

概要

パターン翻訳は、入力された原言語に対し、大量に用意した文パターンを照合し、翻訳文を出力する方法である。通常、パターン翻訳に用いる文パターン辞書の作成は、人手で行うため開発に時間がかかる。しかしパターン翻訳の長所として、入力文が文パターンに適合した場合は翻訳精度の高い文が得られる。これは文パターンが持つ大局的な文法情報により、文全体の構造を保持した翻訳文が得られるためである。

本研究では人手によるコストを無くすため、文パターンをプログラムで自動作成する。そしてパターン翻訳から得られた翻訳候補文に対して、tri-gramを用いた絞込みを行い、翻訳文を選択する。この提案する翻訳システムと従来の統計翻訳システムの比較を行い、提案手法の有効性について調査した。

実験の結果、条件付きではあるが、自動評価方法と人手評価方法の両方から、提案手法の有効性が示された。

目次

第1章	はじめに	1
第2章	英日パターン翻訳システム	3
2.1	パターン翻訳の概要	3
2.2	英日パターン翻訳の手順	3
2.3	文パターン辞書	4
第3章	英日統計翻訳システム	5
3.1	概要	5
3.1.1	英日統計翻訳の枠組み	5
3.1.2	単語に基づく統計翻訳	6
3.1.2.1	IBM 翻訳モデル	7
3.1.2.2	GIZA++	7
3.1.2.3	単語に基づく統計翻訳の問題点	7
3.1.3	句に基づく統計翻訳	8
3.2	翻訳モデル	9
3.3	言語モデル	9
3.4	デコーダ	10
3.5	パラメータチューニング	11
3.6	評価方法	11
第4章	実験における注意点	12
4.1	英日文パターン辞書	12
4.2	英日パターン翻訳	12
4.3	データベース	12

第5章	提案手法	13
5.1	英日対訳単語辞書の作成手順	13
5.2	英日文パターン辞書の作成手順	13
5.3	翻訳文生成の手順	14
第6章	実験条件	17
6.1	単文データ	17
6.2	tri-gram 学習データ	17
6.3	ベースラインシステム	18
6.3.1	翻訳モデルの学習	18
6.3.2	言語モデルの学習	18
6.3.3	デコーダのパラメータ	18
第7章	実験結果	19
7.1	翻訳文に対するランク付け	19
7.2	単文テストデータの自動評価結果	19
7.2.1	英日文パターンに照合した文のみの評価	19
7.2.2	各ランクとベースラインの比較	20
7.2.3	10000 文での自動評価結果	21
7.3	単文テストデータの人手評価結果	21
7.3.1	A ランクの翻訳例	22
7.3.2	B ランクの翻訳例	23
7.3.2.1	提案手法○の翻訳例	23
7.3.2.2	提案手法×の翻訳例	25
7.3.3	C ランクの翻訳例	26
7.3.3.1	提案手法○の翻訳例	26
7.3.3.2	提案手法×の翻訳例	27
7.3.4	D ランクの翻訳例	28
7.3.4.1	提案手法○の翻訳例	28
7.4	単文データのまとめ	30
第8章	重文・複文におけるパターン翻訳	31
8.1	実験条件	31

8.2	重文・複文データ	31
8.3	tri-gram 学習データ	31
8.4	重文・複文テストデータの自動評価結果	31
8.4.1	パターンに照合した文のみの評価	31
8.4.2	A ランクとベースラインの比較	32
8.4.3	10000 文での自動評価結果	33
8.5	重文・複文テストデータの人手評価結果	34
8.5.1	A ランクの翻訳例	34
8.5.2	B ランクの翻訳例	35
8.5.2.1	提案手法○の翻訳例	35
8.5.3	C ランクの翻訳例	37
8.5.3.1	提案手法○の翻訳例	37
8.5.3.2	提案手法×の翻訳例	37
8.5.4	D ランクの翻訳例	38
8.6	重文・複文データのまとめ	39
第9章	考察	40
9.1	単文におけるパターン翻訳の翻訳精度	40
9.2	単文における A ランクの翻訳文の誤訳	40
9.3	単文における類似文パターンの調査	42
9.4	重文・複文における翻訳精度の考察	43
9.5	重文・複文における類似文の調査	44
9.6	翻訳精度とカバー率	44
9.7	翻訳精度向上の手段	45
9.7.1	意味制約を用いた辞書	45
9.7.2	英日対訳単語辞書の翻訳確率	45
9.8	日英方向におけるパターン翻訳との比較	45
9.9	単語に基づく統計翻訳器の検討	46
第10章	おわりに	47

目 次

2.1	英日パターン翻訳の手順	4
3.1	英日統計翻訳の枠組み	6
3.2	デコーダの動作例	10
5.1	英日文パターン辞書の作成手順	14
5.2	翻訳文の生成手順	15
5.3	翻訳文選択の手順	16

表 目 次

2.1	英日文パターンの例	4
3.1	英日方向の単語対応	8
3.2	日英方向の単語対応	8
3.3	言語モデルの例	9
5.1	GIZA++を用いた英日方向の単語辞書の例	13
7.1	各ランクの自動評価結果	19
7.2	A ランクの翻訳文とベースラインの比較	20
7.3	B ランクの翻訳文とベースラインの比較	20
7.4	C ランクの翻訳文とベースラインの比較	20
7.5	D ランクの翻訳文とベースラインの比較	21
7.6	提案手法とベースラインの比較	21
7.7	50 文での人手評価の結果	22
7.8	A ランクにおける提案手法○の翻訳例	22
7.9	A ランクにおける提案手法○の翻訳例	23
7.10	A ランクにおける提案手法○の翻訳例	23
7.11	B ランクにおける提案手法○の翻訳例	24
7.12	B ランクにおける提案手法○の翻訳例	24
7.13	B ランクにおける提案手法○の翻訳例	24
7.14	B ランクにおける提案手法×の翻訳例	25
7.15	B ランクにおける提案手法×の翻訳例	25
7.16	B ランクにおける提案手法×の翻訳例	25
7.17	C ランクにおける提案手法○の翻訳例	26
7.18	C ランクにおける提案手法○の翻訳例	26
7.19	C ランクにおける提案手法○の翻訳例	27

7.20	C ランクにおける提案手法×の翻訳例	27
7.21	C ランクにおける提案手法×の翻訳例	28
7.22	C ランクにおける提案手法×の翻訳例	28
7.23	D ランクにおける提案手法○の翻訳例	29
7.24	D ランクにおける提案手法×の翻訳例	29
7.25	D ランクにおける提案手法×の翻訳例	29
7.26	D ランクにおける提案手法×の翻訳例	30
8.1	各ランクの翻訳文における自動評価結果	32
8.2	A ランクの翻訳文とベースラインの比較	32
8.3	B ランクの翻訳文とベースラインの比較	32
8.4	C ランクの翻訳文とベースラインの比較	33
8.5	D ランクの翻訳文とベースラインの比較	33
8.6	提案手法とベースラインの比較	33
8.7	10 文における人手評価結果	34
8.8	A ランクにおける提案手法○の翻訳例	34
8.9	A ランクにおける提案手法○の翻訳例	35
8.10	A ランクにおける提案手法○の翻訳例	35
8.11	B ランクにおける提案手法○の翻訳例	36
8.12	B ランクにおける提案手法○の翻訳例	36
8.13	B ランクにおける提案手法○の翻訳例	36
8.14	C ランクにおける提案手法○の翻訳例	37
8.15	C ランクにおける提案手法×の翻訳例	37
8.16	C ランクにおける提案手法×の翻訳例	38
8.17	D ランクにおける提案手法×の翻訳例	38
8.18	D ランクにおける提案手法×の翻訳例	39
9.1	A ランクにおける誤訳例 1	40
9.2	A ランクにおける誤訳例 2	41
9.3	A ランクにおける誤訳例 3	41
9.4	A ランクにおける誤訳例 4	42
9.5	A ランクにおける誤訳例 5	42
9.6	類似文の例 1	43

9.7 類似文の例 2	43
9.8 類似文の例 3	43
9.9 類似文の例 4	43
9.10 類似文の例 1	44
9.11 類似文の例 2	44

第1章 はじめに

機械翻訳の歴史は、大きく3つの世代に分類できる。第一世代は1960年代半ばに提案されたパターン翻訳である。パターン翻訳は、人手で作成した大量の文パターンを用意し、入力文が文パターンに適合した場合に翻訳文を出力する方法である。しかし、1966年に、機械翻訳における研究成果を悲観視するALPACレポートが発表され、機械翻訳の研究は衰退した。

第二世代は1980年代に提案された用例翻訳である。用例翻訳は、コーパスに含まれる例文と入力文の類似性を利用した翻訳方法である。入力文に対して類似している用例を参照し、それらの用例を組み合わせて翻訳を行う。用例翻訳は1980年代半ばに、商用化が開始したが、高価であったため、一般には普及しなかった。また、翻訳精度も、評価できるとは言い難かった。

第三世代は1990年代前半に提案された統計翻訳である。統計翻訳は、対訳データより自動的に翻訳規則を獲得し、言語を統計的に扱い翻訳する方法である。現在は、2000年代始めに提案された、統計翻訳に属する句に基づく統計翻訳が主流である。

本研究では、統計翻訳において、単語の対応が自動的にとれることに着目し、英日パターン翻訳の精度調査を行う。パターン翻訳は長所として、入力文が文パターンに適合した場合、翻訳精度の高い文が得られる。しかし、通常文パターン辞書の作成は人手で行うため開発に時間がかかる。

文パターン辞書の開発コストをなくすため、本研究では文パターンをプログラムで自動作成する。そしてパターン翻訳から得られた翻訳候補文に対して、tri-gramを用いた絞込みを行い、翻訳文を選択する。実験の結果から、用いたtri-gramのスコアが高い翻訳文において、翻訳精度の高い文が得られた。

