 自然言語処理, Vol.11, No.3, pp.69-95, 2004.
- [2] 石上真理子, 徳久雅人, 村上仁一, 池原悟. “結合価パターンを用いた動詞句の翻訳可能性の調査”, 言語処理学会第11回年次大会, pp.364-367, 2005.
- [3] Jun Sakata, Masato Tokuhisa, Jin'ichi Murakami. “Phrase-Level Pattern-Based Machine Translation Based on Analogical Mapping Method”, 8th International Conference on NLP(japTAL2012), pp.28-39, Oct. 2012.
- [4] 池原悟 “非線形言語モデルによる自然言語処理 基礎と応用”, 岩波書店, 2009.
- [5] 金澤佑哉, 徳久雅人, 村上仁一, 池原悟 “記号と関数による文型パターン汎化効果の総合評価”, 言語処理学会第13回年次大会発表論文集, pp.51-54, 2007.
- [6] 徳久雅人, 村上仁一, 池原悟. “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, Vol.2006, No.124, pp.9-16, 2006.
- [7] 福安寿規 “非線形言語モデルに基づく単文文型パターン辞書の改良”, 鳥取大学卒業論文, 2008.