

## 概要

本論文では機械翻訳の一種であるパターン翻訳の研究を行う。パターン翻訳は、入力された原言語文に対し、単語辞書と文パターン(原言語文パターンと目的語文パターンの対で構成)を用いて翻訳文を出力する方法である。パターン翻訳の長所として、入力文が適切な文パターンに適合した場合、高品質な翻訳文を得やすい傾向にある。しかし、パターン翻訳には多くの問題点がある。本研究では中でもコストと翻訳精度とカバー率の問題を取り上げる。まず、パターン翻訳は非常にコストがかかる。この理由は単語辞書と文パターンを人手で作成するためである。本研究では単語辞書と文パターンをプログラムで自動作成することでコストの削減を試みた。また、パターン翻訳は翻訳精度が高いとカバー率が低く、カバー率が高いと翻訳精度が低くなる傾向にある。この理由として、入力文に適合した文パターンが保持する字面が関係していると考えた。つまり、文パターンの字面が多ければ翻訳精度が高く(カバー率が低く)、少なればカバー率が高く(翻訳精度が低く)なる。加えて、カバー率には、入力文が文パターンに適合しなければ翻訳ができない問題がある。

翻訳精度の問題に対しては、入力文と文パターンの字面を比較する。そして字面が多く一致する文パターンを優先して選択する。カバー率の問題に対しては、大量の文パターンと単語辞書を作成する。

上記のコストと翻訳精度とカバー率の問題を考慮し、パターン翻訳を行った。その結果、英語入力文 100 文から 24 文の日本語翻訳文を得た。提案手法で得た日本語翻訳文を評価するために、統計翻訳の世界で主流となっている Moses をベースラインとして翻訳精度の調査を行った。評価方法には自動評価法と人手による対比較評価を用いる。実験の結果、全ての自動評価において提案手法がベースラインよりも優れていることがわかった。また対比較評価においても、提案手法○が 6 文に対して提案手法×が 3 文と提案手法の有効性が確認できた。

# 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	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
3.2	単語に基づく統計翻訳	5
3.2.1	IBM 翻訳モデル	5
3.2.1.1	モデル1	6
3.2.1.2	モデル2	8
3.2.1.3	モデル3	8
3.2.1.4	モデル4	9
3.2.1.5	モデル5	10
3.2.2	単語に基づく統計翻訳の問題点	11
3.2.3	GIZA++	12
3.3	句に基づく統計翻訳	12
3.4	翻訳モデル	13
3.5	フレーズテーブル作成法	14
3.6	言語モデル	18
3.7	デコーダ	19
3.8	パラメータチューニング	20
3.9	評価方法	20
3.9.1	BLEU	20
3.9.2	NIST	21

3.9.3	RIBES . . . . .	21
3.9.4	TER . . . . .	22
3.9.5	WER . . . . .	22
3.9.6	METEOR . . . . .	22
<b>第4章</b>	<b>提案手法</b>	<b>24</b>
4.1	提案手法の概略 . . . . .	24
4.2	提案手法の詳細 . . . . .	25
<b>第5章</b>	<b>実験</b>	<b>39</b>
5.1	実験条件 . . . . .	39
5.1.1	データベース . . . . .	39
5.1.2	句に基づく文パターンを用いた英日翻訳 . . . . .	39
5.1.3	ベースラインシステム . . . . .	41
5.1.3.1	翻訳モデルの学習 . . . . .	41
5.1.3.2	言語モデルの学習 . . . . .	41
5.2	実験結果 . . . . .	41
5.2.1	提案手法より得たデータ数 . . . . .	41
5.2.2	日本語翻訳文の出力例 . . . . .	42
<b>第6章</b>	<b>評価結果</b>	<b>44</b>
6.1	自動評価結果 . . . . .	44
6.2	人手評価結果 . . . . .	44
6.3	人手評価における提案手法○の例 . . . . .	45
6.4	人手評価における提案手法×の例 . . . . .	46
6.5	評価結果のまとめ . . . . .	47
<b>第7章</b>	<b>追加実験</b>	<b>48</b>
7.1	実験条件の変更点 . . . . .	48
7.2	実験結果 . . . . .	48
7.3	評価結果 . . . . .	49
7.3.1	自動評価結果 . . . . .	49
7.3.2	人手評価結果 . . . . .	49
7.3.3	人手評価における提案手法○の例 . . . . .	49

7.3.4	人手評価における提案手法×の例 . . . . .	51
7.3.5	提案手法×における日本語翻訳文の改善例 . . . . .	52
7.3.6	評価結果のまとめ . . . . .	52
<b>第8章</b>	<b>考察</b>	<b>53</b>
8.1	翻訳精度の問題 . . . . .	53
8.2	カバー率の問題 . . . . .	54
8.3	翻訳精度とカバー率の問題 . . . . .	54
<b>第9章</b>	<b>おわりに</b>	<b>55</b>

# 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 . . . . .	12
3.2	デコーダの動作例 . . . . .	19
4.1	英日方向の単語辞書の作成手順 . . . . .	26
4.2	日英方向の単語辞書の作成手順 . . . . .	26
4.3	日英方向の単語辞書の作成手順 . . . . .	27
4.4	単語に基づく文パターン辞書の作成手順 . . . . .	28
4.5	対訳フレーズの抽出手順 . . . . .	29
4.6	英日フレーズ確率の付与手順 . . . . .	30
4.7	日英フレーズ確率の付与手順 . . . . .	31
4.8	単語数の差による抽出 . . . . .	33
4.9	句に基づく文パターン辞書の作成手順 . . . . .	34
4.10	文パターン確率の付与手順 . . . . .	35
4.11	文パターン確率の付与手順 . . . . .	36
4.12	日本語翻訳文出力手順 . . . . .	38

# 表 目 次

2.1	英日パターン翻訳の例 . . . . .	4
3.1	英日方向の単語対応 . . . . .	11
3.2	日英方向の単語対応 . . . . .	11
3.3	英日方向の単語対応 . . . . .	14
3.4	日英方向の単語対応 . . . . .	14
3.5	intersection の例 . . . . .	15
3.6	union の例 . . . . .	15
3.7	grow の例 . . . . .	16
3.8	grow-diag の例 . . . . .	16
3.9	grow-diag の例 . . . . .	17
3.10	grow-diag の例 . . . . .	17
3.11	言語モデルの例 . . . . .	18
5.1	4.2 節で得たデータ数 . . . . .	41
5.2	出力例の詳細 1 . . . . .	42
5.3	変数対応フレーズ . . . . .	42
5.4	文パターンの原文 . . . . .	42
5.5	出力例の詳細 2 . . . . .	43
5.6	変数対応フレーズ . . . . .	43
5.7	文パターンの原文 . . . . .	43
6.1	提案手法とベースラインの自動評価結果 . . . . .	44
6.2	提案手法とベースラインの人手評価結果 . . . . .	45
6.3	提案手法○の例 1 . . . . .	45
6.4	提案手法○の例 2 . . . . .	45
6.5	提案手法○の例 3 . . . . .	46

6.6	提案手法×の例1	46
6.7	提案手法×の例2	47
6.8	提案手法×の例3	47
7.1	7.1節で得たデータ数	48
7.2	提案手法とベースラインの自動評価結果	49
7.3	提案手法とベースラインの人手評価結果	49
7.4	提案手法○の例1	50
7.5	提案手法○の例2	50
7.6	提案手法○の例3	50
7.7	提案手法×の例1	51
7.8	提案手法×の例2	51
7.9	提案手法×の例3	51
7.10	提案手法×の改善例	52
8.1	フレーズの例	53
8.2	文パターンの例	54

# 第1章 はじめに

パターン翻訳は1960年代半ばに提案された機械翻訳手法の一種である。パターン翻訳は、大量の文パターン(原言語文パターンと目的語文パターンの対で構成)と単語辞書を用いて、翻訳文を得る方法である。パターン翻訳の長所として入力文が適切な文パターンに適合した場合に精度の高い翻訳文が得られる。しかし、パターン翻訳には多くの問題点がある。以下にパターン翻訳の問題点と原因を示す。

- コストの問題

文パターンと単語辞書の作成を人手で行うため、コストと時間がかかる。

- 翻訳精度とカバー率の問題

パターン翻訳は一般的に、翻訳精度が高ければカバー率が低く、カバー率が高ければ翻訳精度が低くなる傾向にある。

- カバー率の問題

入力文が文パターンに適合しない場合、翻訳ができない。

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[1]が発表され、機械翻訳の研究は次第に行われなくなった。

また、1990年代前半に単語に基づく統計翻訳が提案されたが、翻訳精度が低くあまり研究がされなかった。しかし、2000年代始めに句に基づく統計翻訳が提案され、単語に基づく統計翻訳と比較して翻訳精度が高いことから、現在は句に基づく統計翻訳が主流となっている。句に基づく統計翻訳は、学習データとして対訳文を与えるだけで翻訳が可能である。このため翻訳にかかるコストは非常に低い。加えて、対訳文から対訳単語と単語翻訳確率の自動取得が可能である。

本研究では統計翻訳の特徴である、対訳文から対訳単語と単語翻訳確率の自動取得が可能である点に着目し、パターン翻訳に用いる単語辞書と文パターンを自動的に作成する。そして、パターン翻訳におけるコストの問題を解決する。

また、パターン翻訳は翻訳精度が高ければカバー率が低く、カバー率が高ければ翻訳精度が低くなる傾向にある。この理由として、入力文に適合した文パターンが保持する



字面が関係していると考えた。具体的には、適合した文パターンが保持する字面の数が多ければ翻訳精度は高くなり、少なければ翻訳精度が低くなるということである。そこで本研究では、翻訳精度の問題に対して、入力文と文パターンの字面を比較する。そして字面が多く一致する文パターンを優先して選択する。

さらに、カバー率には入力文が文パターンに適合しなければ翻訳文が出力されない問題がある。本研究ではカバー率の問題に対して、大量の文パターンと単語辞書を作成することで改善を試みる。

本研究では英日方向のパターン翻訳を行う。プログラムで自動作成した単語辞書と文パターンを用いて得た日本語翻訳文に対して、言語翻訳確率 (tri-gram) を用いた絞込みを行い、出力する日本語翻訳文を決定する。最後に日本語翻訳文に対して自動評価と人手評価を行い提案手法の有効性を調査する。

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章で提案する翻訳システムについて説明する。第5章で実験条件を述べ、第6章で実験結果を示す。第7章で追加実験について説明し、第8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 概要

パターン翻訳 [2] とは，機械翻訳手法の一種である．パターン翻訳は，原言語文と目的言語文の対訳文に対して，任意の単語やフレーズを変数化した“文パターン”と“単語辞書”が必要である．原言語入力文と原言語文パターンを照合し，適合する原言語文パターンに対応する目的言語文パターンを得る．そして，文パターンの変数部に対応する単語やフレーズを，単語辞書を用いて翻訳し，目的言語翻訳文を出力する．

パターン翻訳は適切な文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，問題点として，入力文に対するカバー率が低い．

### 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 文パターン辞書と単語辞書を用意する．

手順2 英語入力文と英語側文パターンを照合する．

手順3 変数部に対応する英単語を単語辞書を用いて日本語単語に翻訳する．

手順4 英語側文パターンに対応する日本語側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で得た日本語翻訳文を出力する．

英日パターン翻訳の手順を図??に示す。

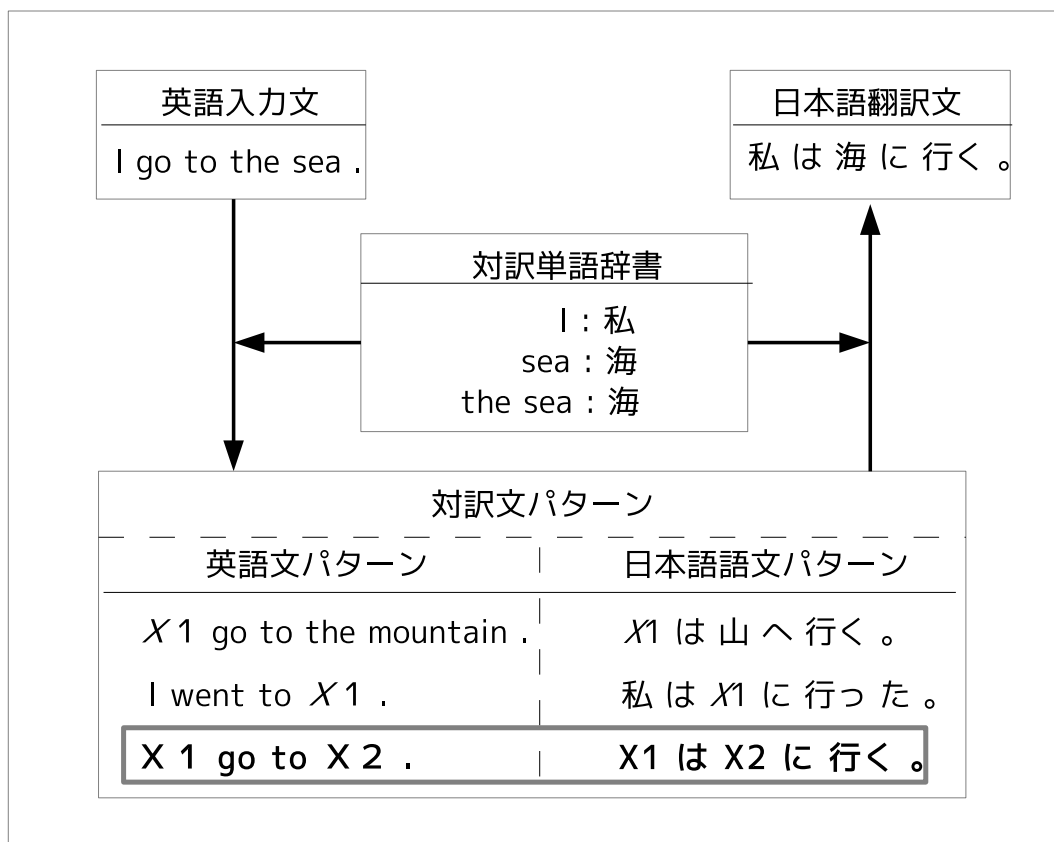


図 2.1 英日パターン翻訳の手順

## 2.3 文パターン辞書

文パターン辞書とは、大量の対訳文から任意の単語やフレーズを変数化して得られる文パターンの集合である。表 2.1 に例を示す。

表 2.1 英日パターン翻訳の例

英語入力文	I go to the sea .
英語文パターン	I go to X1 .
日本語文パターン	私は X1 に 行く 。
日本語翻訳文	私は海に行く。

