

概要

パターン翻訳は、入力された原言語文に対し、フレーズ辞書と文パターン辞書を利用して翻訳文を出力する方法である。パターン翻訳は、入力文が適切な文パターンに適合した場合、翻訳精度の高い文を出力する傾向にある。しかし、フレーズと文パターンを人手作成するために、開発にコストがかかる。

開発コストを削減するため、江木は、フレーズ辞書と文パターン辞書を自動作成する、Pattern Based SMT(以下、全自動手法)を提案した。全自動手法は、統計的手法を用いて、フレーズ辞書と文パターンを自動作成して翻訳を行う。そのため、人手作成よりも、開発の時間とコストが低くなった。しかし、翻訳精度はまだ低い。

そこで、坂東は、翻訳精度の向上のために、人手作成されたフレーズ辞書を用いた Pattern Based SMT(以下、半自動手法)を提案した。その結果、半自動手法は、翻訳精度の高い文が出力されるようになった。しかし、翻訳精度の低い文も出力されるようになり、翻訳文全体で見ると翻訳精度が低くなった。この原因は、人手作成したフレーズ辞書は、自動作成したフレーズ辞書よりも文パターンになるフレーズの数が少ないので、文パターンの数が少なくなる。そのため、未知語を含む文が多く出力されたと考えられる。

本研究では、文パターンの数を増やすことで、翻訳文全体の翻訳精度の向上を目指した。その方法として、半自動手法で得られたフレーズ辞書と全自動手法で得られた文パターン辞書を用いて、パターンに基づく統計翻訳を行い、翻訳精度の調査を行った。また、全自動手法、半自動手法とそれぞれ対比較評価を行った。

対比較評価の結果、提案手法と半自動手法の翻訳精度に大きな差がないことがわかった。しかし、提案手法と全自動手法との比較では、提案手法の方が全自動手法よりも翻訳精度が優れていた。よって、人手作成したフレーズ辞書の有効性を示せた。

目次

第1章	はじめに	1
第2章	パターン翻訳	2
2.1	概要	2
2.2	手順	2
2.3	文パターン辞書	3
第3章	統計翻訳	4
3.1	概要	4
3.2	単語に基づく統計翻訳	5
3.2.1	IBM 翻訳モデル	5
3.2.2	GIZA++	6
3.3	句に基づく統計翻訳	6
3.4	言語モデル	7
3.5	デコーダ	8
第4章	全自動手法	9
4.1	概要	9
4.2	手順	9
4.2.1	単語辞書の作成	9
4.2.2	単語に基づく文パターン辞書(自動)の作成	9
4.2.2.1	日本単語照合	9
4.2.2.2	英語単語照合	9
4.2.2.3	変数化	10
4.2.3	フレーズ辞書(自動)の作成	10
4.2.3.1	パターン照合	10
4.2.3.2	フレーズの抽出	11

4.2.3.3	フレーズ対数確率の計算	11
4.2.4	句に基づく文パターン辞書の作成	13
4.2.4.1	日本語句の照合	13
4.2.4.2	英語句の照合	13
4.2.4.3	変数化	13
4.2.4.4	文パターン対数確率の付与	14
4.2.5	翻訳文の出力	16
4.2.5.1	日本語文パターンの選択	17
4.2.5.2	日本語句の取得	17
4.2.5.3	英語文パターンの取得	17
4.2.5.4	英語句の取得	17
4.2.5.5	英語翻訳候補文の生成	17
4.2.5.6	言語確率 (tri-gram) の算出	17
4.2.5.7	英語翻訳文の選択	17
第 5 章	半自動手法	18
5.1	概要	18
5.2	手順	18
5.2.1	フレーズ辞書 (手動) の作成	18
5.2.2	句に基づく文パターン辞書 (半自動) の作成	18
5.2.3	翻訳文の出力	19
第 6 章	提案手法	21
6.1	フレーズ辞書 (手動) の作成	21
6.2	句に基づく文パターン辞書の作成	21
6.3	翻訳文の出力	22
第 7 章	実験環境	23
7.1	日英対訳文	23
7.2	フレーズ	23
7.2.1	鳥バンク	23
7.2.2	文パターン数	24
7.2.3	フレーズ数	24

7.3	翻訳モデルの学習	25
7.4	言語モデルの学習	25
7.5	実験内容	25
7.6	評価方法	25
第 8 章	実験結果	26
8.1	提案手法と半自動手法の比較	26
8.2	提案手法と全自動手法の比較	27
8.3	英語翻訳文の例	27
8.3.1	提案手法 VS 半自動手法	27
8.3.2	提案手法 VS 全自動手法	33
8.4	実験結果のまとめ	37
第 9 章	考察	38
9.1	提案手法と mooses の比較	38
9.1.1	英語翻訳文の例	38
9.2	翻訳精度	40
第 10 章	おわりに	41

目 次

2.1	日英パターン翻訳の手順	3
3.1	日英統計翻訳の流れ	4
4.1	単語に基づく文パターン辞書 (自動) の作成手順	10
4.2	フレーズの抽出手順	11
4.3	日英フレーズ対数確率の付与手順	12
4.4	句に基づく文パターン辞書 (全自動) の作成手順	14
4.5	文パターン対数確率の付与手順	15
4.6	英語翻訳文の出力手順 (全自動)	16
5.1	句に基づく文パターン辞書 (半自動) の作成手順	19
5.2	英語翻訳文作成手順 (半自動)	20
6.1	英語翻訳文出力手順 (提案手法)	22

表 目 次

2.1	日英パターン翻訳の例	3
7.1	日英対訳文数	23
7.2	鳥バンクから抽出したフレーズの例	23
7.3	文パターン数	24
7.4	日本語文パターン例	24
7.5	英語文パターン例	24
7.6	フレーズ数	25
7.7	フレーズ例	25
8.1	提案手法と半自動手法の対比較評価結果	26
8.2	提案手法と全自動手法との対比較評価結果	27
8.3	提案手法 の例 1	28
8.4	変数対応	28
8.5	提案手法 の例 2	28
8.6	変数対応	29
8.7	提案手法 の例 3	29
8.8	変数対応	29
8.9	半自動手法 の例 1	30
8.10	変数対応	30
8.11	半自動手法 の例 2	31
8.12	変数対応	31
8.13	半自動手法 の例 3	31
8.14	変数対応	32
8.15	提案手法 の例 1	33
8.16	変数対応	33
8.17	提案手法 の例 2	33

8.18	変数対応	34
8.19	提案手法 の例 3	34
8.20	変数対応	34
8.21	提案手法 の例 4	35
8.22	変数対応	35
8.23	全自動手法 の例 1	35
8.24	変数対応	36
8.25	全自動手法 の例 2	36
8.26	変数対応	36
9.1	提案手法と moses との対比較評価結果	38
9.2	提案手法 の例	38
9.3	変数対応	39
9.4	moses の例	39
9.5	変数対応	39

第1章 はじめに

パターン翻訳 [1] は, 入力された原言語文に対し, フレーズ辞書と文パターン辞書を利用して翻訳文を出力する方法である. しかし, フレーズ辞書と文パターン辞書は人手で作成するため, 開発にコストがかかる.

