

概要

日英機械翻訳のための 22 万件の重文・複文文型パターン辞書が作成されている [1]。この文型パターン辞書は、様々なドキュメントから収集した 15 万件の日英対訳コーパスを元に作成されている。言語の多様性に対応するために、文型パターン辞書にはさまざまな表現が収録されているが、問題として、収録されている文型パターンには表現のカバー範囲が重複していることや使用される可能性が低い表現を持つものが多く含まれていることが挙げられる。また、人が文型パターンを用いる場合は、紙媒体での辞書も必要になると考えられるが、22 万件の重文・複文文型パターンを紙媒体で利用することは現実的ではない。そのため、規模を小さくした上で、よく使われる表現を中心に出来るだけ多様な表現を収録した文型パターン辞書が求められている。

そこで、本研究では、使用される可能性の低い文型パターンやカバー範囲が重複する文型パターンに着目して、精度を低下させずに自動的に文型パターン数の削減を行った。精度を維持するために、文型パターンの適合頻度、包含関係および評価用試験文集での照合結果を利用した。さらに、過剰に汎化された日本語文型パターンの削除も行った。

結果、精度を維持しつつ、元の文型パターン辞書の 14.65% (17,973 件) に縮退することが出来た。また、縮退後の文型パターン辞書を用いて、人手による翻訳調査を行った。その結果、日本語文型パターンの包含関係を用いると、正解率が大きく低下するが、意味的な被覆率から見ると影響は小さいと考えられる。さらに、過剰に汎化された日本語文型パターンを削除することで、正解率の向上が見込める。

目次

1	はじめに	1
2	文型パターン辞書と文型パターンパーサ	2
2.1	文型パターン辞書	2
2.1.1	文型パターン記述言語	3
2.2	文型パターン辞書の問題点	7
2.2.1	表現のカバー範囲の重複	7
2.2.2	使用される可能性が低い表現	9
2.2.3	過剰汎化文型パターン	10
2.3	文型パターンパーサ	11
3	文型パターン辞書の削減	13
3.1	包含関係 [7]	13
3.1.1	文型パターン間の包含関係	13
3.1.2	文型パターン要素間の包含関係	13
3.2	評価方法	14
3.2.1	クロスバリデーションテスト	14
3.2.2	再現率	14
3.3	適合頻度	15
3.3.1	文型パターン辞書の適合頻度	15
3.3.2	適合頻度と文型パターンの傾向	19
3.4	提案手法	20
3.4.1	適合頻度 0 の文型パターンの削除 (手順 1)	21
3.4.2	別の評価用試験文集の利用した文型パターンの追加 (手順 2)	21
3.4.3	日本語文型パターンの包含関係による縮退 (手順 3)	22
3.4.4	過剰に汎化された文型パターンを削除 (手順 4)	22
3.5	実験結果	22
3.5.1	実験条件	22
3.5.2	適合頻度 0 の文型パターンの削除 (手順 1)	22
3.5.3	評価用試験文集の照合結果 (手順 2)	23

3.5.4	包含関係による縮退（手順3）	23
3.5.5	過剰汎化文型パターンの削除（手順4）	27
3.5.6	辞書の縮退結果	27
4	人手による翻訳調査	28
4.1	評価対象	28
4.2	評価基準	28
4.2.1	評価 A の例	28
4.2.2	評価 B の例	30
4.2.3	評価 C の例	31
4.2.4	評価 D の例	32
4.2.5	評価 F の例	33
4.3	評価手順	35
4.4	正解率	36
4.5	評価結果	36
5	考察	38
5.1	適合頻度 0 の文型パターン	38
5.2	適合頻度 1 の文型パターン	38
5.3	包含関係による縮退の与える影響	40
5.4	過剰汎化文型パターンの与える影響	42
5.5	適合頻度とパターン適合率と翻訳評価の関係	44
5.5.1	パターン適合率 (C)	44
5.5.2	適合頻度とパターン適合率の関係	45
5.5.3	高頻度文型パターンの翻訳評価	45
5.5.4	特異点にある文型パターンの翻訳評価	46
5.5.5	集中部の文型パターンの翻訳評価	47
6	おわりに	48
7	謝辞	48

図目次

1	適合頻度とパターン適合率の平均の関係	45
---	------------------------------	----

表目次

1	変数一覧	3
2	記号一覧	5
3	離散記号の適合条件	6
4	文型パターン要素間の包含関係 (一部)	14
5	適合頻度の度数分布表	16
6	包含関係判定結果 (辞書 D2)	24
7	縮退結果	27
8	人手による翻訳評価の結果 (クロスバリデーションテスト: 50 文)	37
9	人手による翻訳評価の結果 (オープンテスト: 100 文)	37
10	適合頻度 1 のパターンが適合した文が, 適合するパターン数の分布	38
11	適合頻度上位 8 件の翻訳評価結果	46
12	特異点 8 件の翻訳評価結果	46
13	集中部 19 件の翻訳評価結果	47

1 はじめに

現在，日英機械翻訳のための 22 万件の重文・複文文型パターン辞書がある [1]．この文型パターン辞書は，15 万件の日英対訳コーパスを元に作成されている．言語の多様性に対応するために，文型パターン辞書にはさまざまな表現が収録されている．しかし，問題として，収録されている文型パターンには表現のカバー範囲が重複していることや使用される可能性が低い表現を持つものが多く含まれていることが挙げられる．

また，人が利用する場合，紙媒体での辞書も必要な時がある．現在出版されている結合価パターン辞書 [2] は，1.4 万件の結合価パターンを収録するために 830 ページを用いている．重文・複文文型パターンは結合価パターンよりも情報量が多いことも考慮すると，この文型パターン辞書を紙媒体に出力するためには，1.5 万ページ以上が必要と考えられ，現実的では無い．そのため，規模を小さくした上で，よく使われる表現を中心に出来るだけ多様な表現を収録した文型パターン辞書が求められている．

古くから文型パターン翻訳方式のための文型パターン辞書の開発は行われているが，文型パターン辞書の規模を小さくするという試みは行われていない．それは，文型パターン辞書の開発には膨大なコストがかかり，文型パターン数が 200～300 件程度にとどまっていたためである．

そこで，本研究では，使用される可能性の低い文型パターンやカバー範囲が重複する文型パターンに着目し，精度を低下させないことを目的として，自動的に文型パターン数の削減を行う．

本論文の構成は以下の通りである．第 2 章では文型パターン辞書について説明する．第 3 章では文型パターン辞書の削減方法と削減結果について説明する．第 4 章では人手による翻訳調査について説明する．第 5 章では考察を行い，最後の第 6 章で本研究の結論と今後の課題を述べる．

2 文型パターン辞書と文型パターンパーサ

2.1 文型パターン辞書

本研究では、文型パターン翻訳方式のための重文・複文文型パターン辞書を扱う。この文型パターン辞書には、22 万件の文型パターンが収録されている。文型パターン辞書は、15 万件の日英対訳コーパスから作成されている。日英の各対訳文に含まれる線形要素をそれぞれの値域を持つ変数に置き換え、各種関数・記号を付与することで作成されている。この辞書には、単語レベル(12 万件)、句レベル(8 万件)、節レベル(2 万件)の3つのレベルでそれぞれ汎化された文型パターンが収録されている。以下に例を示す。

- 原文

日本語文： 大阪に着くまでの間、ずっと本を読み続けた。

英語文： I read until I reached Osaka.

- 単語レベル

日本語文型パターン

$/y </tk N1 \text{ は} > /tcfk N2 \text{ に} /cf V3^{\wedge}rentai \text{ (まで | 迄) の! 間、} <! N4 \text{ は} > \#5[!$
 $\text{ずっと}] /tck \text{ 本を} /cf V6 \text{ (つづけ | 続け) た。}$

英語文型パターン

$<I|N4 > V6^{\wedge}past \text{ until } <I|N1 > V3^{\wedge}past N2.$

- 句レベル

日本語文型パターン

$/y </tk N1 \text{ は} > ! VP2^{\wedge}rentai \text{ (まで | 迄) の! 間、} <! N3 \text{ は} > \#4[!$
 $\text{ずっと}] /tck \text{ 本を} /cf V5 \text{ (つづけ | 続け) た。}$

英語文型パターン

$<I|N3 > V5^{\wedge}past \text{ until } <I|N1 > VP2^{\wedge}past.$

- 節レベル

日本語文型パターン

$/y /tcfk CL2^{\wedge}rentai \text{ (まで | 迄) の! 間、} ! CL4.kako .$

英語文型パターン

$CL4^{\wedge}past \text{ until } CL2^{\wedge}past.$

本研究では、単語レベルの文型パターンを対象として用いる。

日英文型パターン翻訳方式では，入力文に適合する日本語文型パターンを検索し，対応する英語文型パターンを用いて訳文を生成する．

2.1.1 文型パターン記述言語

本研究で扱う文型パターンは，字面，変数，関数および記号の4つの要素から構成される．[3]

変数

変数は，線形要素の記述に用い，種類に応じて変域が定義されている．(表1)例えば，名詞変数 N は，全ての名詞単語が適合し，時詞変数 $TIME$ は名詞のうち時詞のみが適合する．

文型パターンでは，変数に番号を付与して用いる．これにより，日英対訳文型パターンの変数の対応関係を識別することが出来る．

表1 変数一覧

変数名	意味	変数名	意味
単語レベル		句レベル	
N	名詞	NP	名詞句
ND	「する」に先行する名詞	VP	動詞句
NUM	数詞	AJP	形容詞句
$TIME$	時詞	$AJVP$	形容動詞句
V	動詞	$ADVP$	副詞句
AJ	形容詞	節レベル	
AJV	形容動詞	CL	節
ADV	副詞		
REN	連体詞		
GEN	限定詞		