## 第3章 英日統計翻訳システム

### 3.1 概要

統計翻訳とは、機械翻訳手法の一種である。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。

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。

### 3.2 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。例として、ある英語文を日本語文に翻訳する場合を考える。英単語を日本語に翻訳し、英単語の語順と同じ並びで日本語単語を並べて翻訳する。単語に基づく統計翻訳は単語対応の確率を得る IBM 翻訳モデルが用いられている。

#### 3.2.1 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 [3] 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。この翻訳モデルは順に複雑な計算を行うモデル 1 から 5 の 5 つのモデルで構成される。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。

$$P(F|E) = \sum_a P(F, a|E) \quad (3.1)$$

アライメントとは仏単語と英単語の対応を意味している。IBM モデルのアライメントでは、各仏単語  $f$  に対応する英単語  $e$  は 1 つあり、各英単語  $e$  に対応する仏単語は 0 から  $n$  個ある。また仏単語  $f$  において適切な英単語と対応しない場合、英語文の先頭に空単語  $e_0$  があると仮定し、その仏単語  $f$  と空単語  $e_0$  を対応づける。

### 3.2.1.1 モデル 1

(3.1) 式は以下の式に分解することができる。  $m$  はフランス語文の長さ、  $a_1^{j-1}$  はフランス語文における、1 番目から  $j-1$  番目までのアライメント、  $f_1^{j-1}$  はフランス語文における、1 番目から  $j-1$  番目まで単語を表している。

$$P(F, a|E) = P(m|E) \prod_{j=1}^m P(a_j|a_1^{j-1}, f_1^{j-1}, m, E) P(f_j|a_1^j, f_1^{j-1}, m, E) \quad (3.2)$$

(3.2) 式ではとても複雑であるので計算が困難である。そこで、モデル 1 では以下の仮定により、パラメータの簡略化を行う。

- フランス語文の長さの確率  $\epsilon$  は  $m, E$  に依存しない

$$P(m|E) = \epsilon$$

- アライメントの確率は英語文の長さ  $l$  に依存する

$$P(a_j|a_1^{j-1}, f_1^{j-1}, m, E) = (l+1)^{-1}$$

- フランス語の翻訳確率  $t(f_j|e_{a_j})$  は、仏単語  $f_j$  に対応する英単語  $e_{a_j}$  に依存する

$$P(f_j|a_1^j, f_1^{j-1}, m, e) = t(f_j|e_{a_j})$$

パラメータの簡略化を行うことで、  $P(F, a|E)$  と  $P(F, E)$  は以下の式で表される。

$$P(F, a|E) = \frac{\epsilon}{(l+1)^m} \prod_{j=1}^m t(f_j|e_{a_j}) \quad (3.3)$$

$$P(F|E) = \frac{\epsilon}{(l+1)^m} \sum_{a_1=0}^l \cdots \sum_{a_m=0}^l \prod_{j=1}^m t(f_j|e_{a_j}) \quad (3.4)$$

$$= \frac{\epsilon}{(l+1)^m} \prod_{j=1}^m \sum_{i=0}^l t(f_j|e_{a_j}) \quad (3.5)$$

モデル1では翻訳確率  $t(f|e)$  の初期値が0以外の場合、Expectation-Maximization(EM) アルゴリズムを繰り返し行うことで得られる期待値を用いて最適解を推定する。EM アルゴリズムの手順を以下に示す。

**手順1** 翻訳確率  $t(f|e)$  の初期値を設定する。

**手順2** 仏英対訳対  $(F^{(s)}, E^{(s)})$  (但し,  $1 \leq s \leq S$ ) において, 仏単語  $f$  と英単語  $e$  が対応する回数の期待値を以下の式により計算する。

$$c(f|e; F, E) = \frac{t(f|e)}{t(f|e_0) + \cdots + t(f|e_l)} \sum_{j=1}^m \delta(f, f_j) \sum_{i=0}^l \delta(e, e_i) \quad (3.6)$$

$\delta(f, f_j)$  はフランス語文  $F$  中で仏単語  $f$  が出現する回数,  $\delta(e, e_i)$  は英語文  $E$  中で英単語  $e$  が出現する回数を表している。

**手順3** 英語文  $E^{(s)}$  の中で1回以上出現する英単語  $e$  に対して, 翻訳確率  $t(f|e)$  を計算する。

1. 定数  $\lambda_e$  を以下の式により計算する。

$$\lambda_e = \sum_f \sum_{s=1}^S c(f|e; F^{(s)}, E^{(s)}) \quad (3.7)$$

2. (3.7) 式より求めた  $\lambda_e$  を用いて, 翻訳確率  $t(f|e)$  を再計算する。

$$\begin{aligned} t(f|e) &= \lambda_e^{-1} \sum_{s=1}^S c(f|e; F^{(s)}, E^{(s)}) \\ &= \frac{\sum_{s=1}^S c(f|e; F^{(s)}, E^{(s)})}{\sum_f \sum_{s=1}^S c(f|e; F^{(s)}, E^{(s)})} \end{aligned} \quad (3.8)$$

**手順4** 翻訳確率  $t(f|e)$  が収束するまで手順2と手順3を繰り返す。

### 3.2.1.2 モデル2

モデル1では、全ての単語の対応に対して、英語文の長さ  $l$  にのみ依存し、単語対応の確率を一定としている。そこで、モデル2では、 $j$  番目の仏単語  $f_j$  と対応する英単語の位置  $a_j$  は英語文の長さ  $l$  に加えて、 $j$  と、フランス語文の長さ  $m$  に依存し、以下のような関係とする。

$$a(a_j|j, m, l) \equiv P(a_j|a_1^{j-1}, f_1^{j-1}, m, l) \quad (3.9)$$

この関係からモデル1における (3.4) 式は、以下の式に変換できる。

$$P(F|E) = \epsilon \sum_{a_1=0}^l \cdots \sum_{a_m=0}^l \prod_{j=1}^m t(f_j|e_{a_j}) a(a_j|j, m, l) \quad (3.10)$$

$$= \epsilon \prod_{j=1}^m \sum_{i=0}^l t(f_j|e_{a_j}) a(a_j|j, m, l) \quad (3.11)$$

モデル2では、期待値は  $c(f|e; F, E)$  と  $c(i|j, m, l; F, E)$  の2つが存在する。以下の式から求められる。

$$c(f|e; F, E) = \frac{t(f|e)}{t(f|e_0) + \cdots + t(f|e_l)} \sum_{j=1}^m \delta(f, f_j) \sum_{i=1}^l \delta(e, e_i) \quad (3.12)$$

$$= \sum_{j=1}^m \sum_{i=0}^l \frac{t(f|e) a(i|j, m, l) \delta(f, f_j) \delta(e, e_i)}{t(f|e_0) a(0|j, m, l) + \cdots + t(f|e_l) a(l|j, m, l)} \quad (3.13)$$

$$c(i|j, m, l; F, E) = \sum_a P(a|E, F) \delta(i, a_j) \quad (3.14)$$

$$= \frac{t(f_j|e_i) a(i|j, m, l)}{t(f_j|e_0) a(0|j, m, l) + \cdots + t(f_j|e_l) a(l|j, m, l)} \quad (3.15)$$

$c(f|e; F, E)$  は対訳文中の英単語  $e$  と仏単語  $f$  が対応付けされる回数の期待値、 $c(i|j, m, l; F, E)$  は英単語の位置  $i$  が仏単語の位置  $j$  に対応付けされる回数の期待値を表している。

モデル2では、EMアルゴリズムで計算すると複数の極大値が算出され、最適解が得られない可能性がある。モデル1では  $a(i|j, m, l) = (l+1)^{-1}$  となるモデル2の特殊な場合であると考えられる。したがって、モデル1を用いることで最適解を得ることができる。

### 3.2.1.3 モデル3

モデル3は、モデル1とモデル2とは異なり、1つの単語が複数対応する単語の繁殖数や単語の翻訳位置の歪みについて考慮する。またモデル3では単語の位置を絶対位置と

して考える。モデル3では以下のパラメータを用いる。

- 翻訳確率  $P(f|e)$   
英単語  $e$  が仏単語  $f$  に翻訳される確率
- 繁殖確率  $n(\phi|e)$   
英単語  $e$  が  $\phi$  個の仏単語と対応する確率
- 歪み確率  $d(j|i, m, l)$   
英語文の長さ  $l$ 、フランス語文の長さ  $m$  のとき、 $i$  番目の英単語  $e_i$  が  $j$  番目の仏単語  $f_j$  に翻訳される確率

さらに、英単語が仏単語に翻訳されない個数を  $\phi_0$  とし、その確率  $p_0$  を以下の式で求める。このとき、歪み確率は  $\frac{1}{\phi_0!}$  で、 $p_0 + p_1 = 1$  で  $p_0, p_1$  は0より大きいとする。

$$P(\phi_0|\phi_1^l, E) = \binom{\phi_1 + \dots + \phi_l}{\phi_0} p_0^{\phi_1 + \dots + \phi_l - \phi_0} p_1^{\phi_0} \quad (3.16)$$

したがって、モデル3は以下の式で求められる。

$$\begin{aligned} P(F|E) &= \sum_{a_1=0}^l \dots \sum_{a_m=0}^l P(F, a|E) \quad (3.17) \\ &= \sum_{a_1=0}^l \dots \sum_{a_m=0}^l \binom{m - \phi_0}{\phi_0} p_0^{m - 2\phi_0} p_1^{\phi_0} \prod_{i=1}^l \phi_i! n(\phi_i|e_i) \\ &\quad \times \prod_{j=1}^m t(f_j|e_{a_j}) d(j|a_j, m, l) \quad (3.18) \end{aligned}$$

モデル3では、全てのアライメントを計算するため、計算量が膨大となるので期待値を近似により求める。

### 3.2.1.4 モデル4

モデル4では、モデル3と異なり、単語の位置を絶対位置ではなく、相対位置で考える。またモデル3では考慮されていない各単語の位置、例えば形容詞と名詞の関係を考慮する。モデル4では歪み確率  $d(j|i, m, l)$  を2つの場合で考える。



- 繁殖数が1以上である英単語に対応する仏単語の中で、最も文頭に近い場合

$$P(\Pi_{[i]1} = j | \pi_1^{[i]-1}, \tau_0^l, \phi_0^l, E) = d_1(j - \odot_{i-1} | \mathcal{A}(e_{[i-1]}), \mathcal{B}(f_j)) \quad (3.19)$$

$\odot_{i-1}$  は  $i-1$  番目の英単語に対応する仏単語の位置を表している。

- それ以外の場合

$$P(\Pi_{[i]k} = j | \pi_{[i]1}^{k-1}, \pi_1^{[i]-1}, \tau_0^l, \phi_0^l, E) = d_{>1}(j - \pi_{[i]k-1} | \mathcal{B}(f_j)) \quad (3.20)$$

$\pi_{[i]k-1}$  は同じ英単語に対応している直前の仏単語を表している。

### 3.2.1.5 モデル5

モデル4では、単語の位置に関して直前の単語以外は考慮されていない。したがって、複数の単語が同じ位置に生じたり、単語の存在しない位置が生成される。モデル5では、この問題を避けるために、単語を空白部分に配置するよう改善が施されている。

- 繁殖数が1以上である英単語に対応する仏単語の中で、最も文頭に近い場合

$$\begin{aligned} P(\Pi_{[i]1} = j | \pi_1^{[i]-1}, \tau_0^l, \phi_0^l, E) \\ = d_1(v_j | \mathcal{B}(f_j), v_{\odot_{i-1}}, v_m - \phi_{[i]} + 1)(1 - \delta(v_j, v_{j-1})) \end{aligned}$$

$v_j$  は  $j$  番目までの空白数、 $\mathcal{A}$  は英語の単語クラス  $\mathcal{B}$  はフランス語の単語クラスを表している。

- それ以外の場合

$$\begin{aligned} P(\Pi_{[i]k} = j | \pi_{[i]1}^{k-1}, \pi_1^{[i]-1}, \tau_0^l, \phi_0^l, E) \\ = d_{>1}(v_j - v_{\pi_{[i]k-1}} | \mathcal{B}(f_j), v_m - v_{\pi_{[i]k-1}} - \phi_{[i]} + k)(1 - \delta(v_j, v_{j-1})) \end{aligned}$$

### 3.2.2 単語に基づく統計翻訳の問題点

以下に，IBM 翻訳モデルを用いて得た英日方向における単語対応の例と，日英方向における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 は日本語単語“は”と“に”に対応する英単語が存在しない．一方で，表 3.2 は全ての単語に対して対応がとれている．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．このため単語翻訳確率の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(句に基づく統計翻訳)が行われている．