なお、過去に西村ら [1] はパターン翻訳と句に基づく統計翻訳を組み合わせた手法(以降、先行手法と述べる)で日英翻訳を行い、翻訳精度を求めている。そこで、考察として先行手法と比較した際の利点、欠点について報告する。また句に基づく統計翻訳との精度比較を行い結果を報告する。実験の結果、条件付きながら、従来の英日統計翻訳システムと比べて提案手法の有効性が確認できた。

本論文の構成は以下の通りである。第2章で従来の英日パターン翻訳システムについて説明し、第3章で英日統計翻訳システムについて説明する。第4章で本研究で使用するデータベースについて説明し、第5章では提案する翻訳システムについて説明する。第6章で実験条件を述べ、第7章で実験結果を示す。第8章で追加実験方法と結果について示し、第9章で本研究の考察を述べる。

第2章 英日パターン翻訳システム

2.1 パターン翻訳の概要

パターン翻訳は、大量の文パターン対を用いて、文パターンの照合を行い翻訳文を出力する方法である。パターン翻訳はパターンが適合した場合、文全体の構造を保持した翻訳精度の高い翻訳文を得ることができる。しかし、一般的なパターン翻訳は文パターンを手で作成するため開発に時間がかかる。また、文パターンに辞書に適合しない場合は翻訳ができないため、入力文に対するカバー率が低いという問題がある。

2.2 英日パターン翻訳の手順

一般的な英日パターン翻訳の手順を以下に示す。

手順1 英日文パターン辞書と英日対訳単語辞書を用意する。

手順2 入力文と英語文パターンを照合する。

手順3 照合に成功した場合、変数部に対応する英単語を英日対訳単語辞書を用いて日本語単語に翻訳する。

手順4 英語文パターンに対応する日本語文パターンの変数部を、翻訳した日本語単語に置き換える。

手順5 手順4で生成した翻訳文を出力する。

英日パターン翻訳の手順を図 2.1 に示す。

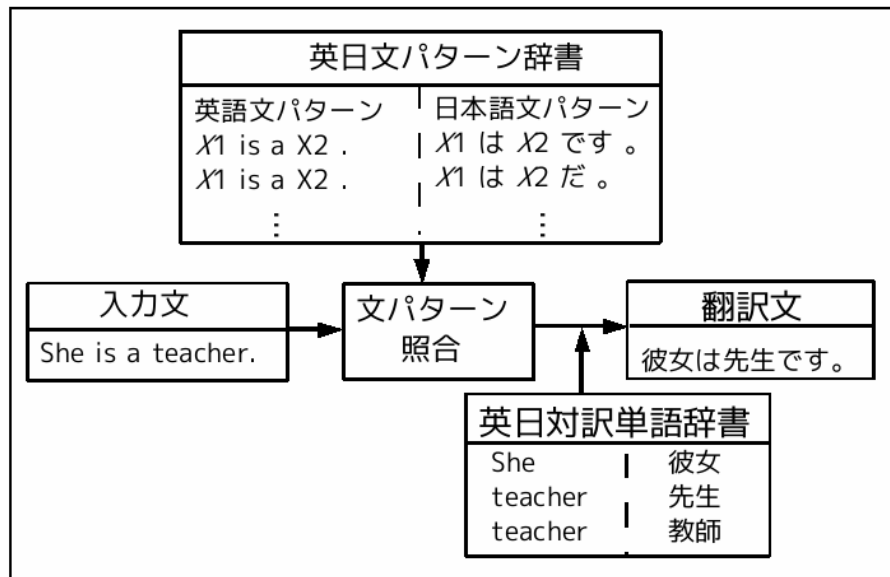


図 2.1 英日パターン翻訳の手順

2.3 文パターン辞書

文パターン辞書とは、大量の対訳文から対応する単語を変数により置き換えることで得られる文パターンの集合である。

表 2.1 に例を示す。なお、本研究における文パターン辞書の作成手順は 5.2 節で示す。

表 2.1 英日文パターンの例

入力文	The fire started in the kitchen .
英語文パターン	The X1 started in the X2 .
日本語文パターン	X1 は X2 から 出た。
翻訳文	火は台所から出た。

第3章 英日統計翻訳システム

3.1 概要

統計翻訳とは、原言語と目的言語の対訳文を大量に収集した対訳データにより、自動的に翻訳規則を獲得し翻訳を行う、機械翻訳手法の一種である。

統計翻訳には”単語に基づく統計翻訳”と”句に基づく統計翻訳”があり、初期の統計翻訳では単語に基づく統計翻訳が用いられていた。しかし近年、句に基づく統計翻訳が提案され、単語に基づく統計翻訳に比べて翻訳精度が高いことがわかった。このため現在は句に基づく統計翻訳が主流となっている。

3.1.1 英日統計翻訳の枠組み

英日統計翻訳システムの枠組みを図3.1に示す。

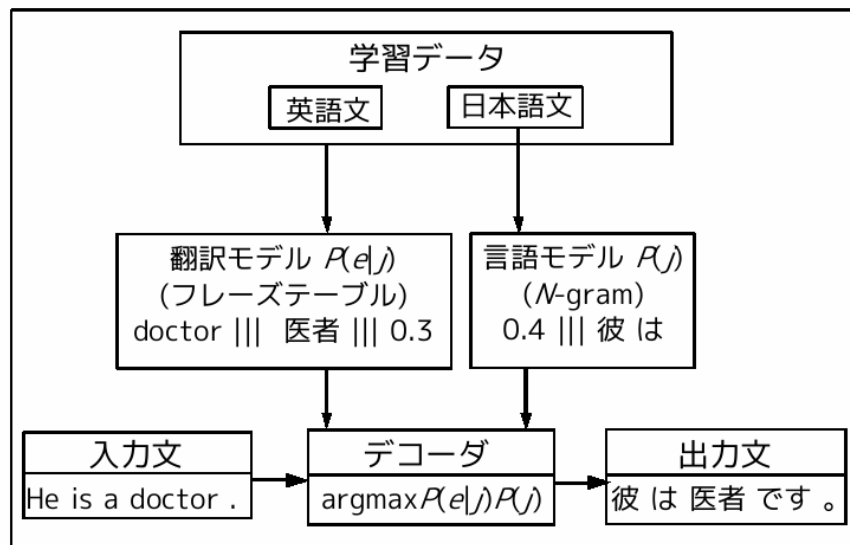


図 3.1 英日統計翻訳の枠組み

英日統計翻訳は、入力文として英語文 E が与えられた場合に、翻訳モデルと言語モデルの組み合わせの中から確率が最大となる日本語文 J を探索することで翻訳を行う。以下にその基本モデルを示す。

$$J = \operatorname{argmax}_e P(j|e) \quad (3.1)$$

$$\simeq \operatorname{argmax}_e P(e|j)P(j) \quad (3.2)$$

ここで $P(e|j)$ は翻訳モデル、 $P(j)$ は言語モデルを示す。 $P(j)$ が単語であれば”単語に基づく統計翻訳”のモデル、 $P(j)$ が句であれば、”句に基づく統計翻訳”のモデルとなる。

また、学習データとは対訳データを大量に用意したものである。学習データに含まれる各々のデータから、それぞれ翻訳モデルと言語モデルを学習する。

3.1.2 単語に基づく統計翻訳

単語に基づく統計翻訳は単語対応の翻訳モデルを用いている。ある英文を日本語文に訳すときに、英単語を日本語に訳し、英単語の語順と同じ並びで日単語を並べて翻訳するという手法である。

3.1.2.1 IBM 翻訳モデル

統計翻訳の代表的なモデルとして、IBMのBrownらによる仏英翻訳モデル [5] がある。IBM 翻訳モデルは、単語に基づく統計翻訳を想定して作成された、単語対応の確率モデルである。IBM 翻訳モデルでは、フランス語から英語への翻訳を想定しているため、以下の説明では仏英翻訳を前提とする。本章では、原言語であるフランス語文を f 、目的言語である英語文を e として定義する。

IBM モデルでは、フランス語文 e 、英語文 f の翻訳モデル $P(F|E)$ を計算するために、アライメント a を用いる。以下に IBM モデルの基本式を示す。

$$P(f|e) = \sum_a P(f, a|e) \quad (3.3)$$

アライメント a は、フランス語単語 F と英単語 E の対応関係を示す。IBM モデルにおいて、英単語は 1:n の対応をもち、フランス語単語は 1:1 の対応をもつと仮定する。またフランス語単語に適切な対応関係をもつ英単語が存在しなかった場合、英語文の先頭に e_0 という特殊文字があると仮定する。そして、フランス語単語と特殊文字の対応付けを行う。

3.1.2.2 GIZA++

GIZA++[2] とは英日方向、日英方向の対訳文について最尤な単語の対応を得るツールである。対訳文を用いて IBM モデルを学習し、単語レベルの確率を得る。

3.1.2.3 単語に基づく統計翻訳の問題点

以下に、英日方向の単語対応の例と、日英方向の単語対応の例を示す。また、●は単語が対応した箇所を示す。

表 3.1 英日方向の単語対応

	I	go	to	see	a	movie
私	●					
は						
映画						●
を					●	
見				●		
に						
行く		●	●			

表 3.2 日英方向の単語対応

	I	go	to	see	a	movie
私	●					
は					●	
映画						●
を			●			
見				●		
に				●		
行く		●				

表 3.1 と表 3.2 を比較すると、表 3.2 は全ての単語に対して対応がとれている。しかし、表 3.1 は日本語単語 ”は” と ”に” に対応する英単語が存在しない。単語に基づく統計翻訳は対応する単語が存在しない場合、何も無い状態から単語の発生確率を計算する。このため単語の確率における信頼性が問題となっている。よって現在は句単位の対応を用いた統計翻訳が行われている。

3.1.3 句に基づく統計翻訳

句に基づく統計翻訳は句対応の翻訳モデルを用いる。ある英文を日本語文に訳すときに、隣接する複数の単語 (フレーズ) を用いて翻訳を行う方法である。

3.2 翻訳モデル

翻訳モデルとは、膨大な量の対訳データを用いて英語のフレーズが日本語のフレーズへ確率的に翻訳を行うためのモデルである。この翻訳モデルはフレーズテーブルで管理されている。以下にフレーズテーブルの例を示す。