関数

関数は，表記上の揺らぎを吸収するために用い，様相関数，語形関数，品詞変換関数がある．

様相関数は，変数に付属した助動詞相当表現（相・態・様相・時制）を指定する．使用

例を以下に示す。

.kako

過去の時制を指定する。「行った」は「*V1.kako*」に適合する。このとき、「行っ」が *V1* に「た」が *.kako* にそれぞれ対応する。

.darou

推量・判断の様相を指定する。「行くだろう」は「*V1.darou*」に適合する。このとき、「行く」が *V1* に「だろう」が *.darou* にそれぞれ対応する。

.sase

使役態を指定する。「食べさせる」は「*V1.sase*」に適合する。このとき、「食べ」が *V1* に「させる」が *.sase* にそれぞれ対応する。

語形関数は、変数に適合する値の形式を指定するために用いる。日本語文型パターンでは活用形の指定し、英語文型パターンでは単語の変形を指定する、使用例を以下に示す。

^rentai

日本語文型パターンにおいて、連体形を指定する。「走る人」は「*V1^rentai* 人」に適合する。このとき、「走る」は動詞の連体形である。

^meirei

日本語文型パターンにおいて、命令形を指定する。「走れ」は動詞の命令形なので、「*V1^meirei*」に適合する。

^past

英語文型パターンにおいて、過去形への変形を指定する。*V1^past* と記述がある場合、動詞変数 *V1* を過去形に変形する。

^st

英語文型パターンにおいて、最上級への変形を指定する。*AJ1^st* と記述がある場合、形容詞変数 *AJ1* を最上級に変形する。

品詞変換関数は、品詞の変換を指定するために用いる。英語文型パターンでのみ用いられる。使用例を以下に示す。

N(AJ1)

形容詞変数 *AJ1* に適合した内容を、名詞に変換する。

V(ADV1)

副詞変数 *ADV1* に適合した内容を、動詞に変換する。

記号

記号は、文型パターン要素の適合の仕方について、任意化、選択、順序変更および記憶という制御を行う。(表2、表3)以下に使用例を示す。

任意記号の例

$N1$ は [本当に] $V2$ 。

この例の場合、「本当に」の部分があっても、無くてもよいことを示す。

順序任意要素指定記号の例

{ $N1$ は, $N2$ と } $V3$ 。

この例の場合、「 $N1$ は $N2$ と」か「 $N2$ と $N1$ は」のどちらからであれば良いことを示す。

離散記号の例

$/cf$

この例の場合、離散記号の適合条件 c は格要素、 f は連用修飾句(副詞句)を表すので、格要素、連用修飾句、または格要素+連用修飾句が適合することを示す。また、離散記号には何も適合しないことも許される。

表2 記号一覧

記号名	表記	意味
離散記号	$/...$	文型に無関係な要素(適合条件については表3参照)
要素選択記号	$(... ...)$	いずれかの要素列と適合
任意記号	$[...]$	文型選択上、任意の要素
補完要素記号	$<...>$	文型パターン作成時の原文に存在しない主語を補った要素
順序任意要素 指定記号	$\{...,...\}$	順序入れ替え可能な範囲
位置変更可能 要素指定記号	$\$n^{\{...\}}$ $\$n$	指定位置に入れ替え可能
文節境界記号	!	文節の境界と適合
記憶記号	$\#n$	適合内容を記憶
標準形表記記号	'...'	字面部分の標準形を指定する

表 3 離散記号の適合条件

表記	適合条件
y	連用節
t	連体節
c	格要素
f	連用修飾句（副詞句）
k	連体修飾句（形容詞，形容動詞連体形，連体詞）

2.2 文型パターン辞書の問題点

2.2.1 表現のカバー範囲の重複

文型パターン辞書において，日本語文型パターン同士でカバー範囲が重複しているという問題がある．以下に例を示す．

重複例 1

- 対訳文型パターン A

日本語文：

彼女は幸福な結婚をしている。

日本語文型パターン：

/ytcfk N1 は /cf AJV2^{rentai} /f ND3 を /cf している。

英語文：

She is happily married.

英語文型パターン：

N1 be ADV(AJV2) V(ND3|V3)^{past}.

- 対訳文型パターン B

日本語文：

私は気楽な独身生活をしている。

日本語文型パターン：

/ytcfk N1 は /cf 気楽な /f N2 ND3 を /cf している。

英語文：

I live in single blessedness.

英語文型パターン：

N1 V(ND3|V3) in AJ(N2) blessedness.

対訳パターン A，B の日本語文型パターンは，変数 *AJV2* と字面「気楽な」のみが異なっている．変数 *AJV2* は形容動詞を表し，「気楽な」の品詞も形容動詞であるため，文型パターン A のカバー範囲に，文型パターン B のカバー範囲が含まれている．

重複例 2

- 対訳文型パターン C

日本語文：

彼は結婚して一家を構えた。

日本語文型パターン：

$/y \$1^{\wedge}/\{tcfk N1 \text{ は } \} /cf (V2|ND2 \text{ をし}) (て | で) \$1 /ytck N3 \text{ を } /cf V4.kako。$

英語文：

He got married and set up house.

英語文型パターン：

$N1 V(V2|ND2)^{\wedge}past \text{ and } V4^{\wedge}past N3.$

● 対訳文型パターンD

日本語文：

彼女は仕返しして恨みを晴らした。

日本語文型パターン：

$/y \$1^{\wedge}/\{tcfk N1 \text{ は } \} /cf (V2|ND2 \text{ をし}) (て | で) \$1 /ytck \text{ 恨みを } /cf \text{ 晴らした。}$

英語文：

She settled an old score and got revenge.

英語文型パターン：

$N1 \text{ settled an old score and } V(V2|ND2)^{\wedge}past.$

対訳パターンC，Dの日本語文型パターン中は、「 $N3$ を $/cf V4.kako$ 」と「恨みを $/cf$ 晴らした」の部分で異なっている。「恨み」は名詞、「晴らした」は動詞+過去を表す助動詞で、それぞれ名詞変数 $N3$ と動詞変数+様相関数の $V4.kako$ に適合する。よって、文型パターンCのカバー範囲に、文型パターンDのカバー範囲が含まれている。

重複例3

● 対訳文型パターンE

日本語文：

岩田君はがんじょうな体をしている。

日本語文型パターン：

$/ytcfk N1 \text{ は } /cf AJV2^{\wedge}rentai /f N3 \text{ を } /cf V4.teiru。$

英語文：

Iwata has a good physique.

英語文型パターン：

$N1 V4 AJ2 N3.$

● 対訳文型パターンF

日本語文：

彼女は質素な服装をしている。

日本語文型パターン：

/ytcfk N1 は /cf AJV2^rentai /f 服装を /cf している。

英語文：

She is plainly dressed.

英語文型パターン：

N1 be ADV(AJV2) dressed.

対訳パターンE, Fの日本語文型パターン中は、「N3を /cf V4.teiru」と「服装を /cf している」の部分が異なっている。「服装」は名詞、「している」は動詞+接続助詞+補助動詞で、それぞれ名詞変数N3と動詞変数+様相関数のV4.teiruに適合する。よって、文型パターンEのカバー範囲に、文型パターンFのカバー範囲が含まれている。

これらの対訳文型パターン例のうち、日本語文型パターンについて着目すると、対訳文型パターンBは対訳文型パターンAで、対訳文型パターンDは対訳文型パターンCで、対訳文型パターンFは対訳文型パターンEで、それぞれ代替できるので削除可能と考えられる。

2.2.2 使用される可能性が低い表現

文型パターン辞書には、さまざまな表現に対応するための文型パターンが多く収録されている。しかし、日常的にはあまり使われない表現に対応する文型パターンが多く含まれている。以下に例を示す。

- 文型パターン例1

日本語文：

頭をそって丸坊主になった。

日本語文型パターン：

/y </tk N1 は> /tcfk (頭 | おつむ) を /cf (そっ | 剃っ) て /ytck 丸坊主に /cf なった。

英語文：

I shaved all my hair off.

英語文型パターン：

<I|N1> shaved all <my|N1^pron^poss> hair off.

「頭をそって丸坊主になった」に相当する部分がほとんど字面になっているが、日常的にはあまり使われないと考えられる。

- 文型パターン例2

日本語文：

泥棒が入って現金を根こそぎ持ってゆかれた。

日本語文型パターン：

/ytcfk N1 が /cf(入っ | はいっ | 入いっ) て </ytck N2 は> /tcfk 現金を /tcfk (根こそぎ | ねこそぎ | 根刮) /f V3.teiku.reru.kako。

英語文：

A burglar broke into my house and took away all the money I had in cash.

英語文型パターン：

N1 broke into <my|N2^{poss}> house and V3^{past} all the money <I|N2^{pron}> had in cash.

「入って現金を根こそぎ」に相当する部分がほとんど字面になっているが、日常的にはあまり使われないと考えられる。

● 文型パターン例 3

日本語文：

ああなると誰もあわれむ者も無い。

日本語文型パターン：

/ytcfk ああ /f(成る | なる) と /tck(誰 | だれ | どなた) も /cf V1^{rentai} 者も /yf(無い | ない)。

英語文：

Nobody takes pity on a man in his condition.

英語文型パターン：

Nobody V1 a man in his condition.

「あわれむ」に相当する部分以外が、全て字面になっている。しかし、「ああなると誰も～者もない」という表現は日常的にはあまり使われないと考えられる。

2.2.3 過剰汎化文型パターン

以下で示すような、過剰に汎化された文型パターンが存在する問題がある。

日本語文：

ニッコリ笑って承知した。

日本語文型パターン：

/y </tk N1 は> #2[/cf V3 て] /ycf(V4.kako|ND4 をした)。

英語文：

She smiled her consent.

英語文型パターン：

<She|N1> V3^{past} <her|N1> N(V4|ND4).