### 3.2.3 GIZA++

GIZA++ [4] とは，統計翻訳で用いることを前提に作られたツールである．IBM 翻訳モデルを用いて，対訳文 (原言語文と目的言語文の対) から対訳単語と単語翻訳確率を自動的に得る．

## 3.3 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．原言語文を目的言語文に翻訳する場合に，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．本研究では英日方向の翻訳を行うため，英日統計翻訳を説明する．英日統計翻訳システムの枠組みを図 3.1 に示す．

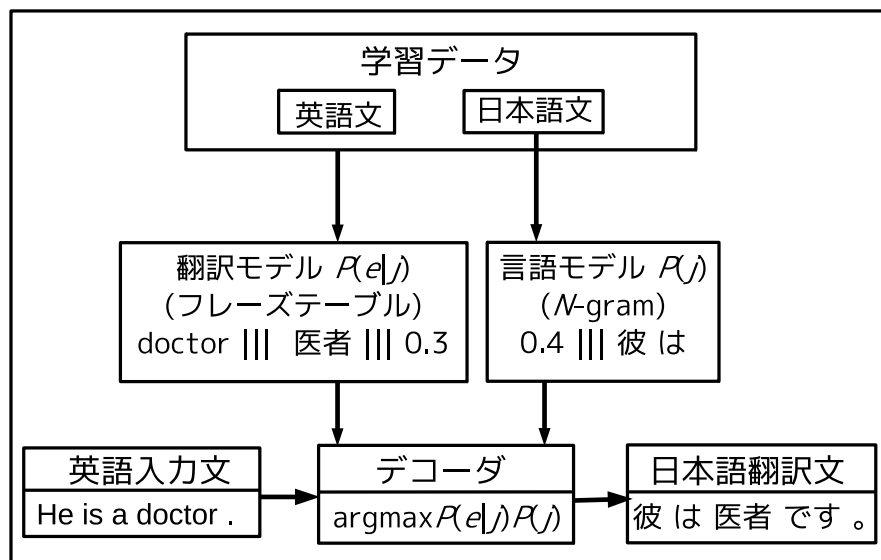


図 3.1 英日統計翻訳の枠組み

英日統計翻訳は、英語入力文  $E$  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翻訳文  $J$  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。

$$J = \operatorname{argmax}_e P(j|e) \quad (3.21)$$

$$\simeq \operatorname{argmax}_e P(e|j)P(j) \quad (3.22)$$

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文(英語文と日本語文の対)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、翻訳モデルと言語モデルを学習する。

### 3.4 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。

### 3.5 フレーズテーブル作成法

まず，GIZA++を用いて学習文から英日，日英方向の双方向で最尤な単語アライメントを得る．英日方向の単語対応の例を表 3.3，日英方向の単語対応の例を表 3.4 に示す．また，●は単語が対応した箇所を示す．

表 3.3 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 3.4 日英方向の単語対応

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

次に，得られた双方向の単語アライメントを用いて，複数単語のアライメントを得る．このアライメントは双方向の単語対応の和集合と積集合から求める．ヒューリスティックスとして双方向ともに対応する単語対応を用いる“intersection”，双方向のどちらか一方でも対応する単語対応を全て用いる“union”がある．表 3.3 と表 3.4 を用いた“intersection”の例を表 3.5，に“union”の例を表 3.6 に示す．

表 3.5 intersection の例

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

表 3.6 union の例

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

また “intersection” と “union” の中間のヒューリスティックスとして “grow” と “grow-diag” がある。これら 2 つのヒューリスティックスでは “intersection” の単語対応と “union” の単語対応を用いる。“grow” は縦横方向，“grow-diag” は縦横対角方向に，“intersection” の単語対応から “union” の単語対応が存在する場合にその単語対応も用いる。“grow” の例を表 3.7 に，“grow-diag” の例を表 3.8 に示す。

表 3.7 grow の例

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

表 3.8 grow-diag の例

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

“grow-diag” の最後に行う処理として “final” と “final-and” がある．“final” は少なくとも片方の言語の単語対応がない場合に，“union” の単語対応を追加する．また，“final-and” は，両側言語の単語対応がない場合に，“union” の候補対応点を追加する．“grow-diag-final” の例を表 3.9 に，“grow-diag-final-and” の例を表 3.10 に示す．

表 3.9 grow-diag の例

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

表 3.10 grow-diag の例

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

得られた単語アライメントから、全ての矛盾しないフレーズ対を得る。このとき、そのフレーズ対に対して翻訳確率を計算し、フレーズ対に確率値を付与することでフレーズテーブルを作成する。



### 3.6 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 3.11 言語モデルの例

tri-gram の単語列	$\log_{10}(P(w_3   w_1 w_2))$ (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)
$w_1 w_2 w_3$		
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。

### 3.7 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” [5] 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。

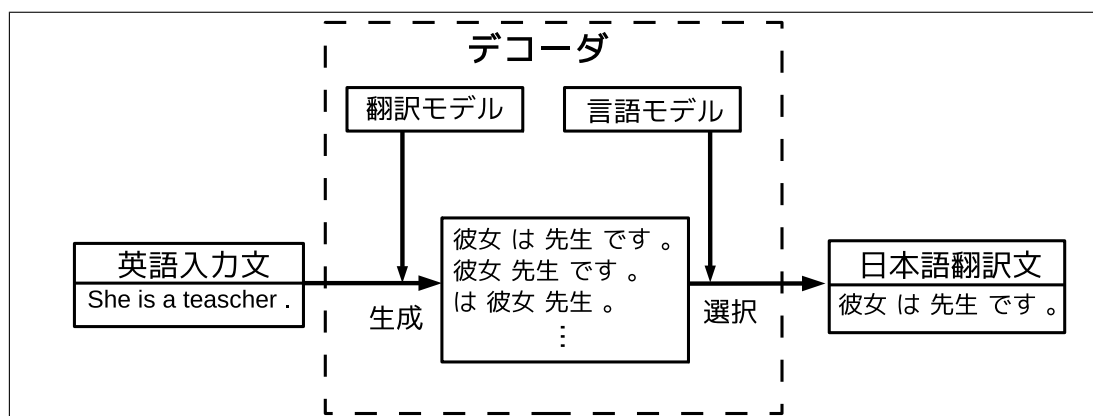


図 3.2 デコーダの動作例

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、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。

## 3.8 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) [6] 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。

## 3.9 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 [7], NIST [8], RIBES [9], TER [10], WER [10], METEOR [11] を用いる。

### 3.9.1 BLEU

自動評価法のなかで最も代表的な手法が BLEU である。BLEU は以下の式を用いて評価値の計算を行う。

$$BLEU = BP_{BLEU} * \exp\left(\frac{1}{N} \sum_{n=1}^N \log p_n\right) \quad (3.23)$$

値域は 0~1 であり、高いほど翻訳文が正解文に近く、翻訳精度が高いことを示す。(3.23) 式の  $N$  には、一般的に  $N = 4$  が用いられる。また、 $BP_{BLEU}$  は翻訳文が正解文と比較して短い場合に用いるペナルティ係数である。これは、翻訳文が正解文よりも短い場合に適合率が上がってしまうことを回避するために用いられる。 $BP_{BLEU}$  は翻訳文の単語数を  $c$ 、正解文の単語数を  $r$  とし、以下の式で計算される。

$$BP_{BLEU} = \begin{cases} 1 & (c > r) \\ e^{1-\frac{r}{c}} & (c \leq r) \end{cases} \quad (3.24)$$

翻訳文の文字数が正解文よりも長い場合には、 $BP_{BLEU} = 1$  であり、BLEU 値に対して影響を及ぼさない。長い場合のペナルティは後述する。翻訳文の文字数が正解文よりも短い場合には、その長さに応じて、 $BP_{BLEU}$  が小さくなり、結果として BLEU 値が小さくなる。なお、単語数  $c$  と  $r$  を算出するときの対象文が 1 文ではなく、コーパス全体

であるので、各文の長さの自由度が残されている。また、(3.23) 式の指数関数の  $p_n$  は以下の式で計算される。

$$p_n = \frac{\sum_{\text{全翻訳文}} \sum_{\text{正解文}} \text{正解文と一致する } N\text{-gram 数}}{\sum_{\text{全翻訳文}} \sum_{\text{正解文}} \text{全 } N\text{-gram 数}} \quad (3.25)$$

$p_n$  は正解文と一致する  $N$ -gram 数を数えるときに正解文の要素を重複して数えることを回避するための処理を行う。例えば、翻訳文が “a a a a” であり、正解文が “This is a book” である場合にこの処理を行わなければ、翻訳文の 4 つの “a” と一致するものとして、正解文の “a” が重複して数えられ  $p_1 = \frac{4}{4}$  となる。この値に対して処理を行うことで、 $p_1 = \frac{1}{4}$  となる。また、正解文に対して、翻訳文の方が長い場合には、 $p_n$  の値は小さくなる。つまり、BLEU 値が小さくなるので、ペナルティが (3.23) 式に含まれていることがわかる。

### 3.9.2 NIST

NIST は、BLEU と同様に  $N$ -gram 適合率で評価を行う。情報量で重み付けしている点が異なる。また、実験では 5-gram を用いる。NIST は 0 から  $\infty$  のスコアを出力し、スコアが大きい方が良い評価である。NIST の計算式を以下に示す。

$$NIST = \sum_{n=1}^N \frac{\sum_i \left( \sum_{\text{出力文 } i \text{ と参照文 } i \text{ に共通する } w_1 \cdots w_n} \text{Info}(w_1 \cdots w_n) \right)}{\sum_i \text{出力文 } i \text{ の中の全 } N\text{-gram 数}} \quad (3.26)$$

$$\text{Info}(w_1 \cdots w_n) = \log_2 \frac{\text{評価コーパス中の } w_1 \cdots w_{n-1} \text{ 数}}{\text{評価コーパス中の } w_1 \cdots w_n \text{ 数}} \quad (3.27)$$

### 3.9.3 RIBES

RIBES は、参照文と出力文との間で、共通単語の出現順序を順位相関係数で評価を行う評価法である。計算式を以下に示す。

$$RIBES = NSR \times P^\alpha \quad (3.28)$$

$$RIBES = NKT \times P^\alpha \quad (3.29)$$

ここで、 $NSR$  はスピアマンの順位相関係数であり、 $NKT$  はケンドールの順位相関係数である。また  $\alpha$  はペナルティに対する重みとして使用され、 $0 \leq \alpha \leq 1$  の値である。単

語の出現順を順位相関係数を用いて評価することで、文全体の語順に着目することができる。なお、RIBESは0から1のスコアを出力し、スコアが大きい方が良い評価である。

### 3.9.4 TER

TERは、Translation Edit Rateの略で翻訳の誤り率を求める評価法である。計算式を以下に示す。

$$TER = \frac{\sum_i (\text{挿入語数 } i + \text{置換語数 } i + \text{削除語数 } i + \text{シフト語数 } i)}{\sum_i (\text{参照文 } i \text{ の平均単語数})} \quad (3.30)$$

分子は参照文と出力文の比較における編集操作数のことである。TERの編集操作は挿入、置換、削除、シフトの4種類の編集を行うことである。なお、TERはスコアが小さい方が良い評価である。

### 3.9.5 WER

WERは、Word Error Rateの略で単語の誤り率を求める評価法である。以下に計算式を示す。

$$WER = \frac{\sum_i (\text{挿入語数 } i + \text{置換語数 } i + \text{削除語数 } i)}{\sum_i (\text{参照文 } i \text{ の平均単語数})} \quad (3.31)$$

分子は参照文と出力文の比較における編集操作数のことである。WERの編集操作は挿入、置換、削除の3種類の編集を行うことである。なお、WERはスコアが小さい方が良い評価である。

### 3.9.6 METEOR

METEORは適合率Pと再現率RからF値を求めた後に、単語の非連続生に基づいてF値を修正して評価を行う自動評価法である。F値と修正に用いるペナルティ関数Penは以下の式で計算される。

$$F = \frac{P \times R}{\alpha \times P + (1 - \alpha) \times R} \quad (3.32)$$

$$METEOR = F \times (1 - Pen) \quad (3.33)$$

$$Pen = \gamma \times \left(\frac{c}{m}\right)^\beta \quad (3.34)$$

METEOR の値域は 0~1 であり，高いほど翻訳精度が高い．(3.34) 式において， $m$  は翻訳文と正解文の一致率であり， $c$  は単語を対象として，翻訳文と一致するものを 1 つに統合したときの数である． $\alpha$ ， $\beta$ ， $\gamma$  はパラメータである．一致率  $c$  を計算するときの類義語などのマッチングには一般に WordNet を用いる．このマッチングでは，意味や概念が近い類義語のマッチング，単語の語形変化を同一のものとして処理することが可能であり，人手評価に近い評価結果を得ることができる．

## 第4章 提案手法

### 4.1 提案手法の概略

パターン翻訳は、原言語文と目的言語文で構成される対訳文に対して、任意の単語やフレーズを変数化した“文パターン”と“単語辞書”を用いて翻訳を行う方法である。パターン翻訳は長所として、入力文が適切な文パターンに適合した場合、高品質な翻訳文を得やすい傾向にある。しかし、パターン翻訳には多くの問題点がある。

本研究では中でもコストと翻訳精度とカバー率の問題を取り上げる。まず、パターン翻訳は問題点として、非常にコストと時間がかかる。この理由は単語辞書と文パターンを手で作成するためである。本研究では単語辞書と文パターンをプログラムで自動作成することでコストの削減を試みる。