フレーズテーブルの例

```
The flower ||| その花 ||| 0.428571 0.0889909 0.428571 0.0907911 2.718
Tonight's concert is ||| 今晚のコンサートは ||| 0.5 0.000223681 0.5 0.0124601 2.718
```

左から英語フレーズ、日本語フレーズ、フレーズの英日方向の翻訳確率 $P(j|e)$ 、英日方向の単語の翻訳確率の積、フレーズの日英方向の翻訳確率 $P(e|j)$ 、日英方向の単語の翻訳確率の積、フレーズペナルティ(値は常に自然対数の底 $e=2.718$) である。

3.3 言語モデル

言語モデルとは、膨大な量の単言語データを用いて単語の列や文字の列が起こる確率を付与するモデルである。統計翻訳では主に N -gram を用いる。以下に言語モデル (tri-gram) の例を示す。表中の w_1 , w_2 , w_3 はそれぞれ tri-gram の単語列における 1 番目, 2 番目, 3 番目の単語を示している。

表 3.3 言語モデルの例

tri-gram の単語列 $w_1 w_2 w_3$	$\log_{10}(P(w_3 w_1 w_2))$ (スムージングなし)	$\log_{10}(P(w_3 w_1 w_2))$ (バックオフスムージング)
痛み が 伴う	-1.382585	-0.3105274
堤防 が 決壊	-1.061585	-0.1920604
天気 が よかつ	-1.768149	-0.1920604
納得 が いか	-0.6635545	-0.1101559
梅雨 が 明け	-0.7214168	-0.1029072
風 が 吹く	-2.222238	-0.1920604

表の 1 行目の例では、左側の数値が、“痛み” と “が” という文字列が連続した後に、“伴う” が出現する確率を常用対数で表した値 “ $\log_{10}(P(\text{伴う} | \text{痛み が})) = -1.382585$ ” を、

中央が tri-gram で表された単語列である ”痛み が 伴う” を，右側の数値はバックオフスムージングにより得られる， ”痛み が” の後に ”伴う” が出現する確率を常用対数で表した値 $\log_{10}(P(\text{伴う} | \text{痛み が})) = -0.3105274$ を示している。

また，バックオフスムージングとは，高次の N -gram が存在しない場合において，低次の N -gram の値を用いて高次の N -gram の値を推定する方法である。

3.4 デコーダ

デコーダは，翻訳モデルと言語モデルを用いて，確率が最大となる翻訳候補を探索し，出力を行う変換器のことである．代表的なデコーダとして，”Moses[4] がある。

入力文として ”She is a teacher .” が与えられたときの翻訳例を図 3.2 に示す。

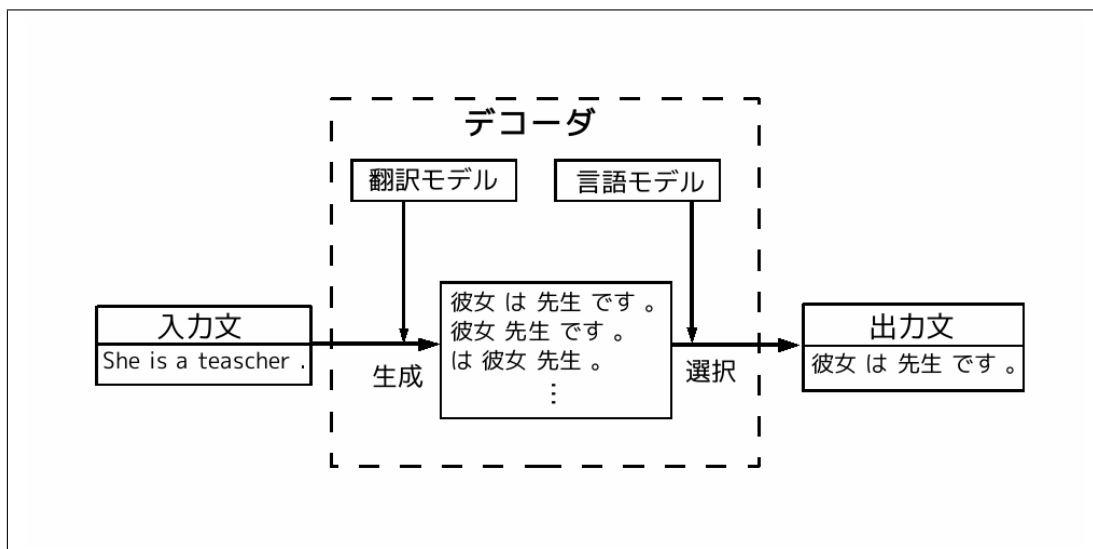


図 3.2 デコーダの動作例

英日統計翻訳において， $\operatorname{argmax}_e P(e|j)P(j)$ の確率が最大となる日本語文を出力するためには，適切な順序で英語と日本語の単語対応を得る必要がある．しかし，適切な英語文を決定するためには，計算量が膨大となり，かつ莫大な時間が必要となる．そこで計算量を削減するために，ビームサーチ法を用いる．

ビームサーチ法とは，翻訳候補の探索において，翻訳確率の低い翻訳候補を枝刈りし，探索範囲を減退する方法である．探索領域の中で一定の確率以上の翻訳候補のみを残し，それ以外の翻訳候補は除外する．

ただし，ビームサーチ法は，切り捨てられた翻訳候補が文章全体で見たときに，最大

の確率を持つ翻訳候補であったという可能性がある。そのため選択した翻訳文が最適解であるとは限らないという問題がある。

3.5 パラメータチューニング

パラメータチューニングは、デコーダで用いるパラメータを最適化することである。一般的に Minimum Error Rate Training(MERT)[10] という手法が用いられる。MERT は目的の評価関数 (一般的には BLEU) を最大にする翻訳結果が選ばれるように、パラメータ調整を行う。この際、ディベロップメントデータと呼ばれる、試し翻訳を行うデータを与える。そして、各文に対して上位 100 個程度の翻訳候補を出力し、よりよい翻訳候補が上位にくるようパラメータを調整する。

3.6 評価方法

機械翻訳システムの翻訳精度を自動的に評価する手法として、事前に用意した参照文と、翻訳システムから得た翻訳文を比較する手法が一般的である。本研究では自動評価方法に BLEU[8] と NIST[9] を用いる。BLEU と NIST は共に語順の正しさで比較を行う。ただし、BLEU は 4-gram, NIST は 5-gram を用いる。BLEU は 0 から 1 までの間で、NIST は 0 から ∞ の間で評価され、語順が正しい場合に高いスコアを出す。BLEU も NIST もスコアが高いほど翻訳精度が高いことを示す。

第4章 実験における注意点

4.1 英日文パターン辞書

過去の研究 [13] に、単文の文パターン辞書作成が行われている。この研究では、変数に意味属性制約を付与して文パターンを作成しているが、本研究で作成する文パターンは意味属性の付与は行わない。

4.2 英日パターン翻訳

英日パターン翻訳において、文パターンの変数に適合する単語が複数の意味を持つ場合、全ての組み合わせの出力を行う。本実験では、5.1節に示す例のように、英単語が日本語単語に翻訳される確率は用いていない。

また、入力文が文パターンとの照合に失敗した場合、翻訳文の出力は行わない。

4.3 データベース

翻訳実験には電子辞書から抽出した単文データベースと重文・複文データベースを用いる [3]。使用するデータベースは以下のことに注意が必要である。

- 実験に使用した単文データは日本語文が単文である。このため、英語文に重文・複文が含まれる場合がある。
- 8節で述べる追加実験に使用した重文・複文データは、日本語文が重文・複文である。このため、英語文に単文が含まれる場合がある。

第5章 提案手法

GIZA++を用いて英日文パターンをプログラムで自動作成する。
以下に手順を示す。

5.1 英日対訳単語辞書の作成手順

手順1 GIZA++を用いて英日方向、日英方向の単語対応を調べ、英日単語辞書、日英単語辞書を作成する。作成した英日単語辞書は表 5.1 に示すように左から英単語、対応する日本語単語、英単語が日本語単語に翻訳される確率が与えられている。

手順2 手順1で作成した英日単語辞書と日英単語辞書の翻訳確率を掛け合わせ一定の閾値 (α) 以上の確率を持つ単語を英日対訳単語辞書として登録する。

表 5.1 GIZA++を用いた英日方向の単語辞書の例

fire	火	0.09
kitchen	台所	0.44

5.2 英日文パターン辞書の作成手順

英日対訳学習文と英日対訳単語辞書を用いて英日文パターン辞書を作成する。
以下に英日文パターン辞書作成の手順を示す。

手順1 英日対訳文における英語文の各単語と英日対訳単語辞書の単語を照合する。照合に成功したら英単語に対応する日本語単語と日本語学習文の単語を照合する。図 5.1 の例では、"I" と "私"，"movie" と "映画" が照合に成功する。

手順2 日本語文の照合に成功したら一致した英日対訳学習文の単語を変数に置き換える。
 今回は”I”と”私”に相当する部分がX1に,”movie”と”映画”に相当する部分がX2に置き換わる。そして英語文パターンとして”X1 go to see a X2.”が,日本語文パターンとして”X1はX2を見に行く。”が得られる。