このような日本語文型パターンは，任意記号内に動詞の1つが記述されているため，単文にも適合する．また，日本語文型パターン中の離散記号 (/y) は，連用節に適合する．そのため，多くの入力文が，これらの日本語文型パターンに「連用節+単文」の形で適合したと考えられる．

しかし，本研究で扱う文型パターン辞書は本来，重文・複文を対象としている．したがって，これらはパターン作成中に過剰に汎化されたと言わざるを得ない．

このような，過剰汎化された文型パターンを削除することで，文型パターン辞書の精度が向上できるのではないかと考えられる．

2.3 文型パターンパーサ

文型パターンパーサ(以下パーサ)[4]は，日本語文と日本語文型パターンとの照合を行い，日本語文に適合する日本語文型パターンを出力するプログラムである．照合方式は，ATN(Augmented Transition Network)[5]をベースとしている．

以下に，パーサが出力した文型パターン例を示す．ここで，バインド値とは，日本語文が文型パターンに適合したとき，文型パターン中の変数が適合した形態素を示すものである．

入力日本語文： 損害を最小限に食い止めることが大事だ。

- 適合例 1

日本語文型パターン：

/ytcfk N1 を /tcfk N2 に /cf (V3^{rentai}|ND3 をする) ! ことが /cf AJV4 #5(.gen-zai|.kako)。

(原文：適材を適所に配することが重要だ。)

バインド値：

N1=損害，N2 = 最小限，V3=食い止める，AJV4=大事だ

英語文型パターン：

It 'is' #5(^present|^past) AJ4 to V(V3|ND3)^{past} N1 in N2.

(原文：It is important to put the right person in the right place.)

- 適合例 2

日本語文型パターン：

/y </tk N1 は> /tcfk N2 を /cf V3^{rentai} ! ことが /cf AJV4 #5(.genzai|.kako)。

(原文：経験を積み重ねることが大切だ。)

バインド値：

N2=損害, /cf = 最小限に, V3=食い止める, AJV4=大事だ

英語文型パターン：

It 'is' #5(^{present}^{past}) AJ4 for <you|N1> to V3^{base} N2.

(原文：It is important for you to pile up experiences.)

- 適合例 3

日本語文型パターン：

/y </tk N1 は> /tcfk N2 に /cf V3^{rentai} ! ことが /cf AJV4 #5(.genzai|.kako)。

(原文：互いに助け合うことがたいせつだ。)

バインド値：

/tcfk=損害を, N2 = 最小限, V3=食い止める, AJV4=大事だ

英語文型パターン：

It 'is' #5(^{present}^{past}) AJ4 for <you|N1> to V3^{base} N2 .

(原文：It is important for you to help each other.)

3 文型パターン辞書の削減

3.1 包含関係 [7]

3.1.1 文型パターン間の包含関係

文型パターン β に適合する入力文の全てが、文型パターン α にも適合するとき、文型パターン β は文型パターン α に包含されると定義する。また、文型パターン α を上位の文型パターン、文型パターン β を下位の文型パターンと呼ぶ。両者の関係を $\alpha \supseteq \beta$ と表記する。

以下に包含関係にある文型パターンの例を示す。

α : $N1$ は $V2$ 。

β : 私 は $V1$ 。

変数 $V1$ は動詞を表すため、文型パターン β は「私は見る。」や「私は行く。」など、「私は(動詞)。」型の文に適合する。また、変数 $N1$ は名詞を表すため、文型パターン α は「彼は見る。」や「地球は回る。」など、「(名詞)は(動詞)。」型の文に適合する。「私」は名詞であることから、文型パターン α も「私は(動詞)。」型の文に適合する。

したがって、文型パターン β に適合する全ての日本語文は、文型パターン α にも適合可能である。

3.1.2 文型パターン要素間の包含関係

前節の定義に従って、文型パターン間の包含関係を判定するためには、全ての入力文に対して適合の可否を調査する必要がある。しかし、全ての入力文に対して調査を行うことは不可能である。そこで、文型パターン自身が適合可能な入力文の領域を表している事に着目して、文型パターンが別の文型パターンに適合するかどうかを調査することで、包含関係を判定する。

文型パターン間の包含関係を考える場合、文型パターンを構成する要素(変数、関数、記号、字面)の包含関係を定義する必要がある。そこで、各要素の定義 [3] に基づいて包含関係を定義する。表 4 に定義した要素間の包含関係の一部を示す。また、付録に全ての定義を示す。

表 4 文型パターン要素間の包含関係 (一部)

上位の要素	下位の要素
<i>N</i> (名詞)	<i>NUM</i> (数詞), <i>TIME</i> (時詞), <i>ND</i> (用言性名詞)
<i>NP</i> (名詞句)	<i>N</i> , <i>N</i> の下位要素
<i>VP</i> (動詞句)	<i>V</i> (動詞)

3.2 評価方法

3.2.1 クロスバリデーションテスト

クロスバリデーションテスト (cross validation test) とは、少数のコーパスで十分な精度のオープンテストを行う手法である。

本研究では、クロスバリデーションテストの入力文として、文型パターン作成に用いた例文を用いる。この入力文と文型パターン辞書をパーサを用いて照合する。この時、各入力文は、その入力文から作成された文型パターン (以下、自己文型パターン) に必ず適合するので無視し、自己文型パターン以外の適合文型パターンを照合結果として用いる。

例

例文 *S1*, *S2*, *S3* があり、それらから文型パターン *X1*, *X2*, *X3* をそれぞれ作成したとする。

例文 *S1* が文型パターン *X1* と *X3* に適合した場合、クロスバリデーションテストでは文型パターン *X3* のみが適合したとする。また、例文 *S2* が文型パターン *X2* にのみ適合した場合、クロスバリデーションテストでは適合文型パターン無しとする。

3.2.2 再現率

再現率 (以下 *R1*) は、文型パターン辞書の精度を評価するためのパラメータの1つで、以下の式で求める。

$$R1 = \frac{\text{自己文型パターン以外の文型パターンが1つ以上適合した文数}}{\text{入力文数}}$$

例

入力文 100 文中，自己文型パターン以外の文型パターンが 1 つ以上適合した入力文が 68 文であった場合， $R1$ は以下の値となる．

$$R1 = \frac{68}{100} = 0.68 (68\%)$$

3.3 適合頻度

適合頻度とは，文型パターンが使用される可能性を表す指標である．

本研究では，以下の手順で求める．

1. 文型パターン辞書作成用の用例と文型パターン辞書を用いて，クロスバリデーションテストを行う
2. 文型パターンそれぞれについて，適合した入力文の数を調査し，その値を適合頻度とする．

例

文型パターン $X1$ が，入力文のうち， $S2, S6, S7$ に適合した場合，文型パターン $X1$ の適合頻度は 3 となる．

3.3.1 文型パターン辞書の適合頻度

予備調査として，本研究で扱う文型パターン辞書の適合頻度の分布を調査した．その結果を表 5 に示す．

各適合頻度ごとに文型パターン例を以下に示す．

- 適合頻度 0 の例 (1)

日本語文：

二人は相性が悪いということで離婚した。

日本語文型パターン：

$/y \$1^{\wedge}/\{cfk N1 \text{ は } \}/\{cfk \text{ 相性が } /cf (\text{悪い} | \text{わるい} | \text{悪い} | \text{悪い}) \text{ という！ことで } \$1 /cf (V2.kako|ND2 \text{ をした}) .$

英語文：

They divorced each other on the grounds of incompatibility.

英語文型パターン：

表 5 適合頻度の度数分布表

適合頻度	文型パターン数 / 割合	
0 文	97,346	79.324%
1 文	8,905	7.256%
2 文以上 10 文未満	9,124	7.435%
10 文以上 100 文未満	5,125	4.176%
100 文以上 1,000 文未満	1,684	1.372%
1,000 文以上 10,000 文未満	532	0.434%
10,000 文以上 100,000 文未満	3	0.002%
計	122,719	-

$N1 V(V2|ND2)^{past}$ each other on the grounds of incompatibility.

● 適合頻度 0 の例 (2)

日本語文：

風を受けてたこが空高く上がって行った。

日本語文型パターン：

$/ytcfk N1$ を！ 受けて $/ytk N2$ が $/cf$ 空高く $/yf V3$ (て | で) $/ycf$ (行っ | ゆっ |
いっ | 往っ) た。

英語文：

Catching the wind, the kite continued to climb higher and higher into the sky.

英語文型パターン：

Catching $N1$, $N2$ continued to $V3^{base}$ higher and higher into the sky.

● 適合頻度 0 の例 (3)

日本語文：

他人の前で若い男女がいちゃいちゃ振る舞うなど昔の日本では考えられな
かった。

日本語文型パターン：

$/y \$1 /tcfk N1$ の $/k N2$ で #3 $[/cf AJ4^{rentai}] /f N5$ が $/cf$ いちゃいちゃ $/f$ 振る舞
うなど $\$1^{/ytk \#6[TIME7 \text{ の}]} /k N8$ では } $/cf V9.rareru.hitei.kako$ 。

英語文：

In the past, Japanese young men and women would never have thought of hug-
ging and kissing in front of other people.

英語文型パターン：

#6[In N7], AJ(N8) #3[AJ4] N5 would never V9^{prp} of hugging and kissing in N2 of N1.

● 適合頻度 1 の例

日本語文：

いったん引き受けた仕事はあくまでもやりぬきなさい。

日本語文型パターン：

/y </tk N1 は> /cf ADV2 /f V3.kako^{rentai} /f N4 は </tcfk N5 は> /tcfk あくまでも /cf (V6^{meirei}|V6.meireigo)。

英語文：

Once you have undertaken a task, you should see it through.

英語文型パターン：

ADV2 <you|N1> V3^{prp} N4, <you|N5> should V6.

● 2 文以上 10 文末満の例 (適合頻度 5)

日本語文：

ここできみに会おうとは夢にも思わなかった。

日本語文型パターン：

/y #1[/cf ADV2] /tck N3 に /cf (会お | あお | 逢お | 會お) うとは </y N4 は> /cf 夢にも /f V5.hitei.kako。

英語文：

I never thought to run into you here.