また、パターン翻訳は翻訳精度が高いとカバー率が低く、カバー率が高いと翻訳精度が低くなる傾向にある。このトレードオフの関係が発生する理由として、入力文に適合した文パターンが保持する字面の数が関係していると考えた。つまり、文パターンの字面が多ければ翻訳精度が高く(カバー率が低く)、少なければカバー率が高く(翻訳精度が低く)なる。また、カバー率には、入力文が文パターンに適合しなければ翻訳ができない問題がある。

改善策として、翻訳精度の問題に対しては、入力文と文パターンの字面を比較する。そして字面が多く一致する文パターンを優先して選択することで翻訳精度の改善を試みる。カバー率の問題に対しては、大量の文パターンと単語辞書を作成することでカバー率の増加を試みる。

本研究では5つのステップを用いて英日パターン翻訳を行う。以下に手順を示す。

#### 手順1 単語辞書

GIZA++を用いて、単語辞書を作成する。

#### 手順2 単語に基づく文パターン辞書

単語辞書を用いて、単語に基づく文パターン辞書を作成する。

### 手順3 フレーズ辞書

単語に基づく文パターン辞書を用いて、フレーズ辞書を作成する。

### 手順4 句に基づく文パターン辞書

フレーズ辞書を用いて、句に基づく文パターン辞書を作成する。

### 手順5 英日パターン翻訳

フレーズ辞書と句に基づく文パターン辞書を用いて、英日パターン翻訳を行う。

なお、単語に基づく文パターンを用いた英日パターン翻訳の実験結果は [12] で報告している。また、4.2 節で用いる対訳学習文とは、電子辞書から抽出した英日対訳文を指す。詳細は 5.1 節で説明する。

## 4.2 提案手法の詳細

### 手順1 単語辞書

単語辞書を作成するために、対訳学習文と GIZA++ を用いる。まず、GIZA++ から英日方向、日英方向の単語対応を調べる。そして、英日方向の単語辞書と日英方向の単語辞書と単語翻訳確率を得る。最後に英日方向の単語翻訳確率と日英方向の単語翻訳確率を掛け合わせる。本研究では英日方向と日英方向の単語翻訳確率を掛け合わせて得た確率値を単語確率と呼ぶ。また、単語辞書は単語確率から得る。次節の単語に基づく文パターン辞書の作成で用いる単語辞書は、文パターン数を抑制するために閾値を設ける。

以降、閾値により選別した単語辞書を“単語辞書(手順1)”，閾値を用いない単語辞書を“単語辞書(GIZA)”と定義する。また、閾値は 5 章で説明する。



図 4.1 に英日方向の単語辞書の作成手順と例を示す。

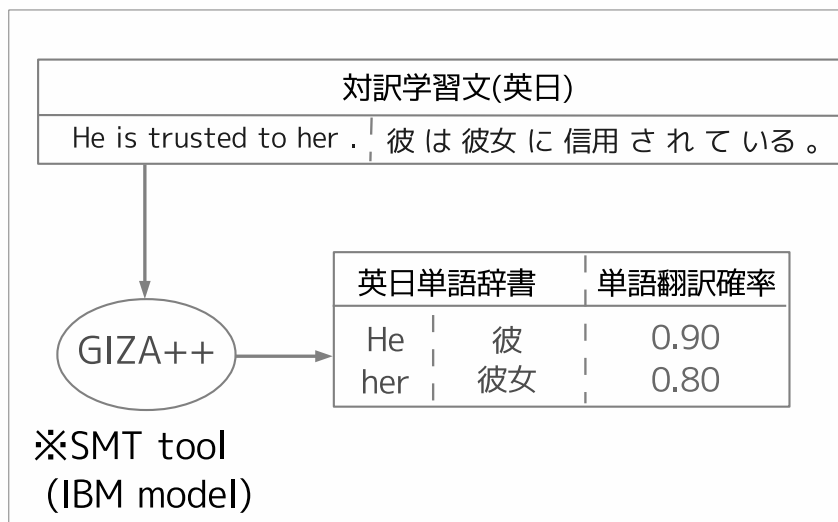


図 4.1 英日方向の単語辞書の作成手順

図 4.2 に日英方向の単語辞書の作成手順と例を示す。

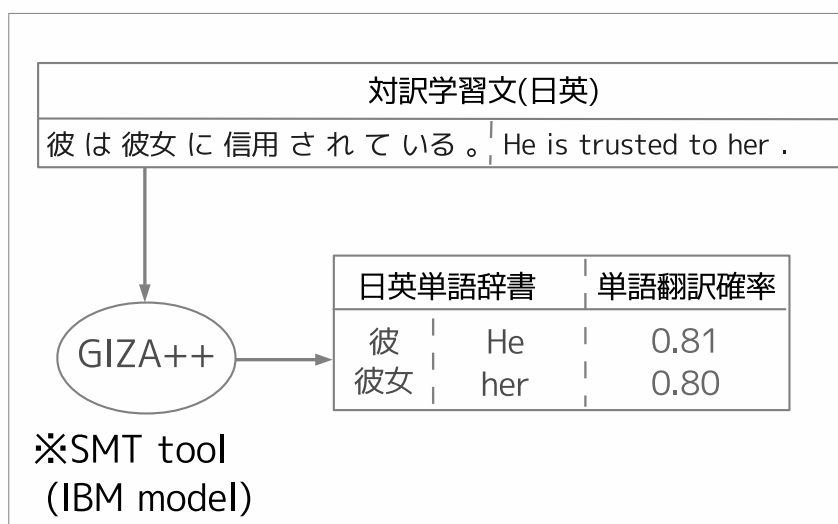


図 4.2 日英方向の単語辞書の作成手順

図 4.3 に単語辞書 (GIZA) の作成手順と例を示す。

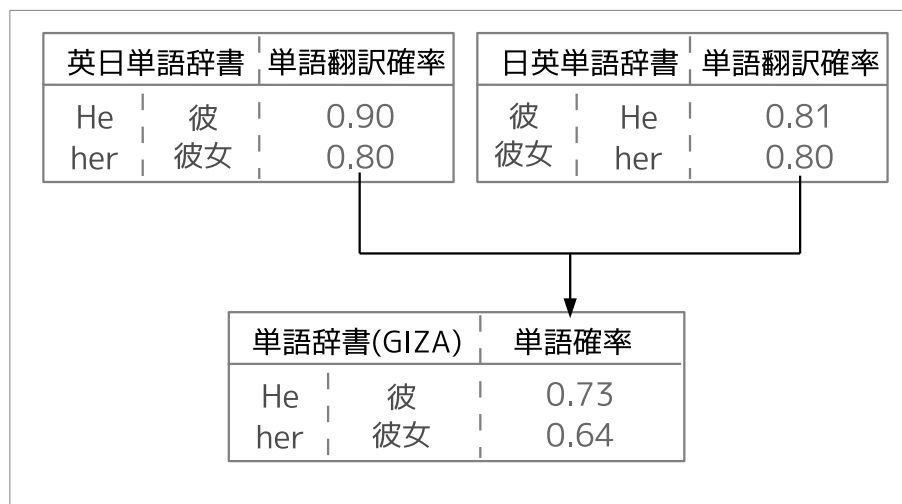


図 4.3 日英方向の単語辞書の作成手順

## 手順2 単語に基づく文パターン辞書

単語に基づく文パターン辞書を作成するために、単語辞書 (手順1) と対訳学習文を用いる。

なお、大量の句に基づく対訳文パターンを生成させるため、可能な限り単語に基づく文パターンを生成する。

### 1. 英単語照合

対訳学習文の各英単語と単語辞書の英単語を照合する。

### 2. 日本語単語照合

英単語に対応する日本語単語と対訳学習文の各日本語単語を照合する。

### 3. 変数化

英日両方の単語が照合に成功した場合、該当箇所を変数化する。変数化する場合、変数の組み合わせを考慮し、可能な限り多くの単語に基づく対訳文パターンを生成する。

図 4.4 に単語に基づく文パターン辞書の作成手順と例を示す。

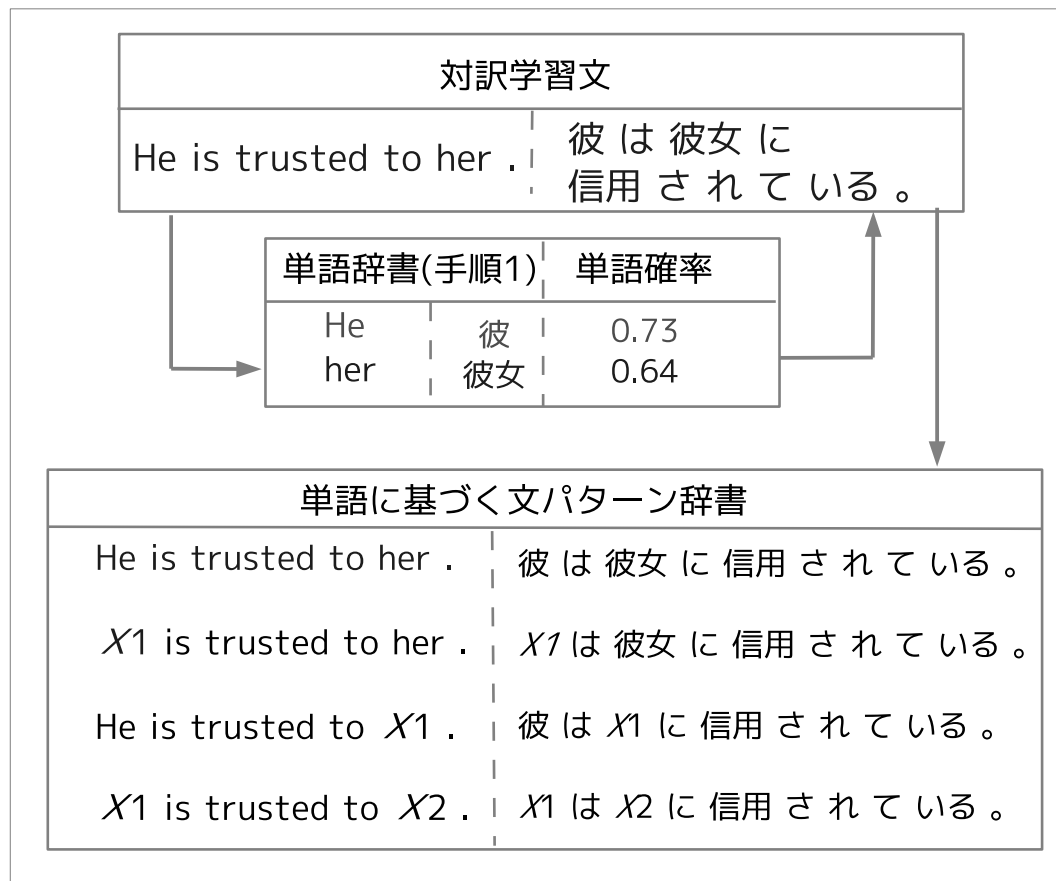


図 4.4 単語に基づく文パターン辞書の作成手順

図 4.4 において変数化される対訳単語は“He | 彼”，“her | 彼女”，である．2つの対訳単語が変数化される場合と変数化されない場合の組み合わせを全て考慮し， $2^2=4$ 通りの単語に基づく文パターンを生成する．

### 手順3 フレーズ辞書

対訳フレーズを抽出するために，単語に基づく文パターン辞書(手順2)と対訳学習文を用いる．

#### 1. パターン照合

対訳学習文と単語に基づく文パターン辞書を照合する．

## 2. 対訳フレーズの抽出

対訳学習文が単語に基づく文パターンに適合した場合、単語に基づく文パターンの変数部に対応する対訳フレーズを抽出する。また、本研究では対訳フレーズの英語側を英語フレーズ、日本語側を日本語フレーズと呼ぶ。

図 4.5 に対訳フレーズの抽出手順と例を示す。

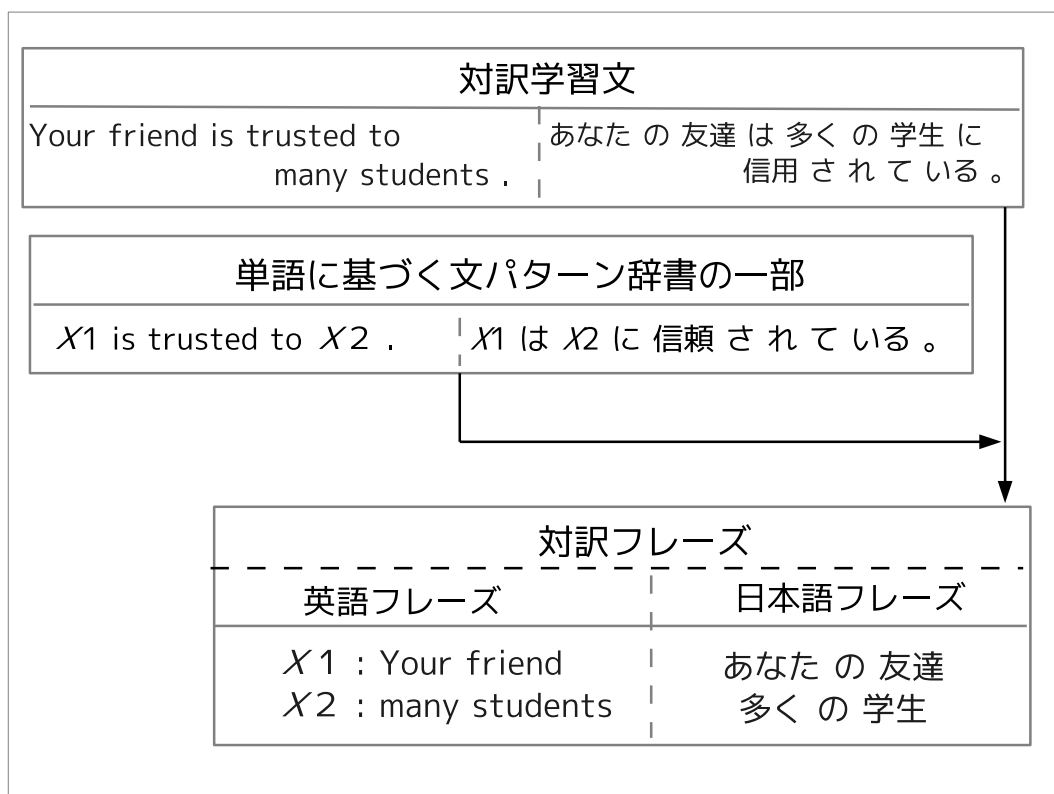


図 4.5 対訳フレーズの抽出手順

## 3. 対訳フレーズの翻訳確率の計算

対訳フレーズの翻訳確率を計算する。対訳フレーズの翻訳確率は英語フレーズが日本語フレーズに翻訳される確率と日本語フレーズが英語フレーズに翻訳される確率を計算する。以下に手順を示す。

### A. 単語の組み合わせの取得

対訳フレーズにおいて、英語フレーズの単語と日本語フレーズの単語の全ての組み合わせを得る。同様に日本語フレーズの単語の組み合わせと英語フレーズの単語の

組み合わせも得る.