図5.1に英日文パターン辞書作成の手順を示す。

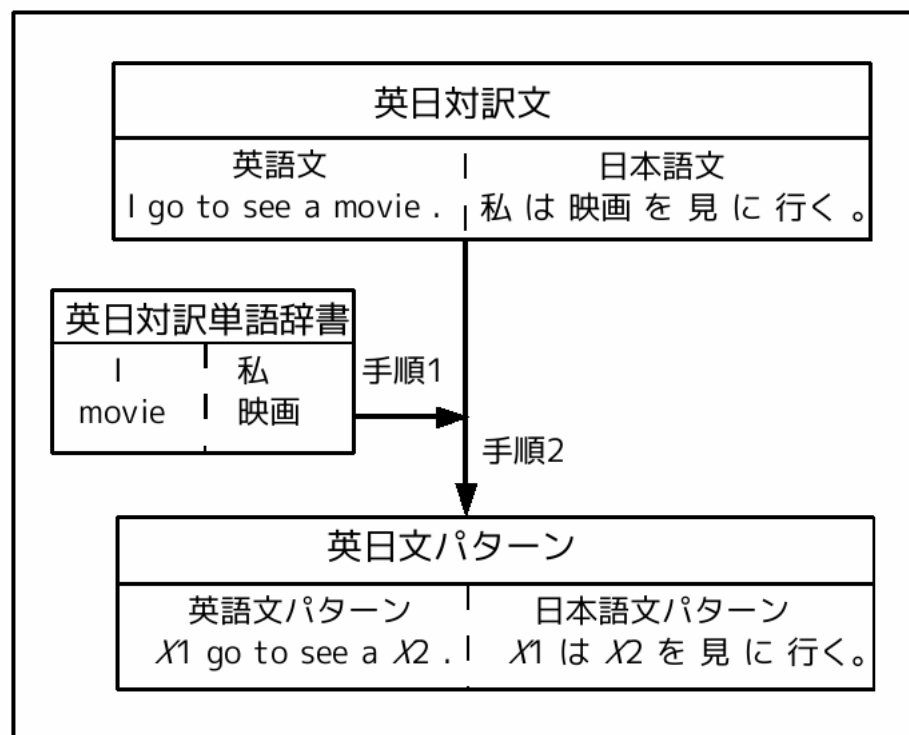


図 5.1 英日文パターン辞書の作成手順

5.3 翻訳文生成の手順

翻訳文生成の手順を以下に示す。なお,手順1から手順3までが翻訳候補生成の手順,手順4と手順5が翻訳候補文から翻訳文を選択する手順である。

手順1 入力文を読み込み、入力文に一致する英語文パターンを選択する。5.2の例では入力文として”She is a teacher .” が読み込まれ、英語文パターン”X1 is a X2 .” が選択される。

手順2 一致する英語文パターンが存在する場合、変数部に対応する英単語を英日対訳単語辞書を用いて日本語単語に翻訳する。5.2では変数部 X1 に対応する”She”を”彼女”に、X2 に対応する”teacher”を”先生”に翻訳する。

手順3 英語文パターンに対応する日本語文パターンの変数部を英日対訳単語辞書を用いて日本語単語に置き換える。5.2では日本語文パターンの X1 に当たる部分を”彼女”に、X2 に当たる部分を”先生”に置き換えている。結果として、”彼女は先生です。”や”彼女は先生だ。”などの翻訳候補文が生成される。

図 5.2 に翻訳候補文生成までの手順を示す。

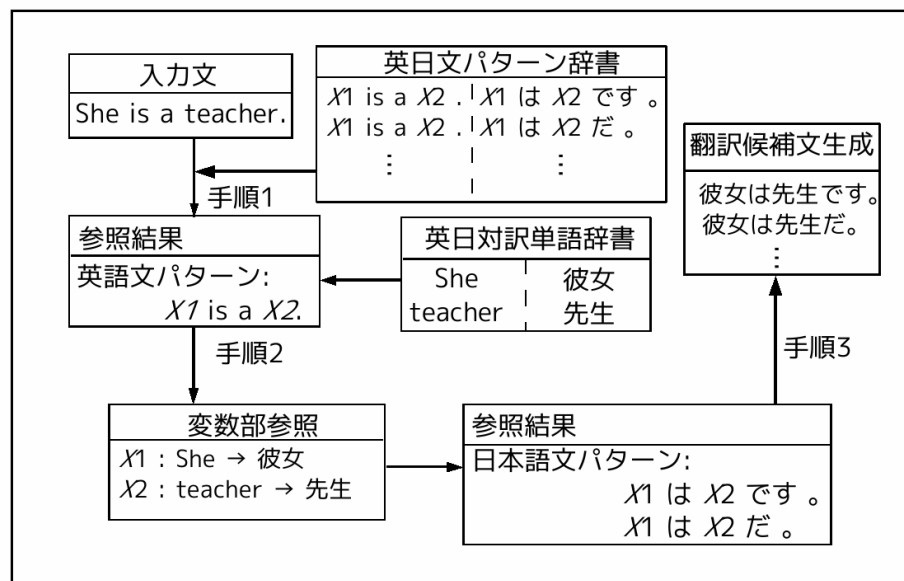


図 5.2 翻訳文の生成手順

手順4 複数の翻訳候補が出力された場合、tri-gram モデルを用いて翻訳候補の絞込みを行う。tri-gram の学習データは英日対訳学習文の日本語学習文を用いる。候補文に対し tri-gram のスコアが存在しない場合はペナルティーとして -1,000 を付与する。5.3 のスコア計算において、候補文を tri-gram モデル化し、学習データと照合する。例えば、翻訳候補文 ”彼女 は 先生 です。” は照合した結果、スコア-999.4 が与えられる。

手順5 tri-gram スコアが最大となる翻訳候補文を翻訳文として出力する。図 5.3 では候補文の中で最もスコアが高い ”彼女 は 先生 です。” が翻訳文として出力される。本実験では翻訳結果を分類するために翻訳文の tri-gram スコアに閾値を設ける。tri-gram 学習データと閾値の詳細は 6.2 節で述べる。

図 5.3 に翻訳文選択までの手順を示す。

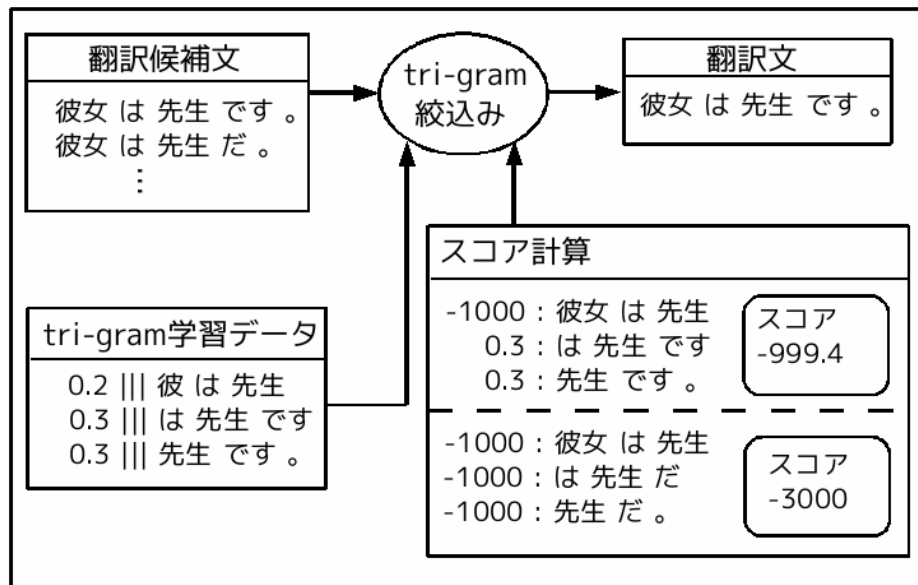


図 5.3 翻訳文選択の手順

第6章 実験条件

本実験では単文データベースから入力テスト文 10,000 文と英日対訳学習文 100,000 文を用いる。

6.1 単文データ

単文テストデータ 10,000 に対し英日パターン翻訳を行う。

英日文パターン辞書の作成には 4.1 節の手順 2 で示す $\alpha=0.05$ (17,128 語登録) を用いる。英日対訳学習文 100,000 文に対し 86,204 文の英日文パターンを得た。また、パターン翻訳には $\alpha=0.005$ (76,202 語登録) の英日対訳単語辞書を用いる。

6.2 tri-gram 学習データ

tri-gram 学習データには英日対訳学習文 100,000 文を使用する。翻訳文の tri-gram スコア (β) に閾値を設け、翻訳文を分類する。翻訳文は

- $-1,000 < \beta \leq 0$
- $-2,000 < \beta \leq -1,000$
- $-3,000 < \beta \leq -2,000$
- $-3,000 \leq \beta$

の 4 つに分類する。

また、以下の場合には翻訳結果を出力しない。

- 入力文が英語文パターンに一致しない
- 5.3 節の手順 2 において変数部に対応する単語が英日対訳単語辞書に登録されていない

6.3 ベースラインシステム

ベースラインシステムには Moses を用いる [4]。また、形態素解析には Mecab を使用する [6]。

6.3.1 翻訳モデルの学習

翻訳モデルの学習には、"train-model.perl" を用いる。"train-model.perl" は GIZA++ を利用し、IBM モデルの計算を行うプログラムである。

6.3.2 言語モデルの学習

言語モデルの学習には SRILM[11] の ngram-count を用いて 5-gram の言語モデルを学習する。なお、 N -gram モデルでは、値が 0 となることを防ぐために、スムージングによる近似を行う。本研究ではスムージングに "-ndiscount" を用いる。

6.3.3 デコーダのパラメータ

翻訳モデルにおいて、日英翻訳確率と英日翻訳確率の共起確率を用いる [12]。したがって、翻訳モデルの重み "weight-t" は "0.5 0.0 0.5 0.0 0.0" とする。また、フレーズの並び替えの重み "distortion weight" を 0.2 とする。本研究において、パラメータチューニングは行わない。

第7章 実験結果

7.1 翻訳文に対するランク付け

6.2節では，英日パターン翻訳から得た翻訳文に対し，tri-gram スコア (β) に閾値を設けた．本研究では，翻訳文をスコア別にランク付けする．以下にランクの定義を示す．

- A ランク : $-1,000 < \beta \leq 0$
- B ランク : $-2,000 < \beta \leq -1,000$
- C ランク : $-3,000 < \beta \leq -2,000$
- D ランク : $-3,000 \leq \beta$

7.2 単文テストデータの自動評価結果

7.2.1 英日文パターンに照合した文のみの評価

入力文 10,000 文に対し，3,252 文の翻訳文を得た．翻訳文 3,252 文に対して A ランクは 379 文，B ランクは 247，C ランクは 292 文，D ランクは 2,334 文であった．各ランクの BLEU と NIST の評価結果を表 7.1 に示す．