英語文型パターン：

<I|N4> never V5^{past} to run into N3 #1[ADV2].

● 10 文以上 100 文末満の例 (適合頻度 59)

日本語文：

あの役人は地位を悪用して金をもうけた。

日本語文型パターン：

/y \$1{/cf GEN1 /k N2 は } /tcfk N3 を /cf V4 (て | で) \$1 /ytck N5 を /cf V6.kako

英語文：

That government official earned money by abusing his position.

英語文型パターン：

AJ1 N2 V6^{past} N5 by V4^{grn} N2^{pron}poss N3.

● 100 文以上 1,000 文末満 (適合頻度 458)

日本語文：

彼らは汗をだらだら流して働いている。

日本語文型パターン：

/y \$1^{/tcfk N1 は } /tcfk N2 を #3[/cf だらだら] /f V4 (て | で) \$1 /ycf V5.teiru
#6(.genzai|.kako)。

英語文：

They are streaming with sweat as they work.

英語文型パターン：

N1 V4^{prog} with N2 as they V5#6(^{present}|^{past}).

- 1,000 文以上 10,000 文未満 (適合頻度 3,227)

日本語文：

彼はディフェンスをすり抜けてシュートした。

日本語文型パターン：

/y \$1^{/tcfk N1 は } /tcfk N2 を /cf V3 (て | で) \$1 /ycf (V4.kako|ND4 をした)。

英語文：

He dodged through the defense and shot.

英語文型パターン：

N1 V3^{past} through N2 and V(V4|ND4)^{past}.

- 10,000 文以上 100,000 文未満 (適合頻度 10,077)

日本語文：

その申し出を蔑むようにはねつけた。

日本語文型パターン：

/y </tk N1 は> #2[/cf GEN3] /k N4 を #5[/cf (蔑む | 貶む) (ように | 様に)] /yf
V6.kako 。

英語文：

She scornfully rejected the proposal.

英語文型パターン：

<She|N1> #5[scornfully] V6^{past} #2[AJ3] N4.

3.3.2 適合頻度と文型パターンの傾向

低頻度文型パターン

全体の約 8 割を占める適合頻度 0 の文型パターンは、字面を多く含む傾向があった。また、字面の表現は日常的にはあまり使われない可能性の高い傾向がある。なお、適合頻度 0 の文型パターンを削除しても、クロスバリデーションテストの再現率には全く影響を与えない。以上より、これらの文型パターンを削除することで、文型パターン数の大幅な削減が期待できる。

高頻度文型パターン

適合頻度が 10,000 以上の文型パターン 3 件を次に示す。

- AN148565-00 (適合頻度：24,961)

日本語文：

ニッコリ笑って承知した。

日本語文型パターン：

/y </tk N1 は> #2[/cf V3 て] /ycf (V4.kako|ND4 をした)。

英語文：

She smiled her consent.

英語文型パターン：

<She|N1> V3^{past} <her|N1> N(V4|ND4).

- WJAF022102-00 (適合頻度：18,246)

日本語文：

朝、旗を上げ、夕方下ろします。

日本語文型パターン：

/y </tk N1 は> #2[/tcfk TIME3、] / #4[N5 を /cf N6、] #7[/ytk TIME8] /f V9 #10[.masu]。

英語文：

We raise the flag in the morning and lower it in the evening.

英語文型パターン：

<We|N1> #4[V(N6) N5] #2[in N3] and V9 N5 #7[in N8].

- AQ205967-00 (適合頻度：10,077)

日本語文：

その申し出を蔑むようにはねつけた。

日本語文型パターン：

/y </tk N1 は> #2[/cf GEN3] /k N4 を #5[/cf (蔑む | 貶む) (ように | 様に)] /yf V6.kako。

英語文：

She scornfully rejected the proposal.

英語文型パターン：

<She|N1> #5[scornfully] V6^{past} #2[AJ3] N4.

これらの文型パターンは、単文にも適合する過剰に汎化された文型パターンである。数多くの日本語文に適合するため、適合頻度が高くなっていると考えられる。しかし、このような文型パターンが翻訳に利用出来る可能性は低い。そのため、文型パターン辞書の精度に悪影響を及ぼしている可能性がある。

3.4 提案手法

本研究で提案する文型パターンの削減方法の基本的な考え方を次に示す。

- 文型パターン辞書の精度を維持
- 適合頻度の低い文型パターンは影響が小さいため削除
- 適合頻度の高い文型パターンは悪影響があるため削除
- カバー範囲が重複する文型パターンを削除するために、包含関係を利用
- 文型パターン辞書の作成に用いた用例集とは別の用例集を利用することで、表現の多様性を確保

以上の考え方に基づいて、文型パターンの削減を行う。以下に手順の概要を示す。

1. クロスバリデーションテストでの適合頻度が0の文型パターンをオリジナルの文型パターン辞書(D0)から削除し、辞書D1を作成
2. 上記で削除した文型パターンのうち、別の評価用試験文集に適合するものを辞書D1に追加し、辞書D2を作成
3. 辞書D2において、包含関係を持つ文型パターンに着目して、辞書D3を作成
4. 辞書D3において、過剰に汎化された文型パターン削除して、縮退版文型パターン辞書D4を作成

各手順の具体的な内容を以下に示す。

3.4.1 適合頻度 0 の文型パターンの削除 (手順 1)

適合頻度 0 の文型パターンは、クロスバリデーションテストでの再現率 (R1) に全く影響を与えない。

そこで、文型パターン辞書 D0 から適合頻度 0 の文型パターンを削除し、文型パターン辞書 D1 を作成する。

3.4.2 別の評価用試験文集の利用した文型パターンの追加 (手順 2)

前の手順で、適合頻度 0 の文型パターンを削除した。しかし、適合頻度を求めるために利用した用例は、文型パターン辞書作成時に用いたものである。そのため、クロスバリデーションテストを行っているが、厳密なオープンデータとは言えない。そこで、文型パターン辞書作成時の用例と全く関係のない、別の評価用試験文集を用いることで、表現の多様性を確保する。手順 1 で削除した文型パターンのうち、この評価用試験文集に適合するものは必要と考えられる。そこで、このような文型パターンを文型パターン辞書 D1 に追加し、文型パターン辞書 D2 を作成する。

評価用試験文集

本手順で用いる評価用試験文集は、日本語独特の言い回しを持つ用例 [6] を集めたもので、3,951 文が収録されている。この評価用試験文集の用例は、文型パターン作成に用いた日英対訳コーパスには含まれていない。そこで、この用例集からランダムに抽出した 100 文を、意味的な精度を評価するためのオープンデータとし、残りの 3,851 文をこの手順で用いるテストデータとする。

以下に、用例の例を示す。

- 用例 1

日本語文：

熱を出した子どもをつききりで看病した。

英語文：

She paid constant attention to nursing the feverish child.

- 用例 2

日本語文：

この映画はみるたびに、涙が出てしかたがない。

英語文：

Every time I watch this movie, I cannot help crying.

- 用例 3

日本語文：

背が高くなければファッションモデルにはなれない。

英語文：

You cannot be a fashion model if you are not tall enough.

3.4.3 日本語文型パターンの包含関係による縮退（手順 3）

カバー範囲が重複する日本語文型パターンを削除するために、包含関係を用いる。包含関係の下位の文型パターンは、上位の文型パターンで代替可能なので、削除しても再現率に影響を与えない。

本手順では、文型パターン辞書 D2 において包含関係を調査し、下位の文型パターンを削除した文型パターン辞書 D3 を作成する。

3.4.4 過剰に汎化された文型パターンを削除（手順 4）

過剰に汎化された文型パターンは、辞書の精度に悪影響を及ぼすと考えられる。このような文型パターンは適合頻度が高い傾向がみられる。そこで、文型パターン辞書 D3 から、適合頻度 10,000 以上の文型パターンを削除して、縮退版文型パターン辞書 D4 を作成する。

3.5 実験結果

3.5.1 実験条件

文型パターン辞書: Pat12.1.0-imi6.0.0.dat (単語レベル, 122,719 件)

文型パターンパーサ: jpp2 version 5.3-kc + pat2map.rb version 1.1

関数定義ファイル: version 8.4a

クロスバリデーション用入力形態素: all_keitaiso+zokusei_20050613toku5 (122,719 文)

評価用試験文集: 日本語文型辞典から抽出した重文・複文 (3,851 文)

3.5.2 適合頻度 0 の文型パターンの削除（手順 1）

表 5 の結果より、オリジナルの文型パターン辞書 (D0) から適合頻度 0 の文型パターン (97,346 件) を削除し、辞書 D1 を作成した。

3.5.3 評価用試験文集の照合結果（手順 2）

手順 1 で削除した文型パターン（97,346 件）と評価用試験文集を，パーサを用いて照合した．その結果，出力された文型パターンは 73 件であった．

この 73 件を必要であると判断して，辞書 D1 に追加し，辞書 D2 を作成した．

以下に文型パターン例を示す．また，付録に追加した 73 件全てを示す．

- 文型パターン例 1

日本語文：

緑地の少ないところには住みたくない。

日本語文型パターン：

/y </tk N1 は> /tcfk N2 の /f AJ3^{rentai} /f N4 には /cf V5.tai /yf (ない | 無い)。

英語文：

I do not want to live in a place devoid of greenery.

英語文型パターン：

<I|N1> do not V5^{base}.want in N4 AJ3 of N2.

- 文型パターン例 2

日本語文：

いくら待っても来なかった。

日本語文型パターン：

/y /cf CL2 (ても | でも) ! CL4.kako。

英語文：

I waited and waited, but he did not appear.

英語文型パターン：

<I|N1> waited and waited, but <he|N2> did V3.not.