### B. 翻訳確率の計算

各英単語に対応する日本語単語の中で，単語辞書 (GIZA) の単語確率が最大となる単語確率を得る．同様に各日本語単語に対応する英単語の中で，単語辞書 (GIZA) の単語確率が最大となる単語確率を得る．

### C. 翻訳確率の付与

得られた単語確率に対して対数を取り，英日方向の翻訳確率の対数値の総和と日英方向の翻訳方向の対数値の総和を求める．英日方向の総和を英日フレーズ確率，日英方向の総和を日英フレーズ確率と呼ぶ．最後に英日フレーズ確率と日英フレーズ確率を足し合わせ，対訳フレーズの翻訳確率として付与する．本研究では計算した確率をフレーズ確率 ( $\alpha$ ) と呼ぶ．

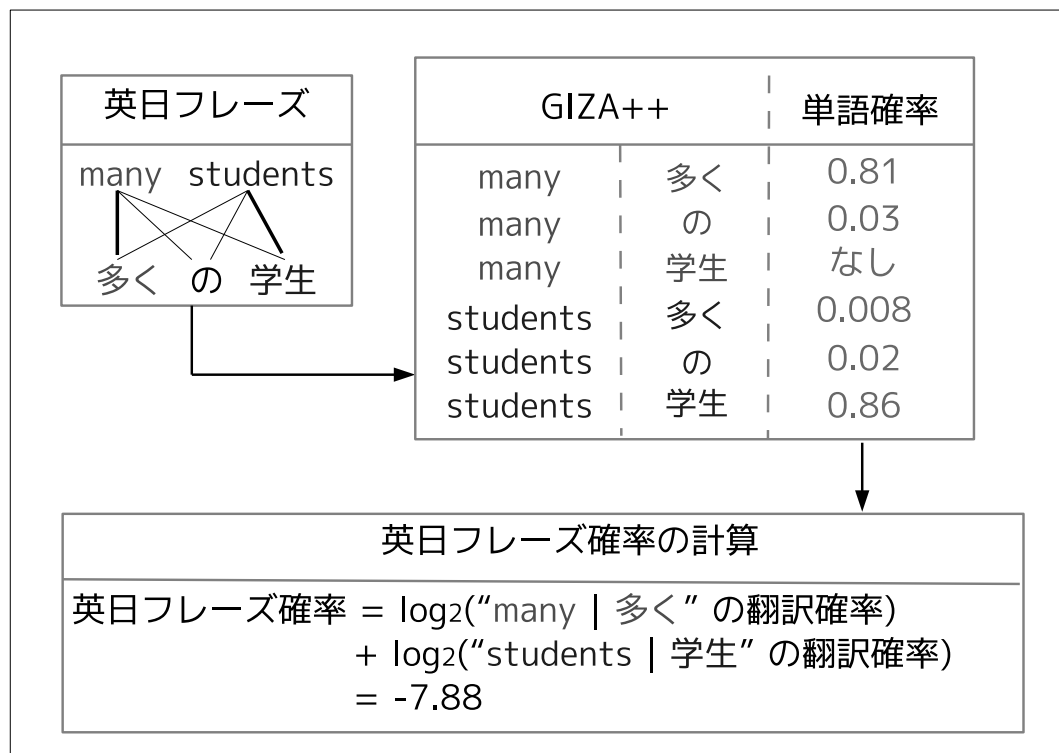


図 4.6 英日フレーズ確率の付与手順

図 4.6 に英日方向の対訳フレーズの例として “many students | 多くの 学生” を

示す。まず、英語フレーズの単語と日本語フレーズの単語の全ての組み合わせを得る。単語辞書 (GIZA) の単語確率を用いて、各組み合わせの中から最大となる単語確率を得る。図 4.6 では “many | 多く” に付与された確率 “0.81” が最も高いため、0.81 に対して対数を取る。“students” も同様に単語確率に対数を取り総和を求める。

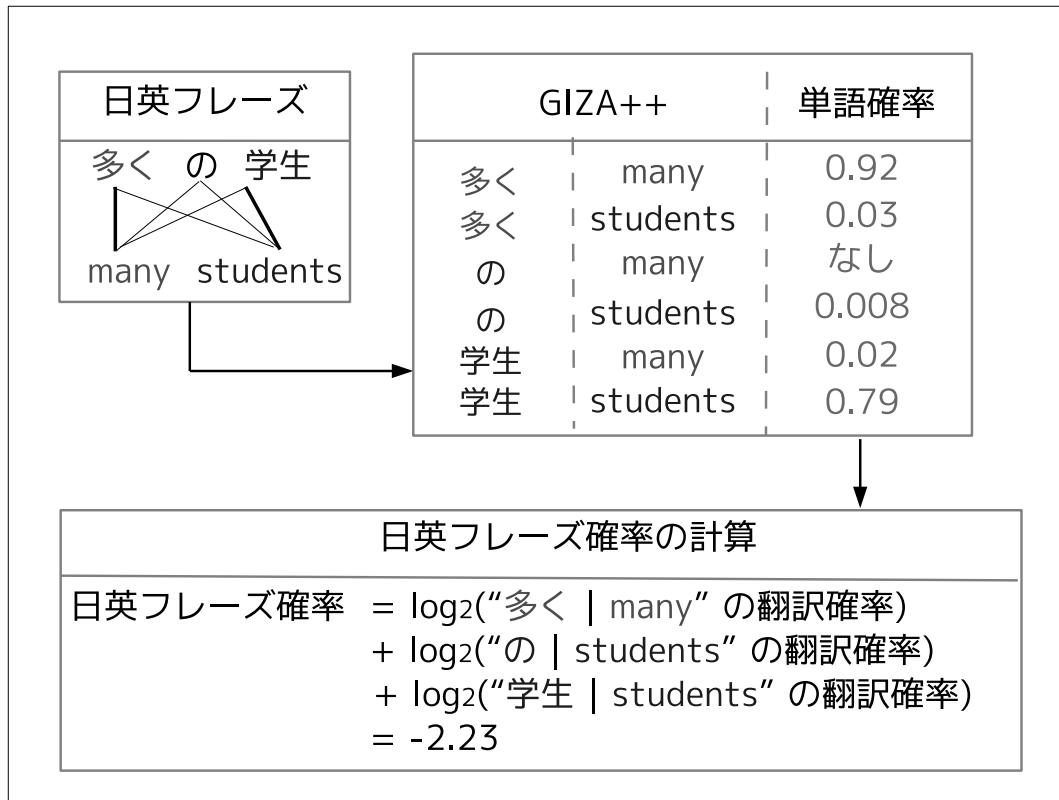


図 4.7 日英フレーズ確率の付与手順

同様に、図 4.7 英日方向の対訳フレーズの例として “多くの学生 | many students” を示す。まず、日本語フレーズの単語と英語フレーズの単語の全ての組み合わせを得る。単語辞書 (GIZA) の単語確率を用いて、各組み合わせの中から最大となる単語確率を得る。図 4.7 では “多く | many” に付与された確率 “0.92” が最も高いため、0.92 に対して対数を取る。“の”, “学生” も同様に単語確率に対数を取り総和を求める。

今回は英日フレーズ確率の -7.88 と日英フレーズ確率の -2.23 の和である -10.11 をフレーズ確率 ( $\alpha$ ) として付与する。

#### 4. 対訳フレーズの選別

次節で作成する句に基づく文パターンを抑制するため、抽出した対訳フレーズの選別を行う。対訳フレーズの選別は2つの手法を用いる。以下に手順を示す。

#### A. フレーズ確率への閾値設定

設定した閾値以上のフレーズ確率を持つフレーズのみを本研究では使用する。閾値の詳細は5章で説明する。

#### B. 対訳フレーズの単語数の差による抽出

英語フレーズと日本語フレーズの単語数の絶対値を計算する。英語フレーズの単語数を基準とし、対応する日本語フレーズ数が±5以内のフレーズのみを抽出する。例えば英語フレーズが2単語、日本語フレーズが3単語であれば抽出するが、英語フレーズが2単語、日本語フレーズが8単語であれば抽出しない。図4.8に単語数の差による抽出方法を示す。

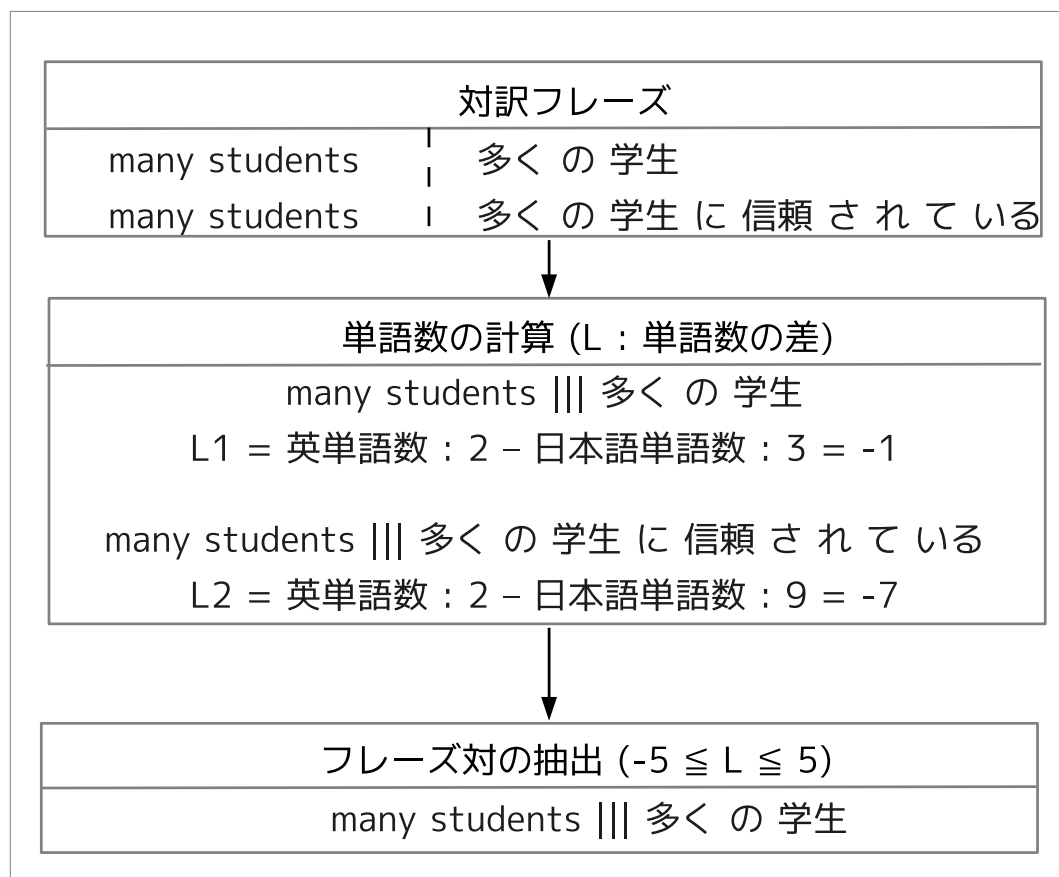


図 4.8 単語数の差による抽出

## 手順4 句に基づく文パターン辞書

句に基づく文パターン辞書を作成するために、フレーズ辞書(手順3)と対訳学習文を用いる。以下に手順を示す。

### 1. 英語フレーズ照合

対訳学習文における英語文の各フレーズとフレーズ辞書(手順3)の英語フレーズを照合する。

### 2. 日本語フレーズ照合

英語フレーズに対応する日本語フレーズと対訳学習文における日本語文の各フレーズを照合する。

### 3. 変数化

対訳フレーズが照合に成功した場合、該当箇所を変数化し、文パターンを生成する。変数化するとき、変数の組み合わせを考慮し、可能な限り多くの句に基づく文パターンを生成する。本研究では、句に基づく文パターンの英語側を英語フレーズ、日本語側を日本語フレーズと呼ぶ。図4.9に句に基づく文パターン辞書の作成手順と例を示す。

図4.9において変数化されるフレーズ対は“Your friend | あなたの友達”，“many students | 多くの学生”である。この2つのフレーズ対が変数化される場合とされない場合の組み合わせを全て考慮し、 $2^2=4$ 通りの句に基づく文パターンを生成する。