表 7.1 各ランクの自動評価結果

	BLEU	NIST
A ランク (379 文)	0.5664	6.9185
B ランク (247 文)	0.2993	4.4478
C ランク (292 文)	0.2228	3.9444
D ランク (2,334 文)	0.0686	3.0176
A~D ランク (3,252 文)	0.1683	3.8956

7.2.2 各ランクとベースラインの比較

単文において、パターン翻訳を用いて翻訳された A ランク 379 文と、ベースラインシステムで翻訳された 379 文の BLEU 値と NIST 値の比較結果を表 7.2 に示す。

表 7.2 A ランクの翻訳文とベースラインの比較

	BLEU	NIST
A ランク	0.5664	6.9185
ベースライン	0.5434	6.7248

パターン翻訳を用いて翻訳された B ランク 247 文と、ベースラインシステムで翻訳された 247 文の BLEU 値と NIST 値の比較結果を表 7.3 に示す。

表 7.3 B ランクの翻訳文とベースラインの比較

	BLEU	NIST
B ランク	0.2993	4.4478
ベースライン	0.3097	4.2523

パターン翻訳を用いて翻訳された C ランク 292 文と、ベースラインシステムで翻訳された 292 文の BLEU 値と NIST 値の比較結果を表 7.4 に示す。

表 7.4 C ランクの翻訳文とベースラインの比較

	BLEU	NIST
C ランク	0.2228	3.9444
ベースライン	0.2466	3.9138

パターン翻訳を用いて翻訳された D ランク 2,334 文と、ベースラインシステムで翻訳された 2,334 文の BLEU 値と NIST 値の比較結果を表 7.5 に示す。

表 7.5 D ランクの翻訳文とベースラインの比較

	BLEU	NIST
D ランク	0.0686	3.0176
ベースライン	0.1674	3.8581

上記の結果から，提案手法より得た A ランクの翻訳文はベースラインと比較して BLEU 値が 2.3% 向上している． B～D ランクはベースラインと比較して BLEU 値の低下が認められたが，単文におけるパターン翻訳は A ランクの翻訳文に関して特に高い翻訳精度が得られた．

7.2.3 10000 文での自動評価結果

提案手法は表 7.1 の A ランクの翻訳文 (379 文) と，ベースラインから得た A ランクを除く 9,621 文の翻訳文を合わせた 10,000 文を評価に用いる．そしてベースラインと比較する．

表 7.6 提案手法とベースラインの比較

	BLEU	NIST
提案手法	0.1381	3.7798
ベースライン	0.1375	3.7743

表 7.6 において，ベースラインと比較して提案手法は BLEU 値が 0.06% 向上した．

7.3 単文テストデータの人手評価結果

表 7.1 の A～D ランクに対し，ランダムに抽出した 50 文を用いて人手による比較を行う．評価基準を以下に示す．

- 提案手法○：提案手法がベースラインよりも良い
- 提案手法×：提案手法がベースラインに劣っている
- 差なし：翻訳品質に差がない

- 同一出力 : 提案手法, ベースラインの出力が同じ

表 7.7 に単文における人手評価の結果を示す.

表 7.7 50 文での人手評価の結果

ランク	提案手法○	提案手法×	差なし	同一出力
A ランク	9	0	19	22
B ランク	4	11	26	9
C ランク	7	9	33	1
D ランク	1	10	39	0

表 7.7 より, A ランクは提案手法○が多く, 提案手法の有効性が認められた. B~D ランクは提案手法×の割合が多く, 提案手法の有効性は認められなかった.

7.3.1 A ランクの翻訳例

以下に A ランクにおける提案手法○の例を示す.

表 7.8 A ランクにおける提案手法○の翻訳例

入力文	A pendulum swings back and forth .
英語文パターン	A X1 swings back and forth .
日本語文パターン	X1 は 左右 に 振動 する 。
提案手法	振り子は 左右 に 振動 する 。
ベースライン	振り子は 左右 に している 。
正解文	振り子は 左右 に 振動 する 。

表 7.8 において提案手法は正解文と同じ翻訳文が得られている. しかしベースラインは”swings”が翻訳されておらず, 文として不適切である.

表 7.9 A ランクにおける提案手法○の翻訳例

入力文	We've never heard any rumors about her love life .
英語文パターン	We've never heard any rumors about X1 love life .
日本語文パターン	X1 には 浮いた話 の一つもない。
提案手法	彼女 には 浮いた話 の一つもない。
ベースライン	彼女 には 浮いた。
正解文	彼女 には 浮いた話 の一つもない。

表 7.10 において、提案手法は正解文と同じ翻訳文が得られている。しかしベースラインは”浮いた”以降が翻訳されていないため、文として不適切である。

表 7.10 A ランクにおける提案手法○の翻訳例

入力文	I am staying with my friend .
英語文パターン	X1 am X2 with my X3 .
日本語文パターン	X1 は X3 のところに X2 している。
提案手法	私は友人のところに滞在している。
ベースライン	友達と しています。
正解文	今友人の所に泊まっている。

表 7.6 において提案手法は正解文と同等の翻訳文が得られている。しかしベースラインは入力文に対する翻訳文として不適切である。

7.3.2 B ランクの翻訳例

7.3.2.1 提案手法○の翻訳例

以下に B ランクにおける提案手法○の例を示す。

表 7.11 B ランクにおける提案手法○の翻訳例

入力文	She called him all the bad names .
英語文パターン	X1 called X2 all the bad names .
日本語文パターン	X1 は X2 を くそみそ に 言った。
提案手法	彼女は 彼 を くそみそ に 言った。
ベースライン	彼女は 彼 を を けなした。
正解文	彼女は さんざん 彼の 悪口 を 言った。

表 7.11 において提案手法は正解文と同等の翻訳文が得られている。しかしベースラインは ”を” が連続して翻訳されており、文として不適切である。

表 7.12 B ランクにおける提案手法○の翻訳例

入力文	Put those examinations on the desk .
英語文パターン	Put those X1 on the X2 .
日本語文パターン	その X1 は X2 の 上 に 載せ なさい。
提案手法	その 試験 は 机 の 上 に 載せ なさい。
ベースライン	それら の 試験 は 机 の 上 に 入れ なさい。
正解文	答案 は この 机 の 上 に 置き なさい。

表 7.12 において提案手法は正解文と同等の翻訳文が得られている。しかしベースラインは ”机の上に置く” と翻訳されており、日本語文として意味が通じず不適切である。

表 7.13 B ランクにおける提案手法○の翻訳例

入力文	She is a music teacher .
英語文パターン	X1 is a X2 X3 .
日本語文パターン	X1 は X2 の X3 だ。
提案手法	彼女は 音楽 の 先生 だ。
ベースライン	彼女は 音楽 先生 を していた。
正解文	彼女は 音楽 の 先生 を している。

表 7.13 において提案手法は正解文と同等の翻訳文が得られている。しかしベースラインは 時制が異なっている。また、 ”音楽” と ”先生” の間に助詞がなく文として適切でない。

7.3.2.2 提案手法×の翻訳例

以下に B ランクにおける提案手法×の例を示す。

表 7.14 B ランクにおける提案手法×の翻訳例

入力文	The telephone is out of order .
英語文パターン	X1 X2 X3 out of order .
日本語文パターン	X1 X2 は 狂い が きて X3 .
正解文	電話 が こわれている。
ベースライン	その 電話 は 故障 している。
提案手法	その 電話 は 狂い が きている。

表 7.14 においてベースラインは正解文と同等の翻訳文が得られている。しかし提案手法は日本語文として意味が通じない不適切な翻訳がされている。

表 7.15 B ランクにおける提案手法×の翻訳例

入力文	The moon came out .
英語文パターン	The X1 came out .
日本語文パターン	X1 が 抜けた。
提案手法	月 が 抜けた。
ベースライン	月 が 出た。
正解文	月 が 出た。

表 7.15 においてベースラインは正解文と同じ翻訳文が得られている。しかし提案手法の翻訳は "came out" が "抜けた" と翻訳されており、不適切な文が翻訳されている。

表 7.16 B ランクにおける提案手法×の翻訳例

入力文	The meeting is in confusion .
英語文パターン	X1 X2 X3 in X4 .
日本語文パターン	X1 X2 は X4 に のって X3 .
提案手法	その 集会 は 混乱 に のっている。
ベースライン	会議 は 混乱 状態 である。
正解文	会議 は 紛糾 している。

表 7.16 においてベースラインは正解文と同等の翻訳文が得られている。しかし提案手法の翻訳は適合した文パターンが不適切であったため不自然な日本語文が得られた。

7.3.3 Cランクの翻訳例

7.3.3.1 提案手法○の翻訳例

以下にCランクにおける提案手法○の例を示す。

表 7.17 Cランクにおける提案手法○の翻訳例

入力文	His pistol went off accidentally .
英語文パターン	His X1 went off accidentally .
日本語文パターン	その 男 の X1 が 暴発 した。
正解文	彼の ピストル が 暴発 した。
ベースライン	彼の ピストル は 誤って 出発 した。
提案手法	その 男 の ピストル が 暴発 した。

表 7.17 において提案手法は正解文と同等の翻訳文が得られている。しかしベースラインは ”went off” が ”出発” と翻訳されており、文として不適切である。

表 7.18 Cランクにおける提案手法○の翻訳例

入力文	My brother went to Tokyo on business .
英語文パターン	My X1 went to X2 on business .
日本語文パターン	X1 は 出張 で X2 まで 出かけた。
提案手法	兄 は 出張 で 東京 まで 出かけた。
ベースライン	兄 は 東京 へ 行った に 商売 を している。
正解文	兄 は 商用 で 東京 へ 行った。

表 7.18 において提案手法は正解文と同等の翻訳文が得られている。しかしベースラインは日本語文として不適切である。

表 7.19 Cランクにおける提案手法○の翻訳例

入力文	The avalanche gathered momentum .
英語文パターン	The X1 gathered X2 .
日本語文パターン	X1はX2を加えた。
提案手法	なだれは勢いを加えた。
ベースライン	雪崩が場が集まった。
正解文	なだれはあっそう加速した。