3.5.4 包含関係による縮退（手順 3）

表 6 に，文型パターン辞書 D2 に対して，日本語文型パターンの包含関係による縮退を行った結果を，文型パターンの適合頻度別に示す．結果より，日本語文型パターン間の包含関係によって，約 7 割に縮退できたことが分かる．

以下に，発見した包含関係の例を示す．

表 6 包含関係判定結果 (辞書 D2)

適合頻度	パターン数	上位パターン	下位パターン	下位削除後 (D3)
0 文	73	0 0.00%	7 9.59%	66 90.41%
1 文	8,905	763 8.57%	1,915 21.50%	6,990 78.50%
2 文以上 10 文未満	9,124	1,212 13.28%	2,598 28.47%	6,526 71.53%
10 文以上 100 文未満	5,125	1,251 24.41%	1,710 33.37%	3,415 66.63%
100 文以上 1,000 文未満	1,684	480 28.50%	852 50.59%	832 49.41%
1,000 文以上 10,000 文未満	532	90 16.92%	388 72.93%	144 27.07%
10,000 文以上 100,000 文未満	3	0 0.00%	0 0.00%	3 100.00%
計	25,446	3,796 14.92%	7,470 29.36%	17,976 70.64%

包含関係例 1

● 上位パターン

日本語文：

先生は学生達を呼び寄せて注意を与えた。

日本語文型パターン：

/y \$1^{/tcfk N1 は } /tcfk N2 を \$1 /cf V3 (て | で) \$1 /ytck N4 を /cf V5.kako 。

英語文：

The teacher called together the students to give them a warning.

英語文型パターン：

N1 V3^{past} N2 to V5^{base} N2^{pron}obj N4.

● 下位パターン (1)

日本語文：

子どもたちは手をつないで輪を作った。

日本語文型パターン：

/y #1{/tcfk N1 は,/tcfk N2 を } /cf つないで /ytck N3 を /cf V4.kako 。

英語文：

The children joined hands to make a circle.

英語文型パターン：

N1 joined N2 to V4^{base} N3.

● 下位パターン (2)

日本語文：

彼は背を屈めて門をくぐった。

日本語文型パターン：

$/y \$1^{\wedge}/\{tcfk N1 \text{ は } \} /tcfk \text{ 背を } /cf \text{ 屈めて } \$1 /ytck N2 \text{ を } /cf V3.kako。$

英語文：

He bent over to go under the gate.

英語文型パターン：

$N1 \text{ bent over to } V3^{\wedge}base N2.$

包含関係例 2

- 上位パターン

日本語文：

屋根に梯子を掛けて上がりました。

日本語文型パターン：

$/y </tk N1 \text{ は}> \#1\{tcfk N2 \text{ に},/tcfk N3 \text{ を } \} /cf V4 (\text{て} | \text{で}) /ycf V5 \#6[.masu].kako$
。

英語文：

He put a ladder up against the roof and climbed it.

英語文型パターン：

$<He|N1> V4^{\wedge}past N3 \text{ up against } N2 \text{ and } V5^{\wedge}past N2^{\wedge}pron^{\wedge}obj.$

- 下位パターン (1)

日本語文：

相場に手を出して懲りた。

日本語文型パターン：

$/y </tk N1 \text{ は}> \#1\{tcfk N2 \text{ に},/tcfk \text{ 手を } \} /cf \text{ 出して } /ycf (\text{懲り} | \text{こり}) \text{ た}.$

英語文：

I have learned by experience what it is to dabble in speculation.

英語文型パターン：

$<I|N1> \text{ have learned by experience what it is to dabble in } N2.$

- 下位パターン (2)

日本語文：

子供に本を読んで寝かしつけた。

日本語文型パターン：

$/y </tk N1 \text{ は}> \#1\{tcfk N2 \text{ に},/tcfk \text{ 本を } \} /cf V3 (\text{て} | \text{で}) /yf V4.kako。$

英語文：

She read the child to sleep.

英語文型パターン：

<She|N1> V3^{past} N2 to V4^{base}.

包含関係例 3

- 上位パターン

日本語文：

車が泥をはねて行った。

日本語文型パターン：

/y #1{/tcfk N1 が, /tcfk N2 を } /cf V3 (て | で) /ycf V4.kako。

英語文：

The car splattered mud as it went by.

英語文型パターン：

N1 V3^{past} N2 as N1^{pron} V4^{past}.

- 下位パターン (1)

日本語文：

敵艦が白旗を掲げて来た。

日本語文型パターン：

/y #1{/cf N1 が, /tcfk N2 を } ! 掲げて /ycf V3.kako。

英語文：

An enemy ship came flying the flag of truce.

英語文型パターン：

N1 V3^{past} flying N2.

- 下位パターン (2)

日本語文：

流れが堤防を越えてあふれ出した。

日本語文型パターン：

/y #1{/tcfk 流れが, /tcfk N1 を } /cf (越え | 逾え) て /ycf V2.kako。

英語文：

The stream overflowed its banks.

英語文型パターン：

The stream V2^{past} its N1.

3.5.5 過剰汎化文型パターンの削除（手順4）

表5の結果より，辞書D3から適合頻度10,000以上の文型パターン（3件）を削除し，縮退版文型パターン辞書（D4）を作成した．

3.5.6 辞書の縮退結果

提案手法を用いて，文型パターン辞書の縮退を行った．

再現率を文型パターン辞書作成用の用例 ($R1_{cross}$) と，評価用試験文集 ($R1_{test}$) で，作成した各辞書について求めた．結果を表7に示す．

結果より，縮退版文型パターン辞書（D4）は，オリジナルの文型パターン辞書（D0）の14.65%の規模になったことが分かる．また，辞書D4は辞書D0と比べ再現率 $R1_{cross}$ ， $R1_{test}$ 共に低下した．これは適合頻度の高い，過剰に汎化された文型パターンを削除したためと考えられる．

表7 縮退結果

文型パターン辞書	パターン数（割合）		再現率	
			$R1_{cross}$	$R1_{test}$
D0：オリジナルの文型パターン辞書	122,719	100.00%	67.98%	43.00%
D1：適合頻度0の文型パターンを削除	25,373	20.68%	67.98%	42.59%
D2：評価用試験文集に適合する文型パターンを追加	25,446	20.74%	67.98%	43.00%
D3：包含関係による縮退	17,976	14.65%	67.64%	42.98%
D4：縮退版の文型パターン辞書	17,973	14.65%	61.68%	35.34%

4 人手による翻訳調査

文型パターン辞書の意味的な精度を確認するために，前章で作成した各文型パターン辞書を用いて，人手による翻訳の評価実験を行った．

4.1 評価対象

以下で示す 2 つの入力文集を用いた．

- 日英対訳コーパスからランダムに抽出した 50 文
- 評価用試験文集からランダムに抽出した 100 文

日英対訳コーパスを用いる場合は，クロスバリデーションテスト，評価用試験文集を用いる場合はオープンテストとなる．

4.2 評価基準

人手で翻訳を行った結果を以下の基準で評価する．

- A:** 品質の高い英文が生成できる
- B:** 重要ではない要素の欠如（ただし，節の欠如は除く）があるが，簡単に修正できる
- C:** 入力文を部分的に訳せている
- D:** 入力文の訳としては使用できない
- F:** 英文の生成ができない

評価の例を以下に示す．

4.2.1 評価 A の例

例 1

入力文：

暑いので窓を開けた。

適合日本語文型パターン：

/y #1[/tcfk TIME2 は] /cf AJ3^{rentai} ので </yc N4 は> ! N5 を /cf V6.kako 。

（原文：今日は寒いのでストーブを焚いた。）

英語文型パターン：

<I|N4> V6^{past} N5 #1[ADV(TIME2)] because it was AJ3.

(原文 : I turned on the heater today because it was cold.)

翻訳結果 :

I opened the window because it was hot.

例 2

入力文 :

彼はライバルを押さえて優勝した。

適合日本語文型パターン :

/y </tk N1 は> /tcfk N2 を #3[/cf きちんと] /f V4 (て | で) #5[/ytk V6 (て | で)] /f
(V7.kako|ND7 をした)。

(原文 : 上着をきちんと直して儀式ばって敬礼した。)

英語文型パターン :

<He|N1> V4^{past} <his|N1^{poss}> N2 and V(V7|ND7)^{past} #5[ADV(V6)].

(原文 : He straightened his jacket and saluted formally.)

翻訳結果 :

He held off his rival and won a victory.

例 3

入力文 :

かぜがこじれて肺炎になった。

適合日本語文型パターン :

/ytcfk N1 が /cf V2 (て | で) /ytck N3 に ! (なっ | 成っ) た。
(原文 : 氷が溶けて水になった。)

英語文型パターン :

N1 V2^{past} into N3.

(原文 : The ice melted into water.)

翻訳結果 :

A cold grew worse into pneumonia.

4.2.2 評価 B の例

例 1

入力文：

彼の家族はみんな風変わりな連中だった。

適合日本語文型パターン：

/y \$1^{/tcfk N1 は } /cf ADV2 \$1 /f AJV3^rentai ! N4.#da.kako。

(原文：それは実に困難な状況でした。)

英語文型パターン：

N1 be^past ADV2 AJ3 N4.

(原文：It was really a difficult situation.)

翻訳結果：

The family were all queer lot.

「彼の」に相当する部分が欠落しているが、その点を除くと品質の高い英文なので評価 B となる。

例 2

入力文：

熱を出した子どもをつききりで看病した。

適合日本語文型パターン：

/y </tk N1 は> /tcfk N2 を /cf V3.kako^rentai /f N4 を /cf V5.kako。

(原文：私たちが救った軍隊を率いた。)

英語文型パターン：

<He|N1> V5^past N4 that V3^apast N2^obj.

(原文：He led the force that relieved us.)

翻訳結果：

He nursed a child that ran a fever.