### 4. 句に基づく文パターンの翻訳確率の付与

文パターンの字面と単語辞書(GIZA)の単語確率を用いて、文パターンに翻訳確率を付与する。翻訳確率の付与は英日文パターンと日英文パターンに対して行う。また、手順3.3で説明したフレーズ確率の付与と同じ手法を用いる。本研究では計算した確率を文パターン確率( $\beta$ )と呼ぶ。図4.10に、英日方向の文パターン確率の付与手順と例を、図??に日英方向の文パターン確率の付与手順と例を示す。

図4.10に英日方向の句に基づく文パターンの例として“ $X1$  is trusted to  $X2$  . |  $X1$  は  $X2$  に信頼されている。”を示す。まず、英語文パターンの単語と日本語文パターンの単語の全ての組み合わせを得る。単語辞書(GIZA)の単語確率を用いて、各組み合わ



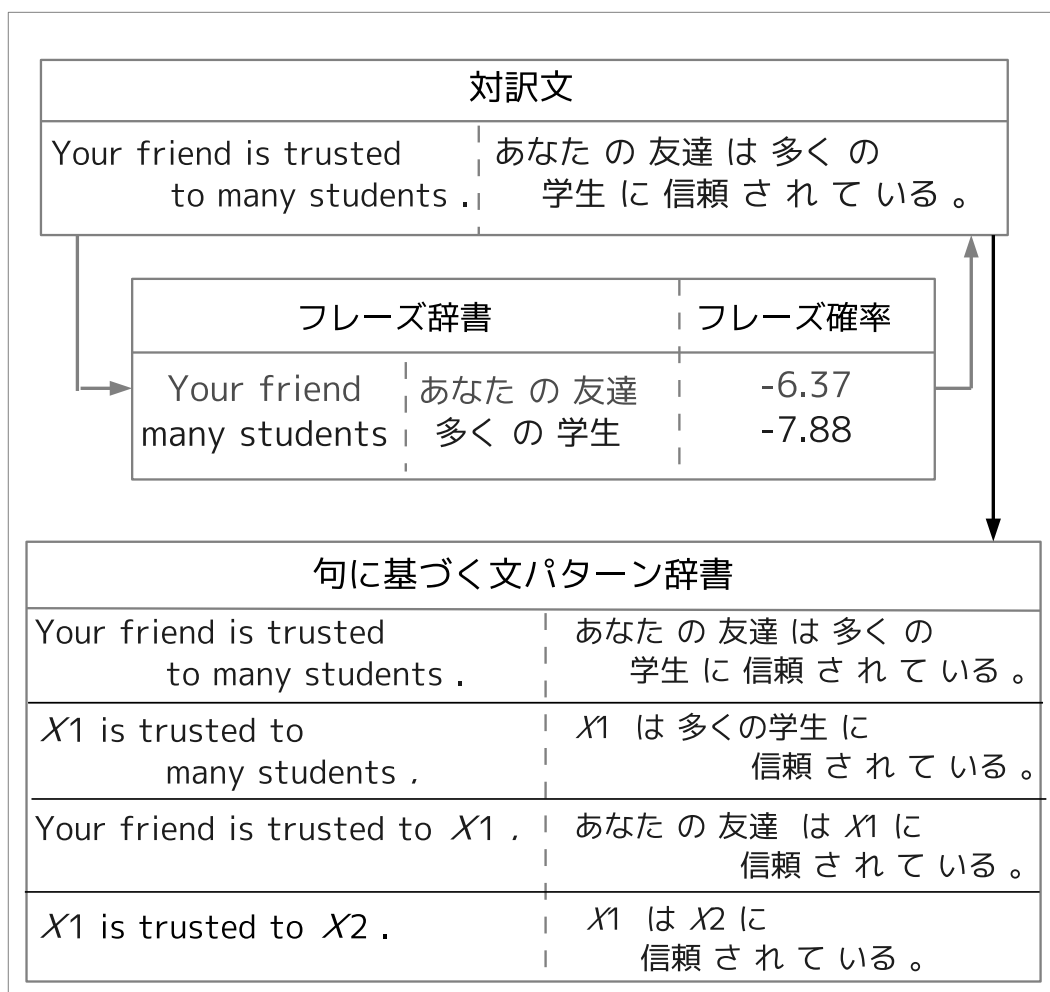


図 4.9 句に基づく文パターン辞書の作成手順

せの中から最大となる単語確率を得る。図 4.10 では “is | は” に付与された確率“0.81”が最も高いため、0.81 に対して対数を取る。“trusted”，“to” も同様に単語確率に対数を取り総和を求める。

同様に、図 4.11 に日英方向の句に基づく文パターンの例として“X1 は X2 に 信頼 されている。| X1 is trusted to X2 .”を示す。まず、日本語文パターンの単語と英語文パターンの単語の全ての組み合わせを得る。単語辞書 (GIZA) の単語確率を用いて、各組み合わせの中から最大となる単語確率を得る。図 4.11 では “は | is” に付与された確率“0.11”が最も高いため、0.11 に対して対数を取る。“に”，“信頼”，“さ”，“れ”，“て”，“いる”，も同様に単語確率に対数を取り総和を求める。今回の例では、英日方向の文パターン確率-1.8 と日英方向の文パターン確率-6.01 の和である-7.81 を文パターン確率 ( $\alpha$ )

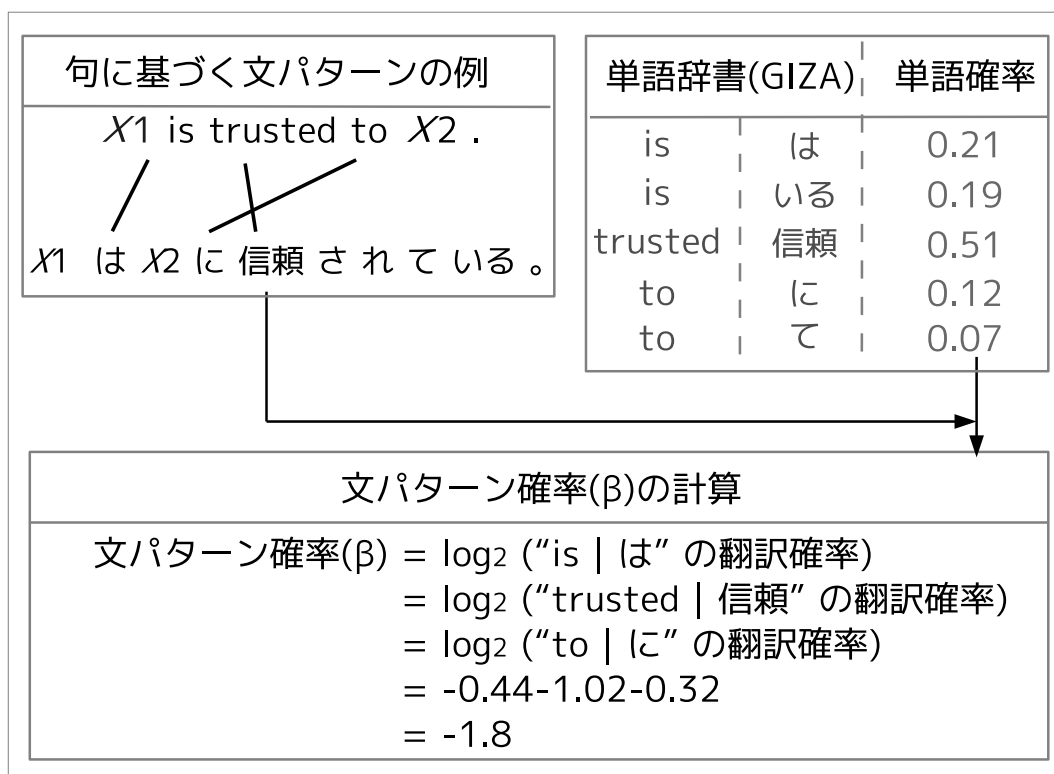


図 4.10 文パターン確率の付与手順

として付与する。

## 手順5 英日パターン翻訳

日本語翻訳文を出力するために、フレーズ辞書(手順3)と句に基づく文パターン辞書(手順4)を用いる。

翻訳精度を向上させるために、翻訳時に英語入力文と英語文パターンの字面を比較する。そして最も多く字面が一致する英語文パターンから優先して選択する。

日本語翻訳文の絞り込みにはフレーズ確率と文パターン確率と言語翻訳確率(tri-gram)を用いる。総和を取り、確率が最大となる日本語翻訳文を出力する。以下に英日パターン翻訳の手順を示す。