表 7.19 において提案手法は正解文と同等の翻訳文が得られている。しかしベースラインは日本語文として不適切である。

7.3.3.2 提案手法×の翻訳例

以下にCランクにおける提案手法×の例を示す。

表 7.20 Cランクにおける提案手法×の翻訳例

入力文	He is bent with age .
英語文パターン	X1 X2 X3 with X4 .
日本語文パターン	X1はX4にX3ていX2。
正解文	彼は年で腰が曲がっている。
ベースライン	彼は年のせいで腰が曲がっている。
提案手法	彼は時代に曲がっていて。

表 7.20 においてベースラインは正解文と同等の翻訳文が得られている。しかし提案手法は "age" が "時代" として翻訳されており、日本語文として意味が通じない。

表 7.21 C ランクにおける提案手法×の翻訳例

入力文	The house burned .
英語文パターン	The X1 X2 .
日本語文パターン	X1 が X2 た。
提案手法	家が焦がした。
ベースライン	その家は燃えていた。
正解文	家が燃えた。

表 7.21 においてベースラインは正解文と同じ翻訳文が得られている。しかし提案手法の翻訳は不適切な文が翻訳されている。

表 7.22 C ランクにおける提案手法×の翻訳例

入力文	He was examined by the police .
英語文パターン	X1 X2 X3 X4 the X5 .
日本語文パターン	X5 X4 X1 が X3 てい X2 。
提案手法	警察で彼が調べていた。
ベースライン	彼は警察の取り調べを受けた。
正解文	彼は警察で取り調べを受けた。

表 7.22 においてベースラインは正解文と同等の翻訳文が得られている。しかし提案手法の翻訳文は文章として不自然な表現である。

7.3.4 D ランクの翻訳例

7.3.4.1 提案手法○の翻訳例

以下に D ランクにおける提案手法○の例を示す。

表 7.23 D ランクにおける提案手法○の翻訳例

入力文	The streets are wet after the rain .
英語文パターン	The X1 X2 X3 X4 the X5 .
日本語文パターン	X1 は X5 X4 X3 て X2 .
正解文	雨のあとで街路がぬれている。
ベースライン	通りにはを潤したを提供してくれた。
提案手法	通りは大雨後ぬれている。

表 7.23 において提案手法は日本語文として意味が通じる翻訳文が得られている。しかしベースラインは助詞が連続して翻訳されており、文として不適切である。

表 7.24 D ランクにおける提案手法×の翻訳例

入力文	I hear a dog howling .
英語文パターン	I X1 X2 X3 X4 .
日本語文パターン	X3 の X4 X2 に X1 た。
提案手法	犬のうなっをに聞こえた。
ベースライン	犬がうなっている。
正解文	犬の遠吠えが聞こえる。

表 7.24 においてベースラインは正解文と同等の翻訳文が得られている。しかし提案手法は日本語文として意味が通じない不適切な文が翻訳されている。

表 7.25 D ランクにおける提案手法×の翻訳例

入力文	Tell me the exact time .
英語文パターン	X1 X2 the X3 X4 .
日本語文パターン	X1 が X3 X4 を X2 た。
提案手法	話しが掴め時間をくれた。
ベースライン	正確な時刻を教えてください。
正解文	正確な時刻を教えてください。

表 7.25 においてベースラインは正解文と同じ翻訳文が得られている。しかし提案手法は不適切な翻訳がされている。

表 7.26 D ランクにおける提案手法×の翻訳例

入力文	I want to travel overseas .
英語文パターン	X1 X2 X3 X4 X5 .
日本語文パターン	X1 は X5 X4 X3 を X2 している。
提案手法	私は 海外 旅行 にを たい している。
ベースライン	海外 旅行 したい。
正解文	海外 旅行 が したい。

表 7.26 においてベースラインは正解文と同じ翻訳文が得られている。しかし提案手法は助詞が連続しており、不適切な翻訳がされている。

7.4 単文データのまとめ

A ランクにおいて、自動評価結果から BLEU 値, NIST 値共に向上し、提案手法の有効性が示された。しかし、B~D ランクにおいて BLEU 値, BIST 値共に低下が認められた。

人手評価結果から、A ランクの翻訳文はベースラインよりも優れている文が多いと言える。一方で B~D ランクの翻訳文はベースラインに劣っている文が多く、A ランクと B~D ランクでは翻訳精度に差異がある。

第8章 重文・複文におけるパターン翻訳

追加実験として、重文・複文を用いたパターン翻訳を行う。

8.1 実験条件

追加実験では、重文・複文データベースから入力テスト文 10,000 文と英日対訳学習文 100,000 を用いる。

8.2 重文・複文データ

重文・複文テストデータ 10,000 に対し英日パターン翻訳を行う。

英日文パターン辞書の作成には $\alpha=0.05$ (16,110 語登録)を用いる。英日対訳学習文 100,000 文に対し 87,674 文のパターン英日文パターンを得た。また、パターン翻訳には $\alpha=0.01$ (49,704 語登録)の英日対訳単語辞書を用いる。

8.3 tri-gram 学習データ

tri-gram 学習データには英日対訳学習文 100,000 文を使用する。そして、単文データを用いたパターン翻訳実験と同様に、翻訳文に閾値を設け、スコア別にランク付けする。閾値、ランクの定義については 6.2 節を参照。

8.4 重文・複文テストデータの自動評価結果

8.4.1 パターンに照合した文のみの評価

入力文 10,000 文に対し、823 文の翻訳文を得た。翻訳文 823 文に対して A ランクは 408 文、B ランクは 31 文、C ランクは 16 文、D ランクは 368 文であった。評価結果を表 8.1 に示す。

表 8.1 各ランクの翻訳文における自動評価結果

	BLEU	NIST
A ランク (408 文)	0.5662	7.7735
B ランク (31 文)	0.4717	4.8247
C ランク (16 文)	0.3517	3.6805
D ランク (368 文)	0.0710	2.3499
A~D ランク (823 文)	0.3630	5.6752

8.4.2 A ランクとベースラインの比較

重文・複文において，パターン翻訳を用いて翻訳された A ランク 408 文と，ベースラインシステムで翻訳された 408 文の BLEU 値の比較結果を表 8.2 に示す。

表 8.2 A ランクの翻訳文とベースラインの比較

	BLEU	NIST
提案手法	0.5662	7.7735
ベースライン	0.5348	7.4597

パターン翻訳を用いて翻訳された B ランク 31 文と，ベースラインシステムで翻訳された 31 文の BLEU 値と NIST 値の比較結果を表 8.3 に示す。

表 8.3 B ランクの翻訳文とベースラインの比較

	BLEU	NIST
B ランク	0.4717	4.8247
ベースライン	0.3573	3.9476

パターン翻訳を用いて翻訳された C ランク 16 文と，ベースラインシステムで翻訳された 16 文の BLEU 値と NIST 値の比較結果を表 8.4 に示す。

表 8.4 C ランクの翻訳文とベースラインの比較

	BLEU	NIST
C ランク	0.3517	3.6805
ベースライン	0.2115	2.8721

パターン翻訳を用いて翻訳された D ランク 368 文と、ベースラインシステムで翻訳された 368 文の BLEU 値と NIST 値の比較結果を表 8.5 に示す。

表 8.5 D ランクの翻訳文とベースラインの比較

	BLEU	NIST
D ランク	0.0710	2.3499
ベースライン	0.1451	3.0121

上記の結果から、A ランクはベースラインと比較して BLEU 値で 3.1%向上している。また、B ランクは BLEU 値で 11.4%、C ランクは BLEU 値で 14.02%の向上が認められた。D ランクはベースラインと比較して BLEU 値の低下が認められたが、重文・複文におけるパターン翻訳は A~C ランクの翻訳文に関して高い翻訳精度が得られた。

8.4.3 10000 文での自動評価結果

提案手法は表 8.1 の A ランクの翻訳文 (408 文) と、ベースラインから得た A ランクを除く 9,592 文の翻訳文を合わせた 10,000 文を評価に用いる。そしてベースラインと比較する。

表 8.6 提案手法とベースラインの比較

	BLEU	NIST
提案手法	0.0999	3.1267
ベースライン	0.0987	3.1100

表 8.6 において、ベースラインと比較して提案手法は BLEU 値が 0.12%向上した。

8.5 重文・複文テストデータの人手評価結果

表 8.1 の翻訳に対し，7.2 節の評価基準を用いて人手評価を行う。尚，重文・複文は単文に比べ英日文パターンに照合した文数が少ないためランダムに抽出した 10 文を用いて人手による比較を行う。

以下に重文・複文における人手評価の結果を示す。

表 8.7 10 文における人手評価結果

ランク	提案手法○	提案手法×	差なし	同一出力
A ランク	3	0	0	7
B ランク	4	0	5	1
C ランク	1	2	6	1
D ランク	0	2	8	0

表 8.7 より，A ランクと B ランクは提案手法○が多く，提案手法の有効性が認められた。C ランクと D ランクは提案手法×の割合が多く，提案手法の有効性は認められなかった。

8.5.1 A ランクの翻訳例

以下に A ランクにおける提案手法○の例を示す。

表 8.8 A ランクにおける提案手法○の翻訳例

入力文	The waves break on the rocks .
英語文パターン	The X1 break on the X2 .
日本語文パターン	X1 が X2 に 当たって 碎ける。
提案手法	波 が 岩 に 当たって 碎ける。
ベースライン	波 が 岩 に 当たった。
正解文	波 が 岩 に 当たって 碎ける。

表 8.8 において提案手法は正解文と同じの意味を持つ翻訳結果が得られている。一方ベースラインの翻訳は意味として通じるが，“碎ける”が翻訳されておらず，提案手法の翻訳文に比べて劣っていると言える。