「つききりで」に相当する部分が欠落しているが、その点を除くと品質の高い英文なので評価 B となる。

例 3

入力文：

彼は彼女の命令を遂行するためにロンドンへ行った。

適合日本語文型パターン：

/y \$1 /tcfk N1 を /cf V2^{rentai} ! ために \$1^{{/tcfk N3 は } /tcfk N4 へ /cf V5.kako 。}

(原文：箔をつけるために彼はイギリスへ行った。)

英語文型パターン：

N3 V5^{past} to N4 in order to V2^{base} N1.

(原文：He went to England in order to gain prestige.)

翻訳結果：

He went to London in order to achieve instructions .

「彼女の」に相当する部分が欠落しているが，その点を除くと品質の高い英文なので評価 B となる。

4.2.3 評価 C の例

例 1

入力文：

病気とは全然知らずに彼を訪れた。

適合日本語文型パターン：

/y </tk N1 は> #2[/tcfk 容赦！なく] #3[/ytck N4 の] ! N5 を /cf V6.kako 。

(原文：容赦なく小説の第一草稿を廃棄した。)

英語文型パターン：

<I|N1> #2[ruthlessly] V6^{past} N5 #3[of N4].

(原文：I ruthlessly destroyed the first drafts of the novel.)

翻訳結果：

I visited him.

前半部分が完全に欠落し，「彼を訪れた」部分のみ翻訳出来ているので評価 C となる。

例 2

入力文：

あの課長は人を使うコツを知っている。

適合日本語文型パターン：

/y \$1^{{/cf #1[REN2] /k N3 は } #4[/cf V5^{rentai}] /f N6 を \$1 /cf V7.teiru 。}

(原文：軽飛行機を操縦する免許をもっている。)

英語文型パターン：

#1[AJ2] N3 V7 N6 #4[to V5^{base}].

(原文：He has a license to pilot a light airplane.)

翻訳結果：

That section chief knows knack to use.

「使う」の目的語「人」が翻訳されていないため，評価 C となる．

例 3

入力文：

彼はライバルを押さえて優勝した。

適合日本語文型パターン：

/y \$1^{{/cf #1[GEN2] /k N3 は } /cf V4 (て | で) \$1 /ycf V5.kako。}

(原文：その家はぐらついて倒れた。)

英語文型パターン：

#1[AJ2] N3 V4^{past} and V5^{past}.

(原文：The house trembled and fell.)

翻訳結果：

He held off and won a victory.

「押さえる」の目的語「ライバル」が翻訳されていないため，評価 C となる．

4.2.4 評価 D の例

例 1

入力文：

彼女の足音がテラスを横切って反響した。

適合日本語文型パターン：

/y </tk N1 は> /tcfk N2 を /cf V3 (て | で) </ycf N4 は> ! (V5.kako|ND5 をした)。

(原文：試験を終えて帰省した。)

英語文型パターン：

<My|N1^{poss}> N2 ADV(V3), <I|N4> V(V5|ND5)^{past}.

(原文：My examination over, I went home.)

翻訳結果：

My terrace across, I echoed.

入力文の意味を全く表現できていないので評価 D となる．

例 2

入力文：

かぜがこじれて肺炎になった。

適合日本語文型パターン：

/y </tk N1 は> #2[/tcfk N3 の] #4[/f 突飛な] /f N5 に /cf V6.kako。

(原文：彼の突飛な発言にまごついた。)

英語文型パターン：

<We|N1> V6^{past}passive at #2[N3^{poss}] #4[absurd] N5.

(原文：We were puzzled at his absurd suggestion.)

翻訳結果：

We were become at pneumonia.

入力文の意味を全く表現できていないので評価 D となる。

例 3

入力文：

彼は怒ってぼくにげんこつを振り上げた。

適合日本語文型パターン：

/y </tk N1 は> #2[/tcfk N3 の] #4[/f AJV5^{rentai}] /f N6 に /cf V7.kako。

(原文：彼のみごとな手跡に感じ入った。)

英語文型パターン：

<We|N1> be^{past} deeply V7^{ed} with #2[N3^{poss}] #4[N(AJV5)] N6.

(原文：We were deeply impressed with his good hand.)

翻訳結果：

We were deeply raised with me.

入力文の意味を全く表現できていないので評価 D となる。

4.2.5 評価 F の例

例 1

入力文：

昔はいざしらず、今は会社を 10 も持つ大実業家だ。

適合日本語文型パターン：

/y \$1^{tcfk N1 は} #2[/tcfk TIME3] \$1 #1{/tck #4[N5 から],/tcfk N6 を} /cf V7^{rentai}

! N8.#da。

(原文：警察は明日家族から事情を聞く予定だ。)

英語文型パターン：

$N1 V(N8)^{prog} to V7^{base} about N6 \#4[from N5] \#2[N3].$

(原文：The police are planning to hear about the situation from the family tomorrow.)

バインド値：

N1=今, N6=会社, V7=待つ, N8=大実業家

日本語文型パターンの名詞変数 N8 には「大実業家」が適合しているが、英語文型パターンでは品詞変換関数 V(N8) が記述されている。そのため、「大実業家」を動詞として訳出しなければならないが、適切な動詞が見つからないため、翻訳をすることが出来ないので評価 F となる。

例 2

入力文：

祖母は自分一人では起き上がることもできないでいる。

適合日本語文型パターン：

$/y </tk N1 \text{は}> \#2[/tcfk TIME3、] / \#4[N5 \text{を} /cf N6、] \#7[/ytk TIME8] /f V9 \#10[.masu].$

(原文：朝、旗を上げ、夕方下ろします。)

英語文型パターン：

$<We|N1> \#4[V(N6) N5] \#2[in N3] and V9 N5 \#7[in N8].$

(原文：We raise the flag in the morning and lower it in the evening.)

バインド値：

N1=祖母, V9=いる

英語文型パターンには、必ず必要な要素として名詞変数 N5 が記述されている。しかし、この適合例では N5 のバインド値がないため、翻訳することができない。このため評価 F となる。なお、これは対訳文型パターンの記述誤りである。

例 3

入力文：

彼はライバルを押さえて優勝した。

適合日本語文型パターン：

$/y </tk N1 \text{は}> /tcfk N2 \text{を} /cf V3 (\text{て} | \text{で}) \#4[/y ADV5] /f (V6.kako|ND6 \text{をした}).$

(原文：それを聞いて大いに安心した。)

英語文型パターン：

<I|N1> be #4[ADV5] AJ(V6|ND6) to V3^base N2.

(原文：I am much relieved to hear that.)

バインド値：

N1=彼，N2=ライバル，V3=押さえ，V6=優勝し(優勝する)

日本語文型パターンの動詞変数 V6 には「優勝し(優勝する)」が適合しているが、英語文型パターンでは品詞変換関数 AJ(V6|ND6) が記述されている。そのため、「優勝し(優勝する)」を形容詞として訳出しなければならないが、適切な形容詞が見つからないため、翻訳をすることが出来ないので評価 F となる。

4.3 評価手順

評価の手順を以下に示す。

1. 入力文を文型パターン辞書と照合し、適合文型パターンを抽出する。

- 入力文：彼は走るのが速い。
- 適合パターン例

日本語文型パターン

/y #1[/tcfk N2 の] /k N3 は /cf V4^rentai ! のが /cf AJ5。

英語文型パターン

#1[N2] N3 V4 ADV(AJ5).

バインド値

N3=彼，V4=走る，AJ5=速い

2. バインド値を、人手で最も適切な英単語に翻訳し、英語文型パターンに代入する。

N3： 彼 he

V4： 走る run

AJ5： 速い fast

翻訳結果： he runs fast.

3. 翻訳結果を評価基準に従って評価する。

評価： A

4.4 正解率

英文の評価結果をもとに，以下の2通りの正解率を算出する．

1. 全ての入力文に対して抽出された全ての文型パターンについて，A および B と評価された文型パターンが存在する割合（文型正解含有率，以下 $P1$ ）
2. 入力文に対して，A および B と評価された文型パターンが1つでも存在する割合（適合文型正解含有率，以下 $P2$ ）

$P1$ および $P2$ は以下の式でそれぞれ求める．

$$P1 = \frac{A, B \text{ と評価されたパターン数}}{\text{抽出された全パターン数}}$$

$$P2 = \frac{\text{1つ以上 } A, B \text{ と評価されたパターンを持つ文数}}{\text{自己パターン以外のパターンが1つ以上抽出された文数}}$$

4.5 評価結果

クロスバリデーションテストの評価結果を表8に，オープンテストの評価結果を表9に示す．また，付録に詳細な評価結果を示す．

表 8 人手による翻訳評価の結果（クロスバリデーションテスト：50 文）

辞書	パターン数	再現率 R1	正解率		R1 × P1	R1 × P2
			P1	P2		
D0	122,719	70% (35/50)	23% (185/797)	49% (17/35)	16%	34%
D1	25,373	70% (35/50)	23% (185/797)	49% (17/35)	16%	34%
D2	25,446	70% (35/50)	23% (185/797)	49% (17/35)	16%	34%
D3	17,976	70% (35/50)	18% (55/308)	46% (16/35)	13%	32%
D4	17,973	68% (34/50)	19% (55/290)	47% (16/34)	13%	32%

表 9 人手による翻訳評価の結果（オープンテスト：100 文）

辞書	パターン数	再現率 R1	正解率		R1 × P1	R1 × P2
			P1	P2		
D0	122,719	44% (44/100)	5% (16/311)	16% (7/44)	2%	7%
D1	25,373	43% (43/100)	5% (15/308)	14% (6/43)	2%	6%
D2	25,446	44% (44/100)	5% (16/310)	16% (7/44)	2%	7%
D3	17,976	44% (44/100)	6% (14/216)	14% (6/44)	3%	6%
D4	17,973	34% (34/100)	7% (14/193)	18% (6/34)	2%	6%