### 1. 英語文パターンの選択

英語入力文を読み込み、英語入力文と英語文パターンの字面を比較する。そして最も多く字面が一致する英語文パターンから優先して選択する。

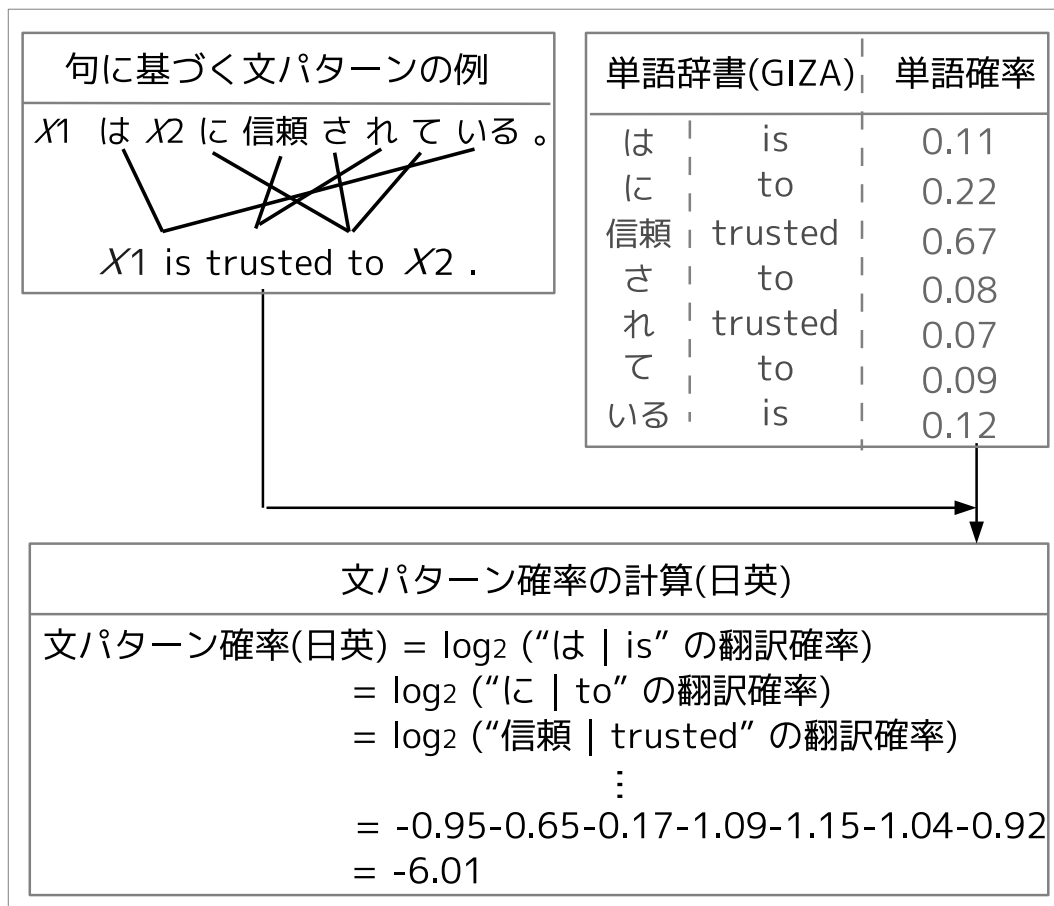


図 4.11 文パターン確率の付与手順

2. 英語フレーズの取得

一致する英語文パターンの変数部に対応する英語フレーズを得る。

3. 日本語文パターンの取得

英語文パターンに対応する日本語文パターンと文パターン確率を得る。

4. 日本語フレーズの取得

日本語文パターンの変数部に対応する日本語フレーズとフレーズ確率を得る。

5. 日本語翻訳文の生成

日本語文パターンの変数部を手順 5.4 で取得した日本語フレーズに置き換える。そして、日本語翻訳文として出力する。

## 6. 言語翻訳確率 (tri-gram) の算出

手順 5.5 の日本語翻訳文に対して言語翻訳確率 (tri-gram) を計算する。

## 7. 日本語翻訳文の選択

フレーズ確率と文パターン確率と言語翻訳確率 (tri-gram) の総和を求め、日本語翻訳文に付与する。最後に総和が最大となる日本語翻訳文を出力する。

図 4.12 に日本語翻訳文を出力するまでの手順を示す。

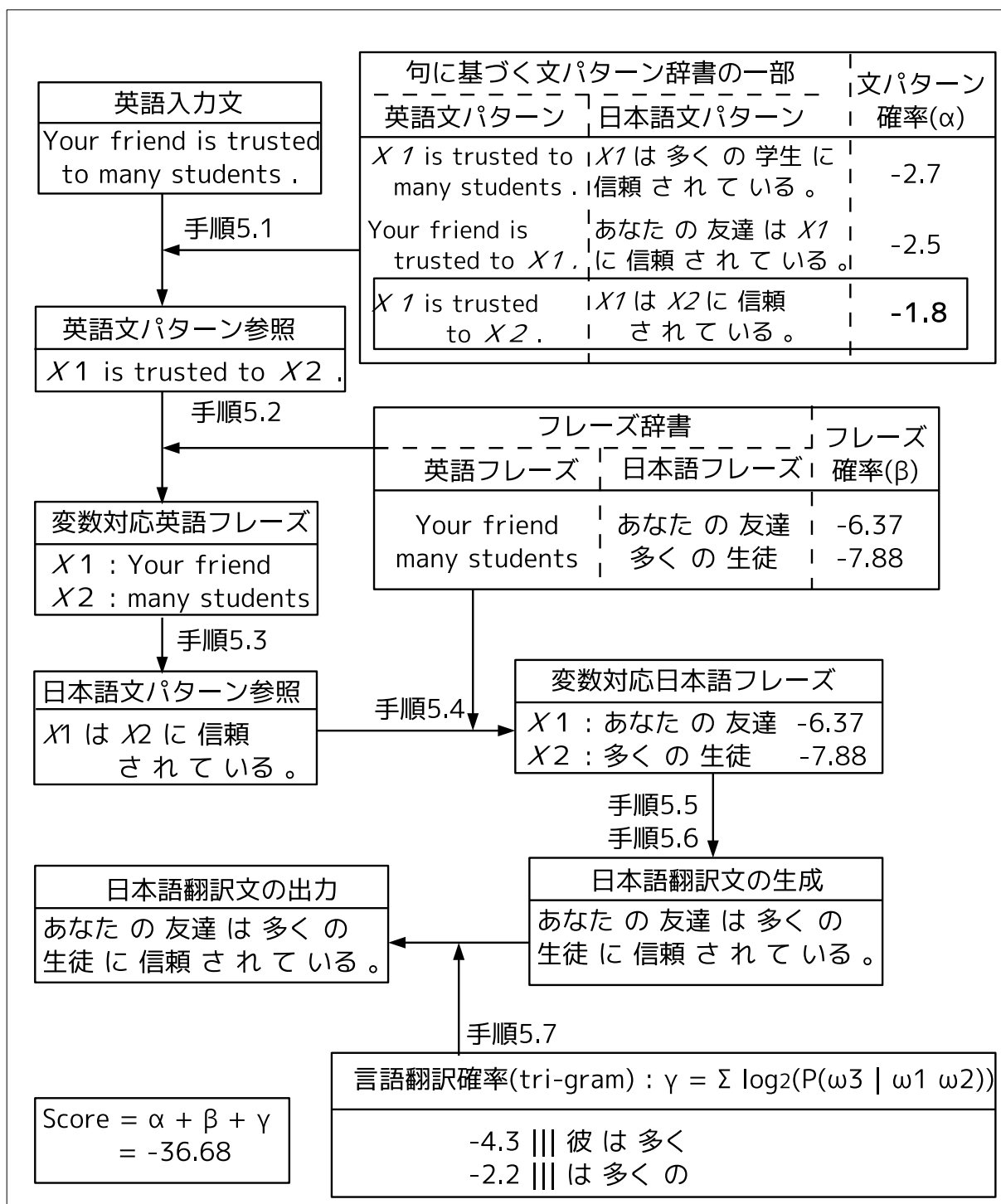


図 4.12 日本語翻訳文出力手順

# 第5章 実験

## 5.1 実験条件

### 5.1.1 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を用いる。以下に注意点を示す。

- 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。
- 本研究では、電子辞書から抽出した英日対訳文を対訳学習文と呼ぶ。

### 5.1.2 句に基づく文パターンを用いた英日翻訳

以下に 4.2 節における実験条件を示す。

#### 手順1 単語辞書

- 対訳学習文 100,000 文を用いる。
- 単語に基づく文パターンの出力数を抑制するため、閾値を 0.1 とする。

#### 手順2 単語に基づく文パターン辞書

- 対訳学習文 100,000 文を用いる。
- 先行研究 [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。しかし、本研究で作成する文パターンには意味属性の付与を行わない。手順4の句に基づく文パターン辞書の作成も同様である。

#### 手順3 フレーズ辞書

- 対訳学習文 100,000 文を用いる。
- 句に基づく文パターンの出力数を抑制するため、閾値を  $-100.0$  とする。

#### 手順 4 句に基づく対訳文パターン辞書

- 対訳学習文 1 文に対し、文パターンの出力数は最大 100,000 パターンまでとする。
- 句に基づく文パターンの作成には、フレーズ辞書 (手順 3) を用いる。ただし、1 つの英語フレーズに対して、付与されたフレーズ確率が高い上位 2 つの日本語フレーズを抽出して利用する。

#### 手順 5 英日パターン翻訳

- 英日パターン翻訳は入力文として英語文 100 文を用いる。
- 英日パターン翻訳には、フレーズ辞書 (手順 3) と句に基づく文パターン辞書 (手順 4) を用いる。ただし、1 つの英語フレーズに対して、付与されたフレーズ確率が高い上位 512 の日本語フレーズを抽出して利用する。
- 英語文パターンの選択
  - － 英語文パターンを選択する際は、英語入力文と英語文パターンの字面を比較し、最も多く字面が一致する英語文パターンを優先して選択する。
  - － 英語入力文 1 文に対し、英語文パターンの選択数は 1,000 文までとする。
- 言語翻訳確率 (tri-gram) の計算
  - － 日本語翻訳文に対して言語翻訳確率 (tri-gram) を計算する。
  - － 言語翻訳確率 (tri-gram) は対訳学習文の日本語文 100,000 文を用いる。
- 日本語翻訳文
  - － 翻訳精度の低い出力文を除外するために、閾値を  $-1,000.0$  とする。

### 5.1.3 ベースラインシステム

ベースラインシステムには Moses [5] を用いる。また、形態素解析には Mecab [14] を使用する。

#### 5.1.3.1 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。

#### 5.1.3.2 言語モデルの学習

言語モデルの学習には SRILM [15] 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。

## 5.2 実験結果

### 5.2.1 提案手法より得たデータ数

表 5.1 に 4.2 節の各手順で得たデータ数を示す。また、表 5.1 に示す英日パターン翻訳は英語入力文 100 文に対し、フレーズ辞書 (手順 3) と句に基づく文パターン辞書 (手順 4) と言語翻訳確率を組み合わせ得た日本語翻訳文の数である。

表 5.1 4.2 節で得たデータ数

単語辞書 (GIZA)	501,010 単語
単語辞書 (手順 1)	22,574 単語
単語に基づく文パターン辞書	1,464,042 パターン
フレーズ辞書	4,128,831 フレーズ
句に基づく文パターン辞書	175,087,300 パターン
英日パターン翻訳	24/100 文



## 5.2.2 日本語翻訳文の出力例

表 5.2 から表 5.7 に提案手法における出力例の詳細を示す. ( $\beta$ :文パターン確率( $\beta$ ), total:フレーズ確率( $\alpha$ )+文パターン確率( $\beta$ )+tri-gram スコア( $\gamma$ ))

表 5.2 出力例の詳細 1

		翻訳確率
英語入力文	The signal changed from green to red .	
英語文パターン	The X00 X01 from X02 to X03 .	$\beta = -10.95$
日本語文パターン	X00 が X02 から X03 に X01 .	
提案手法	信号が青から赤になった。	total = -41.71
参照文	信号が青より赤に変わった。	

表 5.3 に変数に対応するフレーズの翻訳結果を示す.

表 5.3 変数対応フレーズ

変数名	翻訳結果	フレーズ確率( $\alpha$ )
X00	signal : 信号	-1.50
X01	changed : なった	-5.13
X02	green : 青	-2.61
X03	red : 赤	-1.45

表 5.4 に文パターンの原文を示す.

表 5.4 文パターンの原文

英語文パターン原文	The main topic at the Diet session shifted from economic to political questions .
日本語文パターン原文	国会での論点が経済から政治の問題に移った。
英語文パターン	The X00 X01 from X02 to X03 .
日本語文パターン	X00 が X02 から X03 に X01 .

表 5.5 出力例の詳細 2

		翻訳確率
英語入力文	This company has a good staff of talents .	
英語文パターン	This X00 has a X01 X02 of X03 .	$\beta = -27.04$
日本語文パターン	この X00 に X01 X03 X02 ある。	
提案手法	この会社にはすばらしい劇作の才能がある。	total = -81.26
参照文	この会社には人材が揃っている。	

表 5.6 に変数に対応するフレーズの翻訳結果を示す。

表 5.6 変数対応フレーズ

変数名	翻訳結果	フレーズ確率 ( $\alpha$ )
X00	company : 会社	-0.71
X01	good : は	-5.65
X02	staff : が	-9.74
X03	talents : すばらしい劇作の才能	-3.92

表 5.7 に文パターンの原文を示す。

表 5.7 文パターンの原文

英語文パターン原文	This room has a big window in the wall on each of its three sides .
日本語文パターン原文	この部屋には大きな窓が3方の壁にある。
英語文パターン	The X00 X01 from X02 to X03 .
日本語文パターン	X00 が X02 から X03 に X01 。

## 第6章 評価結果

提案手法より得た 24 文の日本語翻訳文に対し、自動評価と人手評価を行った。まず、自動評価結果を以下に示す。

### 6.1 自動評価結果

評価には BLEU [7], NIST [8], RIBES [9], TER [10], WER [10], METEOR [11] の 6 種類の自動評価法を用いる。また、ベースラインシステムには Moses [5] を用いる。表 6.1 に提案手法とベースラインシステムの自動評価結果を示す。

表 6.1 提案手法とベースラインの自動評価結果

	BLEU	NIST	RIBES	TER	WER	METEOR
提案手法	0.32	3.86	0.89	46.54	46.54	0.60
ベースライン	0.28	3.78	0.87	47.53	49.01	0.57

表 6.1 の結果から、提案手法より得た日本語翻訳文はベースラインと比較してスコアが向上している。

### 6.2 人手評価結果

人手による対比較評価を行った。自動評価と同様ベースラインシステムには Moses を用いる。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 : 提案手法とベースラインの出力が同じ

表 6.2 に提案手法とベースラインの人手較評価結果を示す。

表 6.2 提案手法とベースラインの人手評価結果

提案手法○	提案手法×	差なし	同一出力
6	3	10	5

表 6.2 より，提案手法の有効性が確認できた。

### 6.3 人手評価における提案手法○の例

提案手法○の例を表 6.3～表 6.5 に示す。

表 6.3 提案手法○の例 1

英語入力文	The signal changed from green to red .
英語文パターン	The X00 X01 from X02 to X03 .
日本語文パターン	X00 が X02 から X03 に X01 .
提案手法	信号が青から赤になった。
ベースライン	信号が赤に緑を変えた。
参照文	信号が青より赤に変わった。

表 6.3 において，提案手法は参照文と同等の翻訳文が得られている。しかし，ベースラインは翻訳文として不適切である。

表 6.4 提案手法○の例 2

英語入力文	She had her own way .
英語文パターン	She had her X00 .
日本語文パターン	彼女は X00 た。
提案手法	彼女は意地を通した。
ベースライン	彼女は自分でした。
参照文	彼女は我を通した。

表 6.4 において，提案手法は参照文と同等の翻訳文が得られている。一方ベースライ

ンは、参照文の意味を考慮した場合に、提案手法と比較して不適切な翻訳文であると判断した。

表 6.5 提案手法○の例 3

英語入力文	The sun flashed from behind the cloud .
英語文パターン	The X00 X01 from behind the X02 .
日本語文パターン	X00 X02 の陰 から X01 。
提案手法	太陽 が 雲 の 陰 から 現われた 。
ベースライン	太陽 が 雲 の 陰 から さっと 通過 した 。
参照文	太陽 が 雲間 から 突然 顔 を 出した 。

表 6.5 において、提案手法は参照文と同等の翻訳文が得られている。一方、ベースラインの翻訳文は文脈に不自然な表現が含まれている。

## 6.4 人手評価における提案手法× の例

提案手法×の例を表 6.6 から表 6.8 に示す。

表 6.6 提案手法×の例 1

英語入力文	The quarrel lasted for years .
英語文パターン	The X00 lasted for X01 .
日本語文パターン	その X00 X01 続いた 。
提案手法	その けんか でも 何 年 も 続いた 。
ベースライン	その けんか は 何 年間 も 続いた 。
参照文	その けんか は 何 年 も 糸 を 引いた 。

表 6.6 において、提案手法は“けんかでも何年も”と日本語文として不自然な表現である。一方でベースラインは、参照文と同等の翻訳文が得られている。

表 6.7 提案手法×の例 2

英語入力文	The game was rained out .
英語文パターン	The game was X00 .
日本語文パターン	試合は X00 た。
提案手法	試合は中止になった。
ベースライン	試合は雨で中止になった。
参照文	雨で試合が流れた。

表 6.7 において、提案手法は日本語文として意味は通じるが“rained out”が適切に翻訳されていない。一方でベースラインは参照文と同等の翻訳文が得られている。

表 6.8 提案手法×の例 3

英語入力文	This company has a good staff of talents .
英語文パターン	This X00 has a X01 X02 of X03 .
日本語文パターン	この X00 に X01 X03 X02 ある。
提案手法	この会社にはすばらしい劇作の才能がある。
ベースライン	この会社の社員に才能がある。
参照文	この会社には人材が揃っている。

表 6.8 において、提案手法は翻訳文として不適切である。一方でベースラインは参照文と同等の翻訳文が得られている。

## 6.5 評価結果のまとめ

表 6.1 と表 6.2 の結果から、提案手法はベースラインと比較して優れている文が多いことがわかる。よって、評価結果から提案手法の有効性が確認された。

## 第7章 追加実験

追加実験として，実験条件を変更してパターン翻訳を行った．7.1 節に実験条件の変更点を示す．

### 7.1 実験条件の変更点

- 英語文パターンの選択
  - － 英語文パターンを選択する際は，英語入力文と英語文パターンの字面を比較し，最も多く字面が一致する英語文パターンを優先して選択する．
  - － 英語入力文1文に対し，英語文パターンの選択数は10,000文までとする．