表 8.9 A ランクにおける提案手法○の翻訳例

入力文	You may keep the money .
英語文パターン	You may keep X1 X2 .
日本語文パターン	きみは X1 X2 を持っていてよろしい。
提案手法	きみはその金を持っていてよろしい。
ベースライン	君の好きなようにその金をもっている。
正解文	その金を上げるから取っておきなさい。

表 8.9 において提案手法は正解文と同じの意味を持つ翻訳結果が得られている。一方ベースラインの翻訳は意味として通じるが，“may” の意味が翻訳されておらず，提案手法の翻訳文に比べて劣っていると言える。

表 8.10 A ランクにおける提案手法○の翻訳例

入力文	It is foolish to take a hit-or-miss attitude toward exams .
英語文パターン	It X1 foolish to take a hit-or-miss attitude toward exams .
日本語文パターン	試験で一か八かやるの X1 ばかげている。
提案手法	試験で一か八かやるのはばかげている。
ベースライン	で一か八かやるのはばかげている。
正解文	試験で山をかけるのはばかげている。

表 8.10 において提案手法は正解文と同じの意味を持つ翻訳結果が得られている。一方ベースラインの翻訳は語頭が助詞から始まっており，不適切な翻訳である。

8.5.2 B ランクの翻訳例

8.5.2.1 提案手法○の翻訳例

以下に B ランクにおける提案手法○の例を示す。

表 8.11 B ランクにおける提案手法○の翻訳例

入力文	A man was struck dead by lightning .
英語文パターン	A X1 was struck dead by lightning .
日本語文パターン	一人の X1 が雷に打たれて死んだ。
提案手法	一人の男が雷に打たれて死んだ。
ベースライン	一人の男がは雷に打たれて死んだ。
正解文	人が雷に打たれて死んだ。

表 8.11 において提案手法は正解文と同等の意味を持つ出力が得られている。しかしベースラインの出力文は助詞が連続して翻訳されており、文として不適切である。

表 8.12 B ランクにおける提案手法○の翻訳例

入力文	The inscription is worn away and not to be deciphered .
英語文パターン	The X1 is worn away and not to be X2 .
日本語文パターン	X1 が磨滅して X2 。
提案手法	碑文が磨滅して見にくい。
ベースライン	その inscription が磨滅して見にくい。
正解文	碑銘が磨滅して読めない。

表 8.12 において提案手法は正解文と同等の意味を持つ出力が得られている。しかしベースラインの出力文は未知語がそのまま出力されており、文として不適切である。

表 8.13 B ランクにおける提案手法○の翻訳例

入力文	He is a rude fellow .
英語文パターン	He is a X1 X2 .
日本語文パターン	X1 のある X2 だ。
提案手法	失礼のある奴だ。
ベースライン	彼はは乱暴な男だ。
正解文	あいつは無作法なやつだ。

表 8.13 において提案手法は正解文と同等の意味を持つ出力が得られている。しかしベースラインの出力文は助詞が連続して翻訳されており、文として不適切である。

8.5.3 Cランクの翻訳例

8.5.3.1 提案手法○の翻訳例

以下にCランクにおける提案手法○の例を示す。

表 8.14 Cランクにおける提案手法○の翻訳例

入力文	He demonstrated that the world is round .
英語文パターン	X1 demonstrated that X2 X3 X4 round .
日本語文パターン	X1 X4 X3 が丸いということ X2 証明した。
提案手法	彼は世界が丸いということを証明した。
ベースライン	世界はであることがわかったのではないかと心配している。
正解文	彼は地球は丸いことを実証した。

表 8.14 において提案手法は正解文と同等の出力が得られている。しかしベースラインは文として意味が通じず不適切である。

8.5.3.2 提案手法×の翻訳例

以下にCランクにおける提案手法×の例を示す。

表 8.15 Cランクにおける提案手法×の翻訳例

入力文	It is ten minutes ' walk from here to the station .
英語文パターン	It is X1 X2 ' walk X3 X4 to the X5 .
日本語文パターン	X4 X3 X5 まであるいて X1 X2 です。
提案手法	ここから駅まであるいて 0 分です。
ベースライン	ここから駅までは歩いて 1 0 分です。
正解文	ここから駅まで歩いて 1 0 分です。

表 8.15 においてベースラインは正解文と同等の出力が得られている。しかし提案手法は数字が適切に翻訳されず、文として意味が通じない。

表 8.16 C ランクにおける提案手法×の翻訳例

入力文	He was doomed to failure .
英語文パターン	He X1 X2 X3 X4 .
日本語文パターン	X4 X3 X2 て X1 た。
提案手法	失敗 する 運命 て いた。
ベースライン	彼は 結局 失敗 する 運命 に 遭った。
正解文	彼は 命数が 尽きて 失敗 した のだ。

表 8.16 においてベースラインは正解文と同等の出力が得られている。しかし提案手法は”運命 て いた” と不自然な表現となっている。

8.5.4 D ランクの翻訳例

以下に D ランクにおける提案手法×の例を示す。

表 8.17 D ランクにおける提案手法×の翻訳例

入力文	She is a good writer .
英語文パターン	X1 X2 a X3 X4 .
日本語文パターン	X1 X2 異国 ふう に X3 X4 だ。
提案手法	彼女は 異国 ふう に 上手 作家 だ。
ベースライン	彼女は 立派 な 作家 だ。
正解文	彼女は 文章 を 書く の が 上手 だ。

表 8.17 においてベースラインの翻訳文は正解文と異なるが日本語として自然な翻訳である。しかし提案手法は文として意味が通じず日本語文として不適切である。

表 8.18 D ランクにおける提案手法×の翻訳例

入力文	The night passed quietly .
英語文パターン	X1 X2 X3 X4 .
日本語文パターン	X1 X2 X4 では X3 。
提案手法	その夜 静かでは 通り過ぎ。
ベースライン	夜が過ぎた。
正解文	その晩は何事もなくすぎた。

表 8.18 においてベースラインの翻訳文は ”quietly” が翻訳されていないが、が日本語として自然な翻訳である。しかし提案手法は文として意味が通じず日本語文として不適切である。

8.6 重文・複文データのまとめ

A ランク～C ランクにおいて、自動評価結果から BLEU 値, NIST 値共に向上し、提案手法の有効性が示された。

人手評価結果から、A, B ランクの翻訳文はベースラインよりも優れている文が多いと言える。一方で C ランクと D ランクの翻訳文はベースラインの翻訳に劣っている文が多く、A, B ランクと C, D ランクでは翻訳精度に差異がある。A, B ランクに関して人手評価からも提案手法の有効性が示された。

第9章 考察

9.1 単文におけるパターン翻訳の翻訳精度

パターン翻訳と tri-gram を組み合わせた手法から高い精度の翻訳文が得られた。しかし、A ランクとして得た翻訳文は 10,000 文中 379 文と少ない。この原因として tri-gram 学習データ数が不足していることが挙げられる。よって、tri-gram 学習データに新たなコーパスを与えることで A ランクの翻訳文数を増加できると考えられる。

9.2 単文における A ランクの翻訳文の誤訳

パターン翻訳の結果から、A ランクにおける翻訳文 379 文中、誤訳が 5 文あった。以下に A ランクに分類されているが、翻訳文として不適切な文章を示す。

表 9.1 A ランクにおける誤訳例 1

入力文	The accident claimed twenty lives .
英語文パターン	The X1 claimed X2 lives .
日本語文パターン	X1 で X2 人が死亡した。
提案手法	事故で 2 人が死亡した。
正解文	その事故は 20 人の命を奪った。

表 9.1 において、英単語 "twenty" が "20" ではなく "2" と翻訳されている。英単語が誤訳された原因として、英日対訳単語辞書の "twenty" に "20" という意味が与えられていないことがわかった。これは現時点で用いている単語辞書が単語対応であることに起因する。よって数字や固有名詞は単語対応ではなく、句対応の単語辞書を用いるという対策が必要であると考えている。

表 9.2 A ランクにおける誤訳例 2

入力文	We cleared the forest .
英語文パターン	X1 X2 X3 X4 .
日本語文パターン	X1 は X4 に X3 X2 た。
提案手法	私は 森 には 晴れた。
正解文	私たちは 森 を 切り開いた。

表 9.2 において，変数 X2 に当たる "cleared" が "晴れ" に翻訳されている．これは表 9.1 と同様に "cleared" に対応する日本語単語が "晴れ" のみであったことが原因である．

しかし表 9.2 で着目すべき点は，日本語文パターンの動詞が変数化されているということである．表 9.1 では名詞のみが変数化されており，単語辞書に未知語が含まれているが，日本語文として自然な文章である．もし，9.2 の "cleared" が変数化されていなければ，日本語文パターンには，"切り開いた。" という動詞が残される．文法情報が多く含まれる文パターンが生成されるため，翻訳精度が向上すると考えている．また，英語文パターンに "clear" が残されることで適合する文パターン数が減少し，不適切な翻訳候補文数の減少が見込まれると考えられる．動詞を含まない文パターンについては 9.7.1 節で記述する．

表 9.3 A ランクにおける誤訳例 3

入力文	He pulled her hand .
英語文パターン	X1 X2 her X3 .
日本語文パターン	X1 は X3 を X2 た。
提案手法	彼は手を引っ込めた。
正解文	彼女が彼女の手を引っ張った。

表 9.3 において，変数 X2 に当たる "pulled" が "引っ込め" に翻訳されている．英日対訳単語辞書には "pulled" の意味に "引っ張る" という意味も含まれていたが，tri-gram による絞込みにより除外されたと考えられる．これは tri-gram の学習データ量を増加させることで改善できると考えている．また，表 9.2 と同様に動詞が変数化されているため，動詞の文法情報を残しておくことで，翻訳精度の向上が見込める．

表 9.4 A ランクにおける誤訳例 4

入力文	Give me a hand .
英語文パターン	Give X1 a X2 .
日本語文パターン	X1 に X2 をあげなさい。
提案手法	私に手をあげなさい。
正解文	手伝ってください。

表 9.4 において、提案手法の翻訳文は日本語文として不自然である。これは適合した文パターンそのものに問題がある。適合した文パターンの原文を調べたところ、英語文は "Give her a doll ."，対応する日本語文は "彼女に人形をあげなさい。" であった。翻訳精度を向上させるために、文パターン辞書のデータ数を増加する必要がある。

表 9.5 A ランクにおける誤訳例 5

入力文	My brother is in Tokyo .
英語文パターン	X1 X2 is in X3 .
日本語文パターン	X1 の X2 は X3 にある。
提案手法	私の兄は東京にある。
正解文	兄が上京している。

表 9.5 において、提案手法は "兄は東京にある" と翻訳されている。これは日本語文として表現が不適切である。これは先ほどと同様に、適合した文パターンに問題がある。よって、文パターン辞書のデータ数を増加しなければならないと考えている。

9.3 単文における類似文パターンの調査

入力文に類似するが、文パターンに一致しない文章を調査した。入力文と一致しなかった文パターンのうち、特定の 1 単語に対し修正をすることで、入力文が文パターンに一致する文を類似文とする。文パターンと一致しなかった入力文 6748 文からランダムに 50 文を抽出し調査を行った。うち、類似文が 2 文、一致するが単語辞書未登録であったため翻訳されなかった文が 7 文であった。以下に類似文の例を示す。

表 9.6 類似文の例 1

入力文	He made his way along the cliff .
類似文	He made his way down the X1 . He X1 his X2 in the X3 .