開発コストを削減するため, 江木は, フレーズ辞書と文パターン辞書を自動作成する Pattern Based SMT[2](以下, 全自動手法) を提案した. しかし, 翻訳精度はまだ低い. そこで, 坂東は, 翻訳精度の向上のために人手作成されたフレーズ辞書を用いた Pattern Based SMT[3](以下, 半自動手法) を提案した. その結果, 半自動手法は翻訳精度の高い文が出力されるようになった. しかし, 翻訳文全体で見ると翻訳精度が低くなった.

この原因は, 人手作成したフレーズ辞書は自動作成したフレーズ辞書よりもフレーズの数が少ないため, 文パターンの数が少なくなる. そのため, 未知語を含む文が多く出力されたことが考えられる.

そこで, 本研究では, 文パターンの数を増やすことで翻訳文全体の翻訳精度の向上を目指す. その方法として, 半自動手法で得られたフレーズ辞書と, 全自動手法で得られた文パターン辞書を用いる.

本研究では日英方向のパターン翻訳を行う. フレーズ辞書と, 文パターン辞書を用いて得られた英語翻訳文に対して, 言語翻訳確率 (tri-gram) を用いた絞込みを行い, 出力する英語翻訳文を決定する. その後, 対比較実験により, 半自動手法, 全自動手法とそれぞれ 100 文ずつ出力された英語翻訳文を人手評価によってどちらが優れているかを判断し, 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章で従来のパターン翻訳システムについて説明し, 第3章で統計翻訳システムについて説明する. 第4章で全自動手法の説明, 第5章で半自動手法の説明をして, 第6章で提案する翻訳システムについて説明する. 第7章で実験条件を述べ, 第8章で実験結果を示す. 第9章で本研究の考察を述べる.

第2章 パターン翻訳

2.1 概要

パターン翻訳とは、機械翻訳手法の一つである。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、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、翻訳精度の高い翻訳文を得ることができる。しかし、問題点として、一般的なパターン翻訳は文パターンとフレーズを人手作成するため開発に時間がかかる。また、文パターンが適合しない場合は翻訳ができないという特徴がある。

2.2 手順

一般的な日英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人手で作成する。

手順2 日本語入力文と日本語文パターンを照合する。

手順3 変数部に対応する日本語を単語辞書を用いて英語単語に翻訳する。

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

手順5 手順4で得た英語翻訳文を出力する。

図 2.1 に日英パターン翻訳の手順を示す.

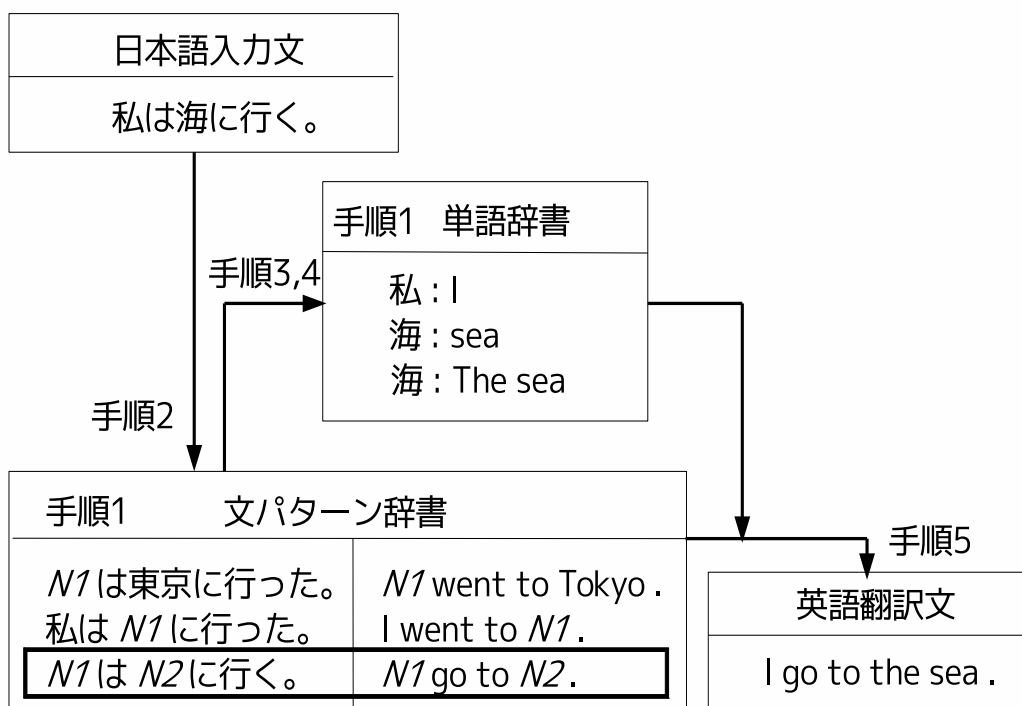


図 2.1: 日英パターン翻訳の手順

2.3 文パターン辞書

文パターン辞書は, 大量の対訳文から任意の単語やフレーズを人手で作成して得られる文パターンの集合である. 表 2.1 に日本語文パターン, 英語文パターンの例を示す.

表 2.1: 日英パターン翻訳の例

日本語入力文	私は海に行く。
日本語文パターン	私は N1 に行く。
英語入力文	I go to the sea .
英語文パターン	I go to N1 .

第3章 統計翻訳

統計翻訳システムを、原言語 (翻訳の対象となる入力された言語) を日本語、目的言語 (翻訳された後に出力される言語) を英語とする日英統計翻訳の場合を例として説明する。(なお、この章の説明は江木孝史: ”句に基づく文パターンを用いた英日翻訳” 第3章と、坂東俊樹: ”人手作成によるフレーズを用いたパターンに基づく統計翻訳” 第3章を参考にしている)

3.1 概要

統計翻訳は、機械翻訳手法の一つであり、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手法である。

日英統計翻訳は、日本語入力文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英語翻訳文を探索することで翻訳を行う。図 3.1 に日英統計翻訳の流れを示す。