### 7.2 実験結果

表 7.1 に 7.1 節で得たデータ数を示す．表 7.1 に示す英日パターン翻訳は，7.1 に則り，英語入力文 100 文に対し，フレーズ辞書 (手順 3) と句に基づく文パターン辞書 (手順 4) と言語翻訳確率を組み合わせ得た日本語翻訳文の数である．

表 7.1 7.1 節で得たデータ数

単語辞書 (GIZA)	501,010 単語
単語辞書 (手順 1)	22,574 単語
単語に基づく文パターン辞書	1,464,042 パターン
フレーズ辞書	4,128,831 フレーズ
句に基づく文パターン辞書	175,087,300 パターン
英日パターン翻訳	50/100 文

## 7.3 評価結果

提案手法より得た 50 文の日本語翻訳文に対し，自動評価と人手評価を行った．まず，自動評価結果を以下に示す．

### 7.3.1 自動評価結果

6.1 節と同様の自動評価法とベースラインを用いる．表 7.2 に提案手法とベースラインシステムの自動評価結果を示す．

表 7.2 提案手法とベースラインの自動評価結果

	BLEU	NIST	RIBES	TER	WER	METEOR
提案手法	0.22	3.50	0.82	57.01	57.0	0.50
ベースライン	0.23	3.67	0.87	55.63	57.47	0.53

表 7.2 の結果から，提案手法より得た日本語翻訳文はベースラインと比較してスコアが低下している．

### 7.3.2 人手評価結果

6.2 節と同様，人手による対比較評価を行った．自動評価と同様ベースラインシステムには Moses を用いる．表 7.3 に提案手法とベースラインの人手評価結果を示す．表 7.3 の

表 7.3 提案手法とベースラインの人手評価結果

提案手法○	提案手法×	差なし	同一出力
10	10	25	5

結果から提案手法○と提案手法×は同数であった．よって，提案手法の有効性は認められなかった．

### 7.3.3 人手評価における提案手法○の例

提案手法○の例を表 7.4 から表 7.6 に示す．



表 7.4 提案手法○の例 1

英語入力文	This is the utmost limit of my power .
英語文パターン	This is the X00 X01 of X02 .
日本語文パターン	X02 X00 X01 だ。
提案手法	私の力の限界だ。
ベースライン	これは私の努力もした。
参照文	これが私にとっては精いっぱいのところだ。

表 7.4 において、提案手法は参照文と同等の翻訳文が得られている。しかし、ベースラインは翻訳文として表現が不適切である。

表 7.5 提案手法○の例 2

英語入力文	Please have a seat .
英語文パターン	X00 have a X01 .
日本語文パターン	X01 X00 。
提案手法	どうぞ腰掛けてください。
ベースライン	座席をしてください。
参照文	どうぞ着席して下さい。

表 7.5 において、提案手法は参照文と同一の翻訳文が得られている。しかし、ベースラインは翻訳文として表現が不適切である。

表 7.6 提案手法○の例 3

英語入力文	The spirits muddled my brain .
英語文パターン	The X00 X01 my X02 .
日本語文パターン	X00 で X02 X01 た。
提案手法	酒で頭がぼんやりした。
ベースライン	私の頭は頭がぼんやりして spirits 。
参照文	私は酒で頭がぼんやりしてしまった。

表 7.6 において、提案手法は参照文と同等の翻訳文が得られている。しかし、ベースラインは未知語が出力されており、翻訳文として不適切である。

### 7.3.4 人手評価における提案手法×の例

提案手法×の例を表 7.7 から表 7.9 に示す。

表 7.7 提案手法×の例 1

英語入力文	The negotiations were conducted in a peaceful atmosphere .
英語文パターン	The X00 X01 in a X02 X03 .
日本語文パターン	X00 は、 X03 X02 X01 。
提案手法	交渉は、 大気中に終わった。
ベースライン	交渉は 平和な 雰囲気に行なわれた。
参照文	交渉は 穏やかに 行われた。

表 7.7 において、提案手法は翻訳文として不適切である。一方でベースラインは参照文と同等の翻訳文が得られている。

表 7.8 提案手法×の例 2

英語入力文	The ring was nowhere to be found .
英語文パターン	The X00 X01 to be X02 .
日本語文パターン	X00 X02 X01 。
提案手法	指輪をしなかった。
ベースライン	リングはどこにも見当たらなかった。
参照文	指輪はどこにも見当たらなかった。

表 7.8 において、提案手法は入力文に対する翻訳として不適切である。一方でベースラインは参照文と同等の翻訳文が得られている。

表 7.9 提案手法×の例 3

英語入力文	The tower stands over 9 0 0 feet high .
英語文パターン	The X00 X01 X02 0 feet high .
日本語文パターン	X00 の X01 は X02 0 フィートの高さになっている。
提案手法	塔の上には 9 0 0 フィートの高さになっている。
ベースライン	その塔は 9 0 0 フィートの高さになっている。
参照文	その塔は 9 0 0 フィート以上ある。

表 7.9 において，提案手法は翻訳文として意味が通じない．一方でベースラインは参照文と同等の翻訳文が得られている．

### 7.3.5 提案手法× における日本語翻訳文の改善例

文パターン選択数を 1,000 文から 10,000 文に変更したことで，表 6.3 における提案手法×の例の日本語翻訳文に変化が見られた．表 7.10 に改善例を示す．

表 7.10 提案手法×の改善例

英語入力文	This company has a good staff of talents .
英語文パターン	This X00 has a X01 of X02 .
日本語文パターン	この X00 に X02 X01 ある。
提案手法	この会社は能力がある。
ベースライン	この会社の社員に才能がある。
参照文	この会社には人材が揃っている。

表 6.3 は日本語翻訳文として意味が通じない文章であった．しかし，表 7.10 はベースラインには劣るが，表 6.3 と比較して優れていると言える．

### 7.3.6 評価結果のまとめ

表 7.2 の結果から提案手法はベースラインに劣っている．また，表 7.3 の結果から提案手法はベースラインと比較して必ずしも精度が良いとは言えない．よって，評価結果から提案手法の有効性は認められなかった．

## 第8章 考察

本研究ではパターン翻訳のコスト，翻訳精度，カバー率の問題に着目し，改善策を提案した．コストの問題はプログラムによる自動作成が可能となりコストの削減に成功した．よって本章では翻訳精度とカバー率，および翻訳精度とカバー率がトレードオフの関係にある問題を考察する．また，

### 8.1 翻訳精度の問題

6章より，提案手法はベースラインよりも優れていることがわかった．Moses とに対比較評価を行ったところ，提案手法○を6文，提案手法×を3文得た．この3文に対して誤り解析を行った結果，不適切な英語文パターンの選択が原因であることがわかった．以下に表 6.6 における誤り解析の結果を述べる．

誤った日本語翻訳文が出力された原因として，英語入力文に対する不適切な文パターンの選択が挙げられる．仮に表 8.1 に示すフレーズ辞書と表 8.2 に示す文パターンを用いた場合，適切な日本語翻訳文が得られることがわかった．なお，表 8.1 に示すフレーズと表 8.2 に示す文パターンは本研究において実際に生成されたが選択されなかったデータである．

表 8.1 フレーズの例

英語フレーズ	The quarrel	years
日本語フレーズ	けんかは	何年も

表 8.2 文パターンの例

英語入力文	The quarrel lasted for years .
英語文パターン	X00 lasted for X01 .
日本語文パターン	その X00 X01 続いた。
提案手法	その けんかは何年も続いた。

以上から、本実験で生成したフレーズ対と文パターンを用いることで適切な日本語翻訳文の出力が可能であることがわかった。よって、文パターンの選択方法の改善が必要である。具体的には文パターン辞書に用いる閾値の見直しが必要であると考えている。

## 8.2 カバー率の問題

本研究では、句に基づく文パターンの選択数を 1,000 文として、英日翻訳を行った。結果、英語入力文 100 文から 24 文の日本語翻訳文を得た。(以降、24 文の日本語翻訳文は“24 文(文パターン選択数:1,000)”と呼ぶ。)また、追加実験として、句に基づく文パターンの選択数を 10,000 文として、英日翻訳を行った。結果、英語入力文 100 文から、50 文の日本語翻訳文を得た。(以降、50 文の日本語翻訳文は“50 文(文パターン選択数:10,000)”と呼ぶ。)よって、文パターンの選択数を増加することで、カバー率の向上が認められた。しかし、50 文(文パターン選択数:10,000)の日本語翻訳文に対して翻訳精度の調査を行ったところ、提案手法の有効性は認められなかった。

## 8.3 翻訳精度とカバー率の問題

翻訳精度を調査したところ、24 文(文パターン選択数:1,000)の日本語翻訳文はベースラインと比較して自動評価、人手評価共に有効性が認められた。しかし、50 文(文パターン選択数:10,000)の日本語翻訳文に対しては自動評価、人手評価共に有効性が認められなかった。原因として、各辞書の作成に用いた閾値に問題があると思われる。今後は翻訳精度の向上を図るために、フレーズ辞書や句に基づく文パターン辞書に用いる閾値の見直しが必要であると考えている。

## 第9章 おわりに

本研究では、句に基づく対訳文パターンをプログラムで自動的に作成した。そして統計的手法を用いて英日パターン翻訳を行った。実験の結果、英語入力文 100 文から日本語翻訳文 24 文を得た。提案手法を評価するために、Moses をベースラインとし、提案手法との対比較評価を行った。対比較評価の結果、提案手法○が 6 文、提案手法×が 3 文であり、提案手法の有効性が認められた。

また、提案手法×の日本語翻訳文に対して誤り解析を行った。誤り解析の結果、不適切な英語文パターンの選択が原因であることがわかった。今後は各辞書に用いる閾値の見直しにより、さらに翻訳精度が向上すると考えている。

## 謝辞

最後に、二年間に渡り、本研究の御指導をいただきました鳥取大学大学院工学研究科  
計算機工学C講座研究室の村上仁一准教授，村田真樹教授，徳久雅人講師に心からお礼  
申し上げます。また，ご多忙の中，助言を頂きました北村章教授に厚くお礼申し上げま  
す。また，計算機工学C講座研究室の皆様，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] Language and Machines - Computers in Translation and Linguistics: A Report by the Automatic Language Processing Advisory Committee, Division of Behavioral Sciences, National Academy of Sciences, National Research Council, Publication 1416, First Printing November 1966.
- [2] Hiroshi Maruyama: "Pattern-Based Translation: Context-Free Transducer and Its Applications to Practical NLP", in Proc. of Natural Language Pacific Rim Symposium, pp.232-237, 1993.
- [3] Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, Robert L. Mercer: "The mathematics of statistical machine translation: Parameter Estimation", Computational Linguistics, 1993.
- [4] Franz Josef Och, Hermann Ney: "A Systematic Comparison of Various Statistical Alignment Models", Computational Linguistics, pp.19-51, 2003.
- [5] Philipp Koehn, Marcello Federico, Brooke Cowan, Richard Zens, Chris Dyer, Ondřej Bojar, Alexandra Constantin, Evan Herbst, "Moses: Open Source Toolkit for Statistical Machine Translation", Proceedings of the ACL 2007 Demo and Poster Sessions, pages 177-180, June 2007.
- [6] Franz Josef Och: "Minimum Error Rate Training in Statistical Machine Translation", In Proceeding of the 41st Annual Meeting of the Association for Computational Linguistics, pp.160-167, 2003.
- [7] Papineni Kishore, Salim Roukos, Todd Ward, Wei-Jing Zhu: "BLEU: a method for automatic evaluation of machine translation", 40th Annual meeting of the Association for Computational Linguistics, pp.311-318, 2002.



- [8] NIST, “Automatic Evaluation of Machine Translation Quality Using N-gram Co-Occurrence Statistics”, 2002.
- [9] 平尾努, 磯崎秀樹, Kevin Duh, 須藤克仁, 塚田元, 永田昌明: “RIBES: 順位相関に基づく翻訳の自動評価法”, 言語処理学会第 17 年次大会発表論文集, pp.1111-1114, 2011.
- [10] Richard Schwartz, Linnea Micciulla, John Makhoul: “A Study of Translation Edit Rate with Targeted Human Annotation”, AMTA, 2006.
- [11] Banerjee Satanjeev, Lavie Alon: “METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments”, Proceedings of Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization at the 43th Annual Meeting of the Association of Computational Linguistics (ACL-2005), pp.65-72, 2005.
- [12] 江木孝史, 村上仁一, 村田真樹, 徳久雅人: “統計的手法を用いた英日パターン翻訳”, 言語処理学会第 18 回年次大会, pp.263-266, 2012.
- [13] 西山七絵, 村上仁一, 徳久雅人, 池原悟: “単文文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.
- [14] Taku Kudo, Kaoru Yamamoto, Yuji Matsumoto: “Applying Conditional Random Fields to Japanese Morphological Analysis”, Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP-2004), pp.230-237, 2004.
- [15] Andreas Stolcke: “SRILM - an Extensible Language Modeling Toolkit”, 7th International Conference on Spoken Language Processing, pp.901-904, 2002.
- [16] 西村拓哉, 村上仁一, 村田真樹, 徳久雅人: “日英対訳文対を用いたパターン翻訳器の自動作成の検討”, 言語処理学会第 18 回年次大会, pp.159-162, 2012.
- [17] 松本拓也, 村上仁一, 徳久雅人: “機械翻訳における人手評価と自動評価の考察”, 言語処理学会第 18 回年次大会, pp.505-508, 2012.