5 考察

5.1 適合頻度 0 の文型パターン

適合頻度 0 の文型パターンを削除した影響を考察するために、オリジナルの文型パターン辞書 (D0) と適合頻度 0 の文型パターンを削除した辞書 (D1) の翻訳評価結果を比較する。

クロスバリデーションテスト (表 8) では評価結果は変化していない。これは、適合頻度 0 の文型パターンは、クロスバリデーションテストにおいては全く利用されないためである。しかし、オープンテスト (表 9) においては、 $R1$ 、 $P1$ 、 $P2$ の全てがわずかに低下した。

ところが、提案手法の手順 2 で追加した文型パターン数は 73 件と、適合頻度 0 の文型パターン数 (97,436 件) に比べてかなり少ない。したがって、適合頻度 0 の文型パターンの削除によって、文型パターン辞書に与える影響はかなり小さいと言える。

5.2 適合頻度 1 の文型パターン

本研究では、適合頻度 0 の文型パターンを、使われる可能性が少ないとして削除した。しかし、適合頻度 1 の文型パターンも使われる可能性が少ないと言える。そこで、適合頻度 1 の文型パターンが唯一適合している入力文が、適合する文型パターン数を調査した。その結果を表 10 に示す。

表 10 適合頻度 1 のパターンが適合した文が、適合するパターン数の分布

適合パターン数	パターン数 / 割合	
1 件	134	1.50%
2 件以上 10 件未満	4,837	54.32%
10 件以上 100 件未満	3,219	36.15%
100 件以上 1,000 件未満	708	7.95%
1,000 件以上 10,000 件未満	7	0.08%
計	8,905	-

結果より、適合パターン数が 1 件の入力文に適合する、適合頻度 1 の文型パターンは 134 件と少ないことが分かる。つまり、この 134 件以外は、削除しても再現率が低下しな

い可能性があると考えられる。

以下に、適合頻度 1 の文型パターン、その文型パターンが適合する入力文および、文型パターンを用いた翻訳結果を示す。

適合頻度 1 の文型パターン

日本語文：

人間は理性のある動物である。

日本語文型パターン：

/ytcfk N1 は /tcfk 理性 (の | が) /f (ある | 在る | 有る) /f N2.#da ある。

英語文：

Man is a rational animal.

英語文型パターン：

N1 be a rational N2.

適合する入力文と翻訳結果

入力文：

人は理性のある動物である。

上記の文型パターンによる翻訳結果：

Man is a rational animal.

評価：

A

この入力文は以下の文型パターンにも適合する。

日本語文：

これは時間の掛かる仕事だ。

日本語文型パターン：

/ytcfk N1 は /tcfk N2 (の | が) /f V3^rentai ! N4.#da #5(.genzai|.kako)。

英語文：

This is a job that takes time.

英語文型パターン：

N1 'is' #5(^present|^past) N4 that V3 N2.

適合頻度：

12

この文型パターンを用いた場合の翻訳結果を以下に示す。

翻訳結果：

Man is an animal that has a reason.

評価：

A

以上の結果より，他の文型パターンを用いても意味的に正しい翻訳結果が得られる場合，適合頻度 1 の文型パターンは削除可能であると言える．文型パターンの意味的に正しい適合を判定する手法が必要となるが，さらなる文型パターン辞書縮退の手がかりの 1 つであると考えている．

5.3 包含関係による縮退の与える影響

包含関係を利用した縮退が，辞書の精度に与える影響を考察するために，オリジナルの文型パターン辞書 (D0) と包含関係による縮退までを行った辞書 (D3) の翻訳評価結果を比較する．

クロスバリデーションテスト (表 8) では，辞書 D3 は辞書 D0 に比べて $P1$ が大きく減少した．しかし，オープンテスト (表 9) では，辞書 D3 の $P1$ が約 1% ではあるが増加した．また，両方のテストで，辞書 D3 は辞書 D0 に比べて， $P2$ が減少している．以下に，クロスバリデーションテストにおいて，正解パターンが無くなった入力文，包含関係によって削除された正解パターン，翻訳結果および評価を示す．

入力文：

彼は彼女の命令を遂行するためにロンドンへ行った。

正解パターン 1

日本語文型パターン：

$/y \$1 /tcfk N1$ を $/cf V2^{rentai}$! ために $\$1^{/tcfk N3}$ は } $/tcfk N4$ へ $/cf V5.kako$ 。

(原文：箔をつけるために彼はイギリスへ行った。)

英語文型パターン：

$N3 V5^{past}$ to $N4$ in order to $V2^{base}$ $N1$.

(原文：He went to England in order to gain prestige.)

バインド値：

N3=彼, N1=命令, V2=遂行する, N4=ロンドン, V5=行っ=行く

翻訳結果:

He went to London in order to achieve instructions.

評価:

B

正解パターン 2

日本語文型パターン:

/y \$1^{/tcfk N1 は } /tcfk N2 を /cf V3^rentai ! ために \$1 /tcfk N4 へ /cf V5.kako 。

(原文: 彼は大統領選を報道するためにアメリカへ渡った。)

英語文型パターン:

N1 V5^past over to N4 to V3^base N2.

(原文: He went over to the U.S. to cover the Presidential election.)

バインド値:

N1=彼, N2=命令, V3=遂行する, N4=ロンドン, V5=行っ=行く

翻訳結果:

He went over to London to achieve instructions.

評価:

B

この2件の正解パターンは, 以下で示す文型パターンに包含されるため, 削除された。

日本語文型パターン:

/y \$1^{/tcfk N1 は } /tcfk N2 を /cf V3^rentai ! ために \$1 /tcfk N4 へ \$1 /cf V5.kako

。

(原文: 彼は情報を収集するためにそこへ行った。)

英語文型パターン:

N1 V5^past ADV(N4) to V3^base N2.

(原文: He went there to glean information.)

しかし, この文型パターンを用いた翻訳結果の評価はFであったので, 正解パターンが存在しなくなった。以下にその翻訳結果と評価を示す。

バインド値:

N1=彼, N2=命令, V3=遂行する, N4=ロンドン, V5=行っ=行く

翻訳結果：

He went ADV(London) to achieve instructions.

評価：

F, “London” を副詞に変換できないため

これら結果は、日本語文型パターンにのみ着目し、英語文型パターンを考慮しなかったためと考えられる。

英語文型パターンを考慮しない場合、日本語文型パターンに適合する日本語文を、意味的に適切な翻訳が行える英語文型パターンが存在しなくなる可能性がある。そのため、 $P1$ 、 $P2$ が影響を受けたと考えられる。

しかし、文型パターン翻訳方式においては、入力文に対して正解パターンが存在することが重要である。そのため、 $P2$ の減少の程度を考慮すると、日本語文型パターンの包含関係による縮退が、文型パターン辞書に与える影響は少ないと考えている。

5.4 過剰汎化文型パターンの与える影響

過剰に汎化されたと考えられる文型パターン（3件）が与える影響を考察するために、包含関係による縮退までを行った辞書（D3）と、この3件を削除した、縮退版文型パターン辞書（D4）を比較する。

表7より、辞書D4は辞書D3と比べて、再現率（ $R1_{cross}$ 、 $R1_{open}$ ）が大きく減少したことが分かる。つまり、この3件の文型パターンのみが適合する入力文が多く存在することが言える。^{*1}

しかし、表8、9の結果から意味的な被覆率（ $R1 \times P2$ ）は、辞書D3、D4のどちらも同じ値となっている。さらに、正解率 $P1$ 、 $P2$ の両方ともが向上している。

したがって、過剰に汎化された文型パターンを削除することで、再現率は低下するが、意味的な精度が変化しなかったため文型パターン辞書としての精度は向上したと考えている。

そこで、3.3節で調査した、適合頻度上位10件の文型パターンを調査した。その結果、過剰に汎化された文型パターンを、本研究で削除した3件に加え、新たに4件発見した。

- 適合頻度：9,245

日本語文：

*1 クロスバリデーションテストで約6%（約7,700文）、オープンテストで約8%（約300文）

すばらしく見事な身のこなしてステージを移動した。

日本語文型パターン：

/y </tk N1 は> #2[/cf すばらしく /f 見事な /f 身の /k こなしで] /tcfk N3 を /cf
V4.kako。

英語文：

He moved across the stage with marvelous deftness.

英語文型パターン：

<He|N1> V4^{past} across N3 #2[with marvelous deftness].

- 適合頻度： 7,067

日本語文：

聴衆は思ったように反応した。

日本語文型パターン：

/y \$1^{/tcfk N1 は } #2[/cf (思っ | おもっ | 惟っ | 意っ | 憶っ | 懐っ | 想っ | 念っ)
た (ように | 様に)] \$1 /yf (V3.kako|ND3 をした)。

英語文：

The audience reacted predictably.

英語文型パターン：

N1 V(V3|ND3)^{past} #2[predictably].

- 適合頻度： 7,067

日本語文：

彼女は涙ぐみながらほほえんだ。

日本語文型パターン：

/y \$1^{/tcfk N1 は } #2[/cf V3 ながら] \$1 /yf V4.kako。

英語文：

She smiled in a mist of tears.

英語文型パターン：

N1 V4^{past} #2[in a mist of N(V3)].

- 適合頻度： 7,062

日本語文：

街路に群がり集まった。

日本語文型パターン：

/y </tk N1 は> /tcfk N2 に #3[/cf (群がり | 叢り | 簇り | 羣がり)] /yf V4.kako。

英語文：

They swarmed the streets.

英語文型パターン：

<They|N1> V4^{past} N2.