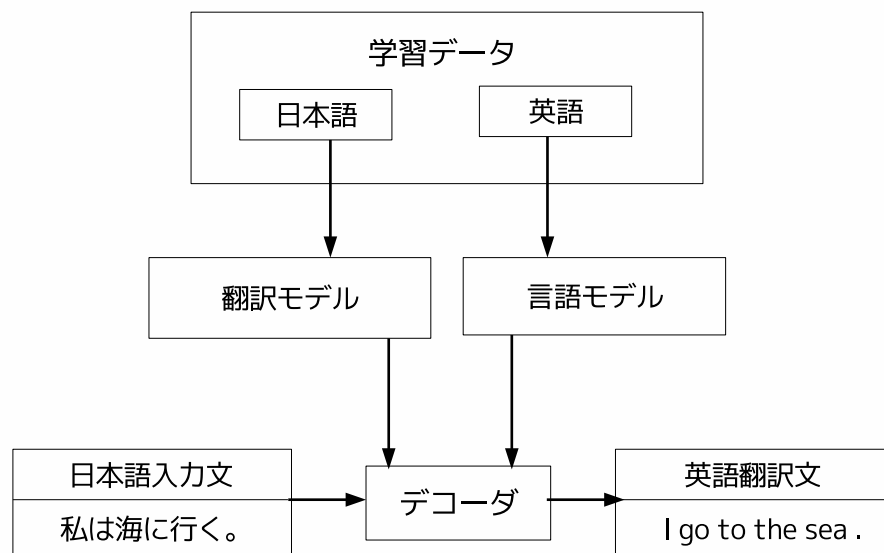


図 3.1: 日英統計翻訳の流れ

また、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る。以下に、単語に基づく統計翻訳と句に基づく統計翻訳の説明をそれぞれ示す。

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。日英翻訳の場合、日本単語を英語に翻訳し、日本単語の語順と同じ並びで英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

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。統計翻訳における単語対応を得るため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 [4] がある。IBM 翻訳モデルは、 $P(j|e)$ の近似方法の違いによって、model1 から model5 までの、順に複雑になる 5 つのモデルから構成されている。各モデルのおおまかな違いを以下に示す。

model1 目的言語における、ある単語が原言語の単語に対応する確率のみを用いる

model2 model1 に加えて、目的言語における、ある単語に対応する原言語の単語の原言語文中での位置の確率（以下、permutation 確率と呼ぶ）を用いる

model3 model2 に加えて、目的言語における、ある単語が原言語の何単語に対応するかの確率を用いる

model4 model3 における permutation 確率を改良して用いる（model2 の絶対位置に対して、相対位置）

model5 model4 における permutation 確率を更に改良して用いる

IBM 翻訳モデルは、仏英翻訳を前提としている。しかし、本研究では日英翻訳を扱っているため、本論文では、日英翻訳の場合を前提に説明する。

IBM 翻訳モデルでは、日本語文を j 、英語文を e として定義する。また、IBM 翻訳モデルにおいて、日本語文 J と英語文 E の翻訳モデル $P(j|e)$ を計算するため、アライメント a を用いる。以下に IBM 翻訳モデルの基本式を示す。なお、アライメントとは、ある日本語単語 j と英単語 e の対応関係のことを示す。

$$P(j|e) = \sum_a P(j, a|e)$$

IBM 翻訳モデルでは、日英統計翻訳の場合、英単語は 0 から n 個の対応を持ち、日本語の単語は 1 つの英単語のみと対応すると仮定する。また、日本語の単語の対応関係として適切な英単語がなかった場合、英語文の文頭の特殊文字 e と対応付けを行う。

3.2.2 GIZA++

GIZA++[5] は主に統計翻訳で使用されるツールであり、IBM 翻訳モデルを用いて、原言語と目的言語の対訳文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

2000 年代初期に提案された句に基づく統計翻訳は、句の対応を翻訳モデルに用いる。句に基づく統計翻訳は、句を構成する単語の数が、翻訳文の句と目的文の句で一致する必要がなくなり、単語に基づく統計翻訳の単語対応の問題を解決した。また、並べ替えにおいても、単語に基づく統計翻訳よりも優れている。そのため、近年では句に基づく統計翻訳が主流となっている。

3.4 言語モデル

言語モデルは、大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。日英翻訳では、言語モデルと翻訳モデルを用いて英語として自然な文を選出するために用いる。統計翻訳では主に N -gram を用いる。

・ N -gram モデル

統計翻訳では一般的に、 N -gram モデルを用いる。 N -gram モデルは、“単語列 $w_1^n = w_1, w_2, w_3, \dots, w_n$ の i 番目の単語 w_i の生起確率 $P(w_i)$ は直前の $(N-1)$ の単語列 $w_{i-(N-1)}, w_{i-(N-2)}, w_{i-(N-3)}, \dots, w_{i-1}$ に依存する” という仮説に基づくモデルである。また、 $N = 1$ のモデルを uni-gram, $N = 2$ のモデルを bi-gram, $N = 3$ のモデルを tri-gram と呼ぶ。(なお、 $N = 4$ 以上は 4-gram など数値を用いて呼ぶ。)

計算式を以下に示す。

ここで、 w_i^j は i から j 番目までの単語列を表す。

$$P(w_1^n) = P(w_1) \times P(w_2|w_1) \times P(w_3|w_2|w_1) \times \dots P(w_n|w_1^{n-1}) \quad (3.1)$$

$$\approx P(w_1) \times P(w_2|w_1) \times P(w_3|w_2|w_1) \times \dots P(w_n|w_{n-(N-1)}^{n-1}) \quad (3.2)$$

$$= \prod_{i=1}^n P(w_i|w_{i-(N-1)}^{i-1}) \quad (3.3)$$

たとえば、“I play baseball .” という単語列に対して $N = 2$ とした bi-gram モデルの言語モデルを適合した場合、単語列が生成される確率は以下の式で計算される。

$$P(I \text{ play baseball .}) \simeq P(I) \times P(\text{play}|I) \times P(\text{baseball}|\text{play}) \dots P(.|\text{baseball}) \quad (3.4)$$

tri-gram モデルであれば、 $P(\text{baseball}|I \text{ play})$ 、4-gram モデルであれば $P(.|I \text{ play baseball})$ となる。3-gram の場合を考えると、“I play” という単語列の次に “baseball” が来る確率を考える。しかし、 N -gram モデルは局所的な情報であり、文法構造の情報を持たない。したがって、異なる文法構造間の翻訳は、同じ文法構造間の翻訳と比較して、翻訳精度が低下する傾向がある。そのため、 N -gram モデルにおいて、信頼できる値を算出するためには、大規模なコーパスを用いる必要がある。そこで、出現頻度の少ない単語列をモデルの学習から削除(カットオフ)する方法や、確率が0となるのを防ぐために、大きい確率を小さく、小さい確率を大きくするスムージング手法が提案されている。スムージングの代表的な手法にバックオフ・スムージングがある。バックオフ・スムージングは学習データに出現しない N -gram の値をより低い次数の $(N-1)$ -gram の値から推定する。