表 9.6 は類似文の "down" と "in" がそれぞれ "along" もしくは変数であればパターンに適合する。

表 9.7 類似文の例 2

入力文	He is out of his mind .
類似文	X1 X2 out of the X3 .

表 9.7 は類似文の "the" が "his" もしくは変数であればパターンに適合する。
以下に単語辞書が原因の類似文の例を示す。

表 9.8 類似文の例 3

入力文	He is always fidgety .
類似文	X1 X2 X3 X4 .

表 9.9 類似文の例 4

入力文	the dinner party was absolutely deadly .
類似文	X1 X2 X3 X4 X5 X6 .

表 9.8 は fidgety が、表 9.9 は absolutely が単語辞書未登録であったため翻訳されなかった。

9.4 重文・複文における翻訳精度の考察

重文・複文において、A～C ランクの翻訳精度が向上した理由として、単文に比べて英日文パターン辞書と英日対訳単語辞書に用いた単語数が少なかったことが挙げられる。特に英日文パターン辞書の作成段階で、文パターン中に変数化された箇所が少なく、文字情報が多く残されていた可能性がある。よって入力文が英語文パターンと一致した際に正解文と同等の翻訳文が多く得られたと考えられる。また、重文・複文において、A ランクの誤訳は0文であった。

9.5 重文・複文における類似文の調査

9.3 節と同様に類似文の調査を行う。文パターンと一致しなかった入力文 9177 文からランダムに 50 文を抽出し調査した。うち、文パターンに一致するが単語辞書未登録であったため翻訳されなかった文が6文であった。以下に類似文の例を示す。

表 9.10 類似文の例 1

入力文	The leaves shrivel in cold weather .
類似文	The X1 X2 X3 X4 X5 .

表 9.11 類似文の例 2

入力文	His mouth twisted into a grin .
類似文	X1 X2 X3 X4 X5 X6 .

表 9.10 は shrivel が、表 9.11 は into が単語辞書未登録であったため翻訳されなかった。

9.6 翻訳精度とカバー率

パターン翻訳において、翻訳精度の性能に特化した場合、カバー率が低下する。また逆に、カバー率の向上に努めた場合、翻訳精度が低下するという問題が生じる。

今後はこの問題に対し、9.3 節と 9.5 節で述べた結果から、試験的にパターン翻訳に用いる単語数を増加させ、翻訳精度とカバー率へ与える影響の調査を行う。

9.7 翻訳精度向上の手段

9.7.1 意味制約を用いた辞書

4.1 節で述べたように、意味制約を用いた日英文パターン辞書の研究が過去に行われたが、本研究では文パターンに意味制約を付与していない。また、翻訳実験の結果から、英日文パターンにおいて動詞を変数化すると、表 9.2 のように不適切な位置に動詞が置かれ、翻訳精度が低下することがわかった。

そこで、パターン翻訳の翻訳精度向上のため、動詞を含まない英日対訳単語辞書を提案する。この単語辞書を用いることで、作成される英日文パターンの動詞は変数化されない。よって、動詞を含まない英日対訳単語辞書を用いた英日文パターンは、本研究の手法で得られた英日文パターンに比べて、文法情報がより多く含まれていると考えられる。このため、出力される翻訳文に対しても翻訳精度の向上が見込める。ただし、英日文パターンに含まれる変数が減少するため、パターンに一致する入力文数が減少する可能性がある。そのため、カバー率に対しても対策が必要である。

9.7.2 英日対訳単語辞書の翻訳確率

4.2 節でも述べたが、本研究ではパターン翻訳における単語の翻訳に翻訳確率を用いていない。そこで、今後はパターン翻訳に翻訳確率を付与した場合、翻訳精度にどのような影響を及ぼすかをについても調査していく。

9.8 日英方向におけるパターン翻訳との比較

過去に西村らは 1 節の先行手法を用いて日英翻訳実験を行った。単文テストデータ 10,000 文に対し、閾値 0.01(53,185 語登録) の日英対訳単語辞書を用いた日英文パターン辞書と閾値 0.02(34,587 語登録) の日英対訳単語辞書を用いて先行手法を行い、2,103 文が日本語文パターンに照合した。

文パターンに照合した文数を比較すると、先行手法が 2,103 文であるのに対し本実験では 3252 文と英日翻訳の方が 1,000 文以上多く文パターンに照合している。先行手法と本実験に使用した文パターン辞書と対訳単語辞書の閾値および登録単語数が同じでないため単純な比較はできないが、日英翻訳に比べ英日翻訳の方がパターンに照合し易いと考えられる。

この理由として英語と日本語間での表現方法の違いが挙げられる。例えば英文 "I climb the mountain ." は単語を入れ替えて "I the climb mountain." と表現すると文章として成立しなくなる。しかし日本語文は "私は山に登る。" の単語を入れ替えて "山に私は登る。" と表現しても文章として成立する。このように日本語文は語順が自由であり、複数の表現が生成されるため文パターンに照合しにくい。一方で英文は表現方法が限られているため英語文パターンに照合し易かったと思われる。

9.9 単語に基づく統計翻訳器の検討

統計翻訳の初期は単語に基づく統計翻訳であった。しかし単語に基づく統計翻訳は句に基づく統計翻訳に比べて翻訳精度が低い。そのため近年は句に基づく統計翻訳が主流になっている。

本実験のパターン翻訳は英日対訳単語辞書と対訳パターンが翻訳モデルに、tri-gram が言語モデルに相当すると考えることができる。よって実験で用いたパターン翻訳器は単語に基づく統計翻訳のデコーダとして考えることができる。

第10章 おわりに

本研究では，文パターンの自動作成と tri-gram による翻訳候補の絞込みを組み合わせた手法から，翻訳文を生成した．自動評価の結果，パターン翻訳と tri-gram を組み合わせることにより，条件付きながら，単文において BLEU 値で 2.3%，NIST 値で 0.2 の翻訳精度の向上が認められた．また，重文・複文においても，BLEU 値で 3.2%，NIST 値で 0.3 の翻訳精度の向上が認められた．よって，提案手法は有効であるといえる．tri-gram の学習データはモノリンガルであるため，今後は学習データに新聞記事コーパスを与え翻訳結果への影響を調査する．

謝辞

最後に、一年間に渡り、本研究の御指導をいただきました鳥取大学工学部知能情報工学科計算機講座C研究室の村田真樹教授，村上仁一准教授，徳久雅人講師に深く感謝するとともに厚くお礼を申し上げます。また，計算機工学講座C研究室の皆様に厚くお礼を申し上げます。また，参考にさせて頂いた論文の著者の方々に対して，深く感謝します。

参考文献

- [1] 西村拓哉, 村上仁一, 徳久雅人, 池原悟, ”文パターンを用いた統計翻訳”, 言語処理学会第 16 回年次大会, pp.676-679. 2010.
- [2] GIZA++, Training of statistical translation models
<http://www.fjoch.com/GIZA++>
- [3] 村上仁一, 徳久雅人, ”日英対訳データベース作成のための 1 考察”, 言語処理学会第 17 回年次大会, pp.979-982, 2011.
- [4] Philipp Koehn, Marcello Federico, Brooke Cowan, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, Evan Herbst, “Moses: Open Source Toolkit for Statistical Machine Translation”, Proceedings of the ACL 2007 Demo and Poster Sessions, pp.177-180, 2007.
- [5] Peter F.Brown, Stephen A.Della Pietra, Vincent J.Della Pietra, Robert L.Mercer, “The mathematics of statistical machine translation: Parameter Estimation”, Computational Linguistics, pp.263-311, 1993.
- [6] MeCab, 日本語形態素解析器
<http://mecab.sourceforge.net/>
- [7] SRILM, The SRI Language Model Toolkit
<http://www.speech.sri.com/projects/srilm/>
- [8] BLEU, NIST Open MT Scoring
<http://www.itl.nist.gov/iad/894.01/tests/mt/>
- [9] NIST, Automatic Evaluation of Machine Translation Quality Using n-gram Co-Occurrence Statistics
<http://www.itl.nist.gov/iad/mig/test/mt/>

- [10] Franz Josef Och, “Minimum error rate training for statistical machine translation”, Proceedings of the ACL, 2003.
- [11] SRILM, The SRI Language Modeling Toolkit
<http://www.speech.sri.com/projects/srilm/>
- [12] Jin’ichi Murakami, Masato Tokuhisa, Satoru Ikehara, “Statistical Machine Translation using Large J/E Parallel Corpus and Long Phrase Tables”, International Workshop on Spoken Language Translation 2007, pp.151-155, 2007
- [13] 西山七絵, 村上仁一, 徳久雅人, 池原悟, “単文句型パターン辞書の構築”, 言語処理学会第 11 回年次大会, pp.372-375, 2005