したがって、今後さらに過剰に汎化された文型パターンを削除することで、意味的に不適切な文型パターンへの適合が減少し、辞書の精度が向上すると考えられる。

5.5 適合頻度とパターン適合率と翻訳評価の関係

5.5.1 パターン適合率 (C)

パターン適合率は、入力文の文字が文型パターンの離散記号以外の要素に適合する割合を表す。以下に計算例を表す。

例 1

入力文：

彼は学校へ行った。(9 文字)

適合文型パターン：

/y N1 は /tcfk N2 へ /ycf V3.kako 。

バインド値：

N1=彼, N2=学校, V3=行っ, .kako=た

パターン適合率：

$$C = \frac{9}{9} = 1.0 (100\%)$$

例 2

入力文：

彼は学校へ行った。(9 文字)

適合文型パターン：

/y N1 は /ycf V2.kako 。

バインド値：

N1=彼, /ycf=学校へ, V3=行っ, .kako=た

パターン適合率：

$$C = \frac{6}{9} = 0.66 (66\%)$$

5.5.2 適合頻度とパターン適合率の関係

第 3.3.1 項で調査した文型パターンの適合頻度と、その文型パターンが入力文と適合するときのパターン適合率の平均値との関係を調査した。作成した分布図を図 1 に示す。図

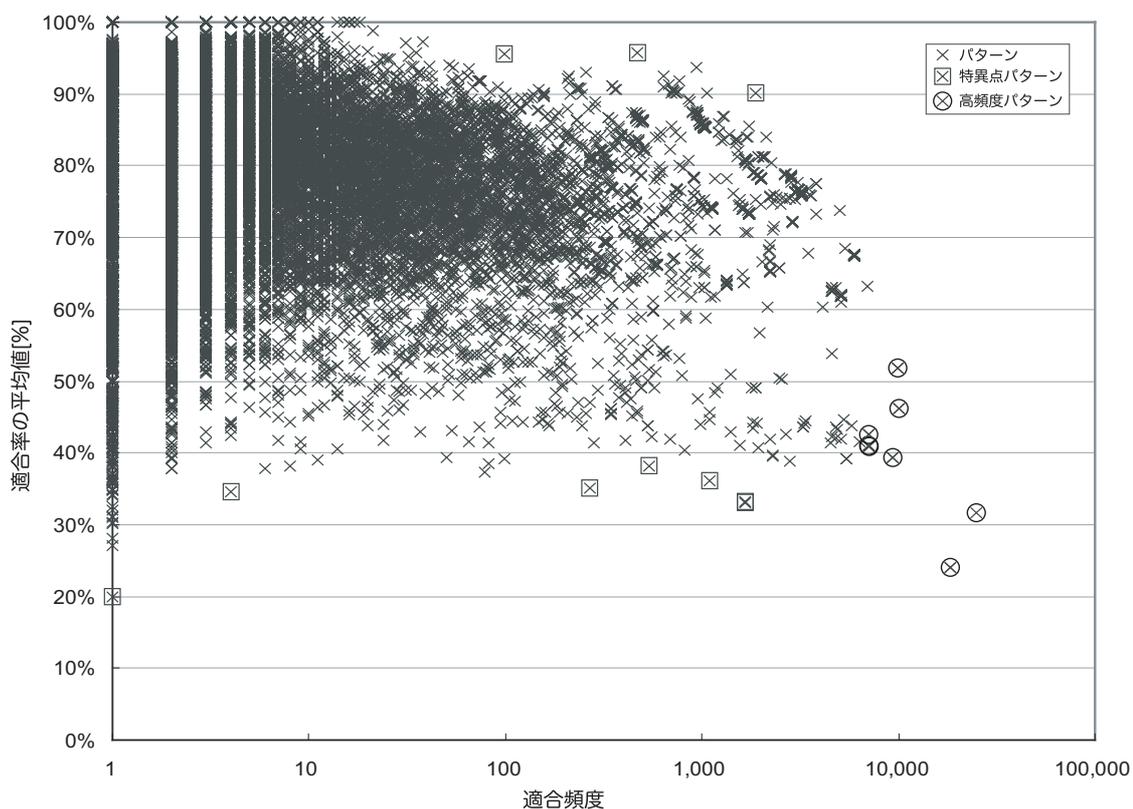


図 1 適合頻度とパターン適合率の平均の関係

より、多くの文型パターンが適合頻度 1 から 100、パターン適合率の平均値が 60% から 90% の範囲に集中していることが分かる。

5.5.3 高頻度文型パターンの翻訳評価

図 1 の ⊗ で示す点は、適合頻度が上位 8 件の文型パターンを表す。この 8 件の文型パターンを用いて翻訳評価を行った。入力文として、文型パターンが適合する文からランダムに 10 文を抽出し、第 4 章と同じ評価基準 (A, B, C, D の 4 段階) で評価した。なお、適合文数が 10 文未満の場合は全ての適合文を用いた。

評価結果を表 11 に示す。また、具体的な結果を付録に示す。

表 11 適合頻度上位 8 件の翻訳評価結果

評価	A	B	C	D	F
件数	0	2	35	28	15

評価の結果，評価値の高い英文はほとんど得られなかった．前節で述べたとおり，適合頻度の高い文型パターンは過剰に汎化されている傾向があるため，品質の高い英文を得ることは難しいと考えられる．

5.5.4 特異点にある文型パターンの翻訳評価

図 1 の特異点にある文型パターンを 8 件選択し，入力文として，文型パターンが適合する文からランダムに 10 文を抽出して評価した．選択した文型パターンを図中の ☒ で示す．

評価結果を表 12 に示す．また，具体的な結果を付録に示す．

表 12 特異点 8 件の翻訳評価結果

評価	A	B	C	D	F
件数	1	10	34	40	0

結果より，高評価となった英文も存在するが，以下で示す文型パターンに集中しており，その他の文型パターンの評価は低いことが分かった．

日本語文：

彼女は死んだごきぶりを見て食欲をなくした。

日本語文型パターン：

/ytcfk N1 は #2[/cf V3.kako^rentai] /f N4 を /cf V5 (て | で) </ytck N6 は> ! N7 を /cf V8.kako 。

英語文：

She lost her appetite when she saw a dead cockroach.

英語文型パターン：

<He|N1> V4^past across N3 #2[with marvelous deftness].

適合頻度：

470 件

パターン適合率の平均値：

95.69%

10 文の評価：

A:1, B:6, C:2, D:1, F:0

よって、特異点に存在する文型パターンが、翻訳に利用できる可能性は低いと考えられる。

5.5.5 集中部の文型パターンの翻訳評価

図 1 の文型パターンが集中している部分（頻度：1～100 件，パターン適合率の平均値：60%～90%）からランダムに 19 件抽出し，入力文として，文型パターンが適合する文からランダムに 10 文を抽出して評価した。なお，適合文数が 10 文未満の場合は全ての適合文を用いた。

評価結果を表 13 に示す。また，具体的な結果を付録に示す。

表 13 集中部 19 件の翻訳評価結果

評価	A	B	C	D	F
件数	15	15	21	8	0

評価の結果，A または B 評価の翻訳結果が他の部分に比べ多い傾向があるため，集中している部分の文型パターンを用いることで，評価の高い英文が得られる可能性が高いと考えられる。

また，入力文と文型パターンの原文が似ている場合に評価が高いという傾向が見られた。

6 おわりに

本研究では，精度を維持したまま文型パターン辞書を縮退する方法を検討し，単語レベルの文型パターン辞書(122,719件)の縮退を行った．その結果，17,973件(14.65%)に縮退することが出来た．また，再現率，正解率はともに低下したが，文型パターンの削減幅を考えると，問題は少ないと考えている．

適合頻度0の文型パターンは，日常的には使われない表現が多く，削除が辞書の精度に与える影響は小さいことが分かった．しかし，日本語文型パターンの包含関係を用いると正解率が低下することが分かった．そのため，英語文型パターンも考慮し，日英両方の文型パターンの包含関係を用いる必要がある．また，過剰に汎化された文型パターンを辞書から削除すると，再現率は減少したが，意味的な被覆率が変化しなかったことから，このような文型パターンが辞書に悪影響を与えていたことが確認出来た．

今後は，再現率を維持したまま，正解率も維持する辞書の削減方法を考えていく．そのためには，日英の文型パターンが共に包含関係にあるものに着目する方法が考えられる．

7 謝辞

本論文作成に際して，多大なる検討と助言をしてくださった池原悟教授ならびに村上仁一助教授，徳久雅人助手，そして計算機工学講座C研究室の方々に心から感謝します．さらに，本論文について多くの御意見，御助言を頂きました，木村周平助教授に深く感謝します．

最後に，参考にさせて頂いた文献の著者の方々に対して感謝します．

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複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] 池原悟, 宮崎正弘, 白井諭, 横尾昭男, 中岩浩巳, 小倉健太郎, 大山芳史, 林良彦: 日本語彙大系, 岩波書店, 1997.
- [3] 池原悟, 宮崎正弘, 佐良木昌, 池田尚志, 白井諭, 村上仁一, 徳久雅人: 機械翻訳のための日英文型パターン記述言語, 電子情報通信学会技術研究報告, TL2002-48, pp. 1–6, 2003.
- [4] 徳久雅人, 村上仁一, 池原悟: 文型パターンパーサの試作, 言語処理学会 第 10 回年次大会発表論文集, pp. 608–611, 2004.
- [5] James Allen: *Natural Language Understanding(2nd Edition)*, The Benjamin/Cummings Publishing Company, Inc., pp. 101–106, 1994.
- [6] 砂川有里子, 駒田聡, 下田美津子, 鈴木睦, 筒井佐代, 蓮沼昭子, ベケシュ・アンドレイ, 森本順子: 日本語文型辞典, くろしお出版, 1998.
- [7] 片山慶一郎, 村上仁一, 徳久雅人, 池原悟: 日本語文型パターンの縮退方法, 言語処理学会 第 12 回年次大会発表論文集, pp. 568–571, 2006.
- [8] 片山慶一郎, 村上仁一, 徳久雅人, 池原悟: 日英機械翻訳のための文型パターン辞書の圧縮に関する検討, 情報処理学会研究報告, 2006-NL-178, 2007 (発表予定).