3.5 デコーダ

デコーダは、翻訳モデルと言語モデルの全ての組み合わせの中から確率が最大となる出力文を探索して翻訳を行う。代表的なデコーダに Moses[8] がある。

第4章 全自動手法

4.1 概要

全自動手法は、統計的手法を用いて、フレーズ辞書と、文パターン辞書の両方を自動作成する手法である。

4.2 手順

江木によって提案された全自動手法は大きく分けて5つのステップで翻訳を行う。日英翻訳の場合の手順を以下に示す。

4.2.1 単語辞書の作成

対訳学習文と GIZA++ を利用して単語辞書を作成する。

4.2.2 単語に基づく文パターン辞書 (自動) の作成

対訳学習文と単語辞書を用いて、単語に基づく文パターン辞書 (自動) を作成する。まず、GIZA++ を用いて日英方向と英日方向の対訳単語と単語翻訳確率を得る。次に、日英方向の単語翻訳確率と英日方向の単語翻訳確率を掛け合わせる。

4.2.2.1 日本単語照合

対訳文の日本単語と単語辞書の日本単語を照合する。

4.2.2.2 英語単語照合

日本単語に対応する英語単語と対訳文の英語単語を照合する。

4.2.2.3 変数化

日英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。

単語に基づく文パターン辞書(自動)の作成手順と例文を図4.1に示す。

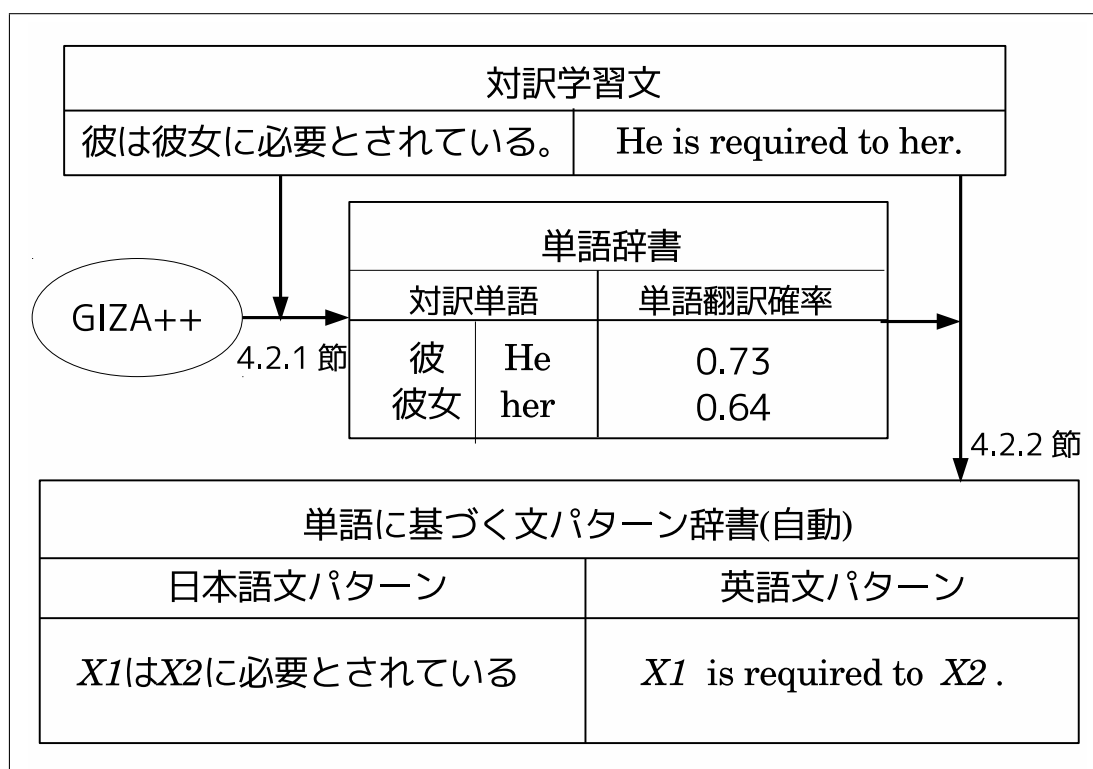


図 4.1: 単語に基づく文パターン辞書(自動)の作成手順

4.2.3 フレーズ辞書(自動)の作成

対訳学習文と単語に基づく文パターン辞書(自動)を用いて、フレーズを抽出する。次に、単語翻訳確率を用いて、フレーズにフレーズ対数確率を付与して、フレーズ辞書(自動)を作成する。以下にフレーズ辞書(自動)の作成手順を示す。

4.2.3.1 パターン照合

対訳文と単語に基づく文パターン辞書を照合する。

4.2.3.2 フレーズの抽出

対訳文と単語に基づく文パターン辞書を用いて、フレーズを抽出する。文パターン辞書対訳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フレーズを抽出する。図 4.2 にフレーズの抽出手順と例を示す。

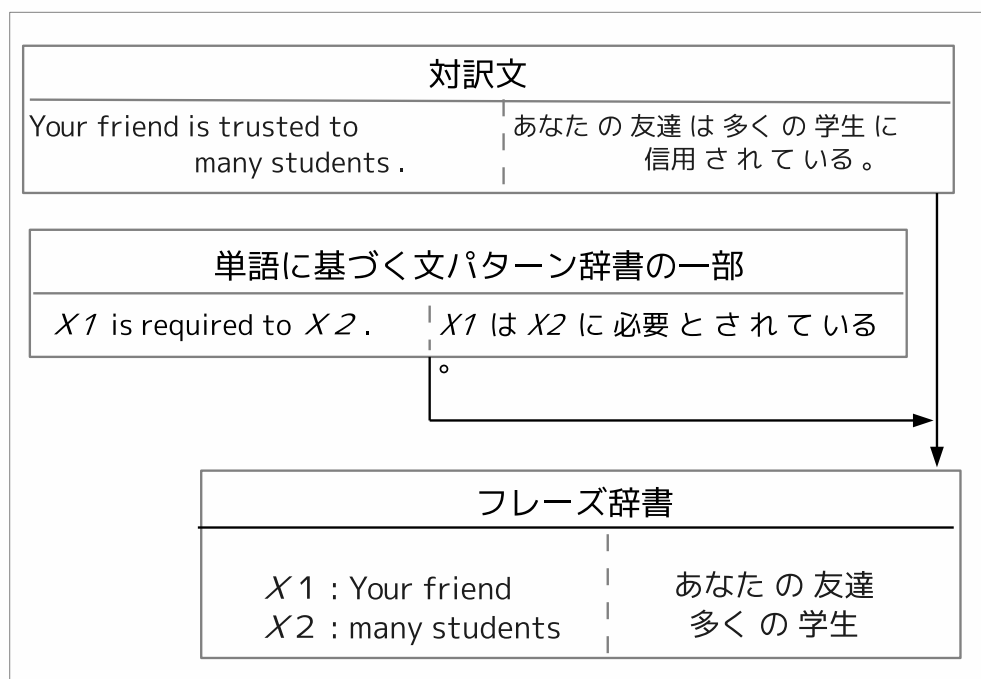


図 4.2: フレーズの抽出手順

4.2.3.3 フレーズ対数確率の計算

対訳フレーズと単語辞書を用いて、フレーズ対数確率を計算する。以下にフレーズ対数確率の計算手順を示す。

1. 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。

2. 翻訳確率の計算

日本語単語に対応する英語単語の中で、単語翻訳確率が最大となる確率を得る。同様に英語単語に対応する日本語単語の中で、単語翻訳確率が最大となる確率を得る。

3. フレーズ対数確率の付与

得られた確率に対して対数を取り、日英方向の単語翻訳確率の対数値の総和と英日方向の単語翻訳確率の対数値の総和を求める。次に、日英方向の総和と英日方向の総和を足し合わせて、フレーズのフレーズ対数確率として付与する。

図 4.3 に日英フレーズ対数確率の付与手順を示す。

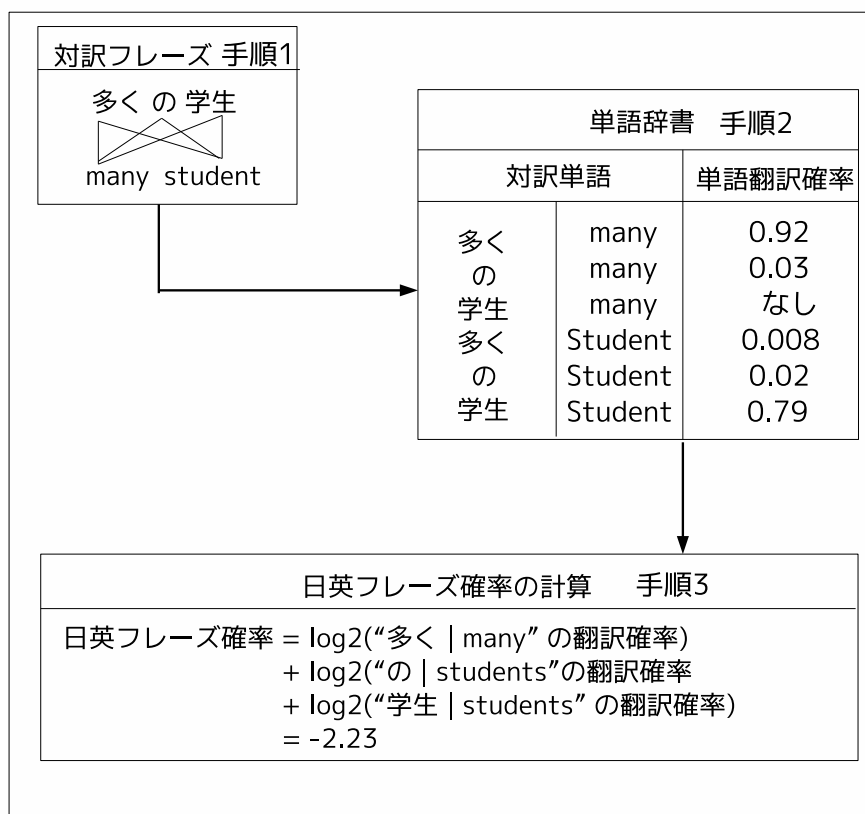


図 4.3: 日英フレーズ対数確率の付与手順

図 4.3 に英日方向のフレーズの例として“”多くの学生 | many students”を示す。まず、日本語句の単語と英語句の単語の全ての組み合わせを得る。次に、単語翻訳確率を用いて、各組み合わせの中から最大となる単語翻訳確率を得る。図 4.3 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” も同様に単語翻訳確率に対数を取り総和を求める。

4.2.4 句に基づく文パターン辞書の作成

対訳学習文とフレーズ辞書(自動)を用いて、句に基づく文パターン辞書(全自動)を作成する。次に、単語翻訳確率を用いて文パターン対数確率を付与し、句に基づく文パターン辞書を作成する。以下に句に基づく文パターン辞書の作成手順を示す。

4.2.4.1 日本語句の照合

対訳文における日本語文の各句とフレーズ辞書の日本語句を照合する。

4.2.4.2 英語句の照合

日本語句に対応する英語句と対訳文における英語文の各句を照合する。

4.2.4.3 変数化

フレーズが照合に成功した場合、該当箇所を変数化し、句に基づく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て、可能な限り多くの句に基づく文パターンを生成する。図 4.4 に句に基づく文パターンの作成手順と例を示す。

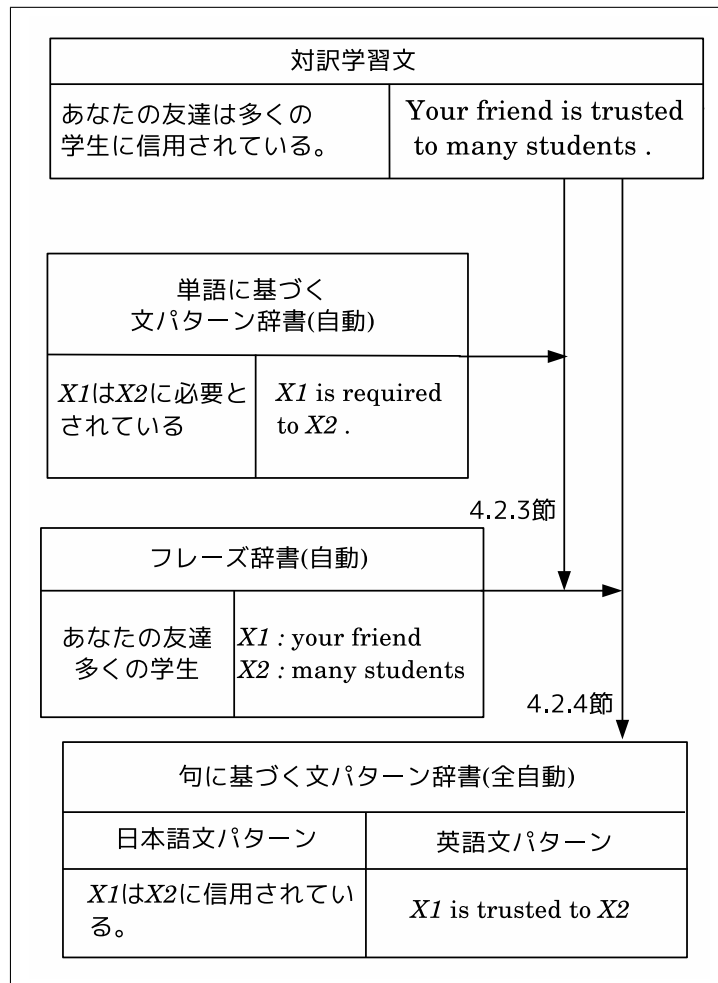


図 4.4: 句に基づく文パターン辞書 (全自動) の作成手順

4.2.4.4 文パターン対数確率の付与

文パターンの字面と単語翻訳確率を用いて、文パターンに翻訳確率を付与する。翻訳確率の付与は日英文パターンと英日文パターンに対して行う。計算した確率を文パターン対数確率と呼ぶ。なお、文パターン対数確率の付与は 4.2.3.3 節で説明したフレーズ対数確率の付与と同じ手順である。

図 4.5 に、日英方向の文パターン対数確率の付与手順と例を示す。

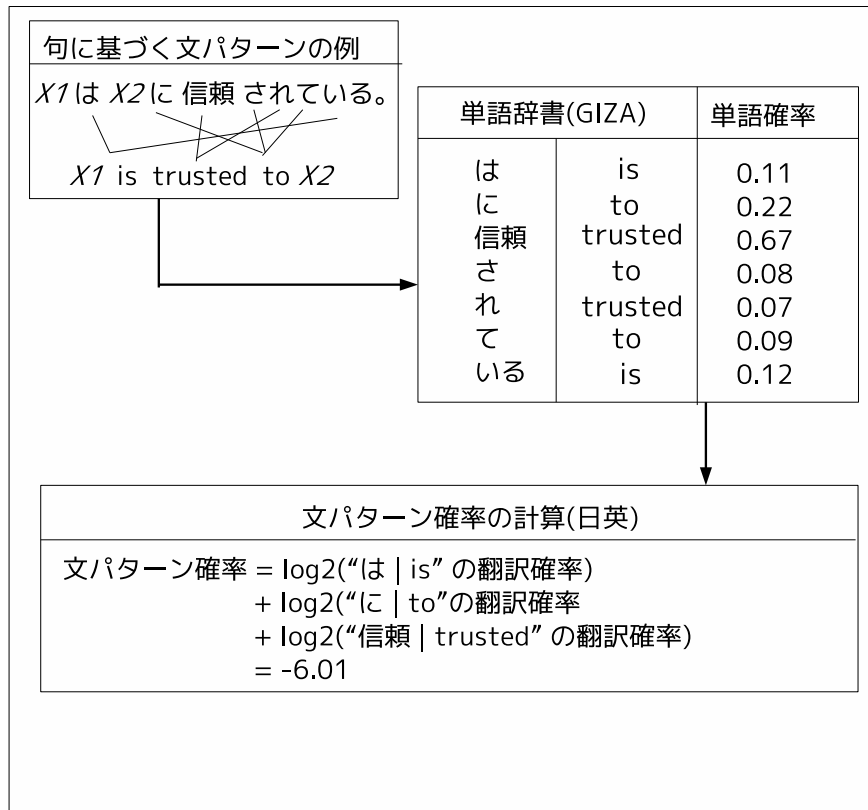


図 4.5: 文パターン対数確率の付与手順

図 4.5 に英日方向の句に基づく文パターンの例として“N1 は N2 に 信頼 されている。 | N1 is trusted to N2 .”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。次に、単語翻訳確率を用いて、各組み合わせの中から最大となる単語翻訳確率を得る。図 4.5 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”も同様に単語翻訳確率に対して対数を取り総和を求める。

4.2.5 翻訳文の出力

日本語文を入力して、フレーズ辞書(自動)と句に基づく文パターン辞書(全自動)を用いて、翻訳文を出力する。

翻訳精度を向上させるために、翻訳時に日本語入力文と日本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日本語文パターンから優先して選択する。図 4.6 日本語翻訳文を出力するまでの手順を示す。

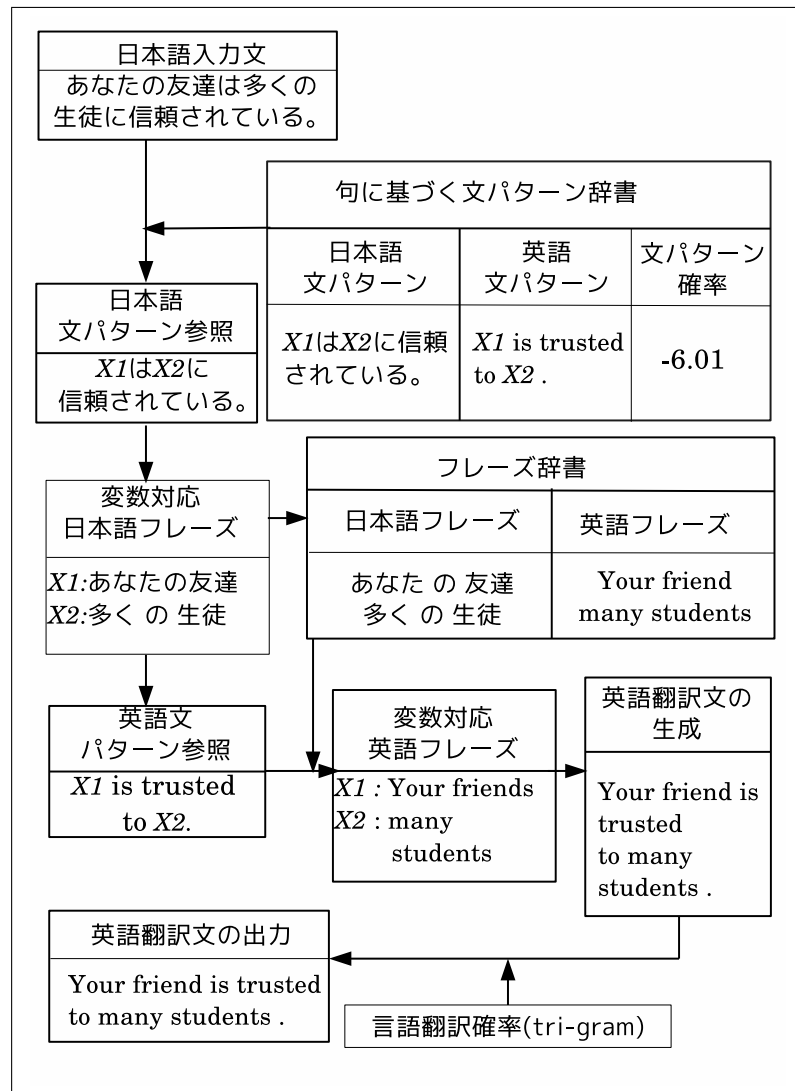


図 4.6: 英語翻訳文の出力手順(全自動)

4.2.5.1 日本語文パターンの選択

日本語文を入力とし、日本語入力文と日本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日本語文パターンから優先して選択する。図 4.6 では、“X1 は X2 に信頼されている。”がもっとも字面が一致する。

4.2.5.2 日本語句の取得

最も多く字面が一致する日本語文パターンの変数部に対応する日本語句を得る。図 4.6 では、“X1”には“あなたの友達”、“X2”には“多くの生徒”に対応する。次に、“X1 は X2 に信頼されている。”に対応する英語文パターン“X1 is trusted to X2.”を得る。

4.2.5.3 英語文パターンの取得

日本語文パターンに対応する英語文パターンを得る。図 4.6 では、“X1 は X2 に信頼されている。”に対応する英語文パターン“X1 is trusted to X2.”を得る。

4.2.5.4 英語句の取得

英語文パターンの変数部に対応する英語句を得る。図 4.6 では、“X1”には“Your friends”、“X2”には“many student”に対応する。

4.2.5.5 英語翻訳候補文の生成

英語文パターンの変数部を取得した英語句に置き換える。そして、英語翻訳候補文として生成する。

4.2.5.6 言語確率 (tri-gram) の算出

英語翻訳候補文に対して言語確率 (tri-gram) を計算する。

4.2.5.7 英語翻訳文の選択

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、英語翻訳候補文に付与する。最後に総和が最大となる英語翻訳文を出力する。

第5章 半自動手法

5.1 概要

半自動手法は、人手作成されたフレーズを用いてフレーズ辞書を作成することで、文パターン辞書を半自動手動で翻訳を行う。

5.2 手順

坂東によって提案された半自動手法の、日英翻訳の場合の手順を以下に示す。

5.2.1 フレーズ辞書 (手動) の作成

人手作成されたフレーズを用いて、フレーズ辞書 (手動) を作成する。次に、単語翻訳確率を用いて、フレーズにフレーズ対数確率を付与して、フレーズ辞書を作成する。

5.2.2 句に基づく文パターン辞書 (半自動) の作成

対訳学習文とフレーズ辞書 (手動) を用いて、句に基づく文パターン辞書 (半自動) を作成する。次に、単語翻訳確率を用いて文パターン対数確率を付与し、句に基づく文パターン辞書を作成する。半自動手法の句に基づく文パターン辞書 (半自動) の作成手順を図 5.1 に示す。

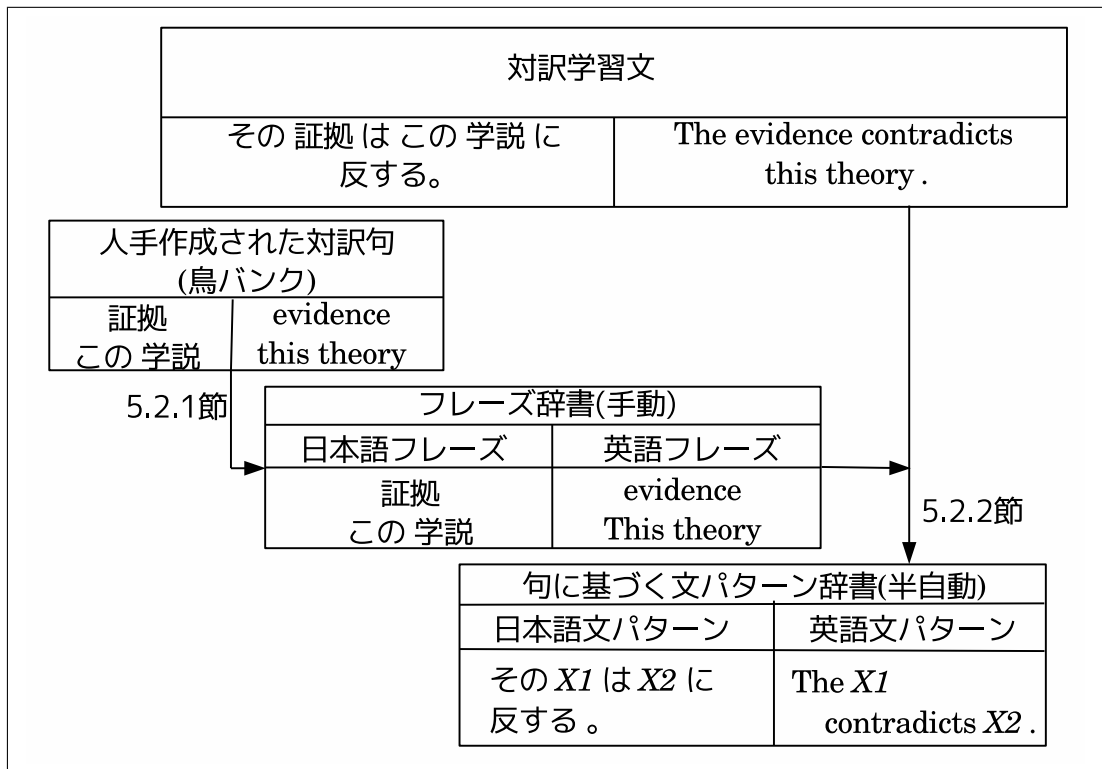


図 5.1: 句に基づく文パターン辞書 (半自動) の作成手順

5.2.3 翻訳文の出力

日本語文を入力して, フレーズ辞書 (手動) と句に基づく文パターン辞書 (半自動) を用いて, 翻訳文を出力する.

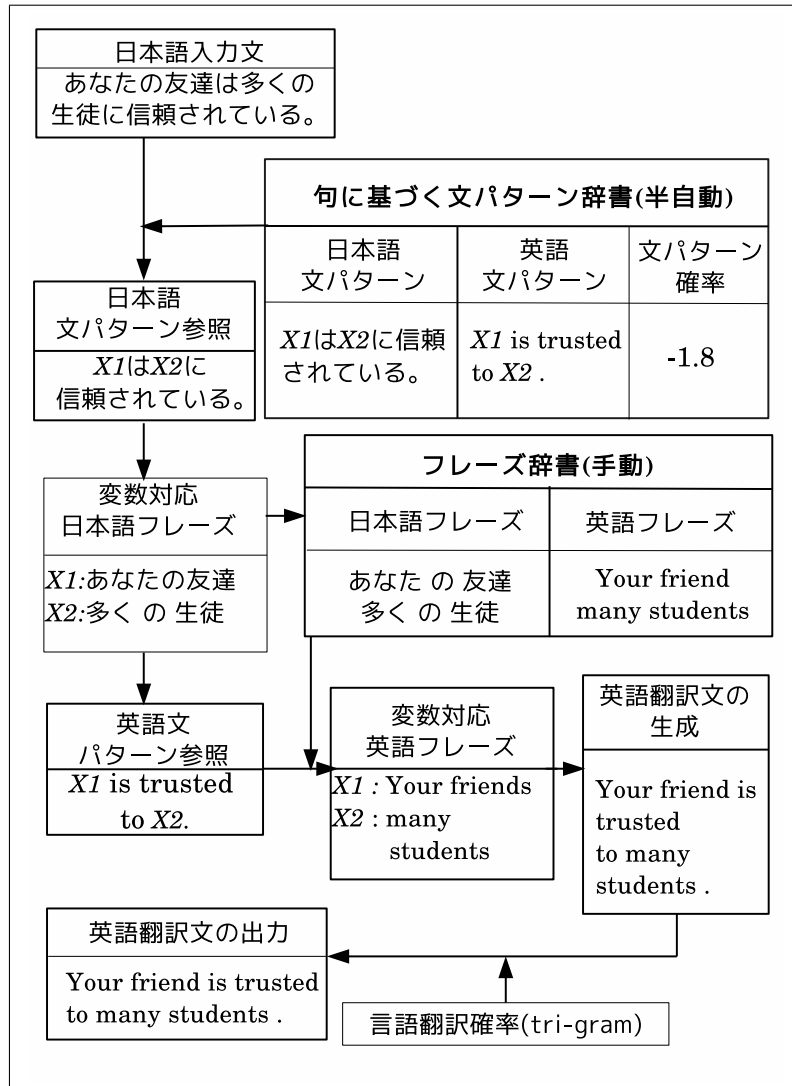


図 5.2: 英語翻訳文作成手順 (半自動)

第6章 提案手法

半自動手法は、翻訳精度の高い文が出力されるようになった。しかし、翻訳精度の低い文も出力されるようになった。そして、翻訳文全体で見ると翻訳精度が低くなった。この原因は、以下のことが考えられる。

- ・ 人手作成したフレーズ辞書は、自動作成されたフレーズ辞書よりもフレーズの数が少ない。
- ・ 半自動手法の文パターン辞書は全自動手法よりも文パターンの数が少ない。

これらより、半自動手法で翻訳精度の低い文が多く出力されたと考えられる。

そこで、文パターン辞書を増やすことで、翻訳文全体の翻訳精度を向上させる方法を提案する。具体的には、全自動手法で得られた文パターン辞書(全自動)と、半自動手法で得られたフレーズ辞書(手動)を用いて翻訳する。提案手法の手順を以下に示す。

6.1 フレーズ辞書(手動)の作成

半自動手法より、人手作成されたフレーズを用いてフレーズ辞書(手動)を作成する。次に、単語翻訳確率を用いて、フレーズにフレーズ対数確率を付与して、フレーズ辞書を作成する。なお、フレーズ辞書(手動)を作成する手順は5.2.1節と同様である。

6.2 句に基づく文パターン辞書の作成

対訳文とフレーズ辞書を用いて、句に基づく文パターン辞書(全自動)を作成する。次に、単語翻訳確率を用いて文パターン対数確率を付与し、句に基づく文パターン辞書を作成する。なお、句に基づく文パターン辞書の作成手順は4.2.4節と同様である。

6.3 翻訳文の出力

日本語文を入力とし、フレーズ辞書(手動)と句に基づく文パターン辞書(全自動)を用いて英語翻訳文を出力する。

翻訳精度を向上させるために、翻訳時に日本語文と日本語文パターンの字面を比較する。そして、最も多くの字面が一致する日本語文パターンを優先して選択する。英語翻訳文を出力するまでの手順は4.2.5節と同様である。

提案手法による翻訳手順を図6.1に示す

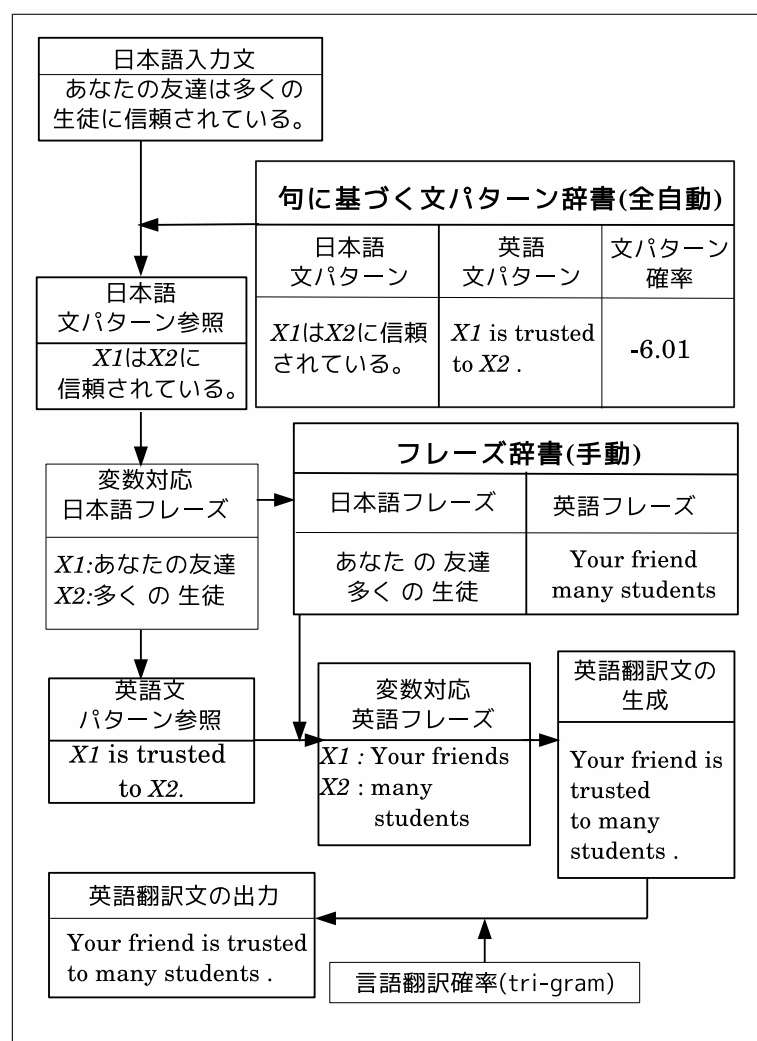


図 6.1: 英語翻訳文出力手順 (提案手法)

第7章 実験環境

7.1 日英対訳文

本研究では, 日英対訳文として, 単文コーパス [6] を用いる. 本研究では, 単文コーパスを表 7.1 の内訳で用いる.

表 7.1: 日英対訳文数

対訳文	100,000 文対
テストデータ	100 文

7.2 フレーズ

本研究では, 人手作成されたフレーズとして鳥バンクを用いる.

7.2.1 鳥バンク

鳥バンク [7] は, 日本語の重文と複文を対象として人手作成されたパターン辞書である. 本研究では, このパターン辞書から抽出したフレーズを用いる. フレーズの例を表 7.2 に示す.

表 7.2: 鳥バンクから抽出したフレーズの例

あなたのお父さん
Your father
息子の話
son's story
移民政策
immigration policy

7.2.2 文パターン数

本研究で文パターンとして用いる鳥バンクと自動作成の文パターン数を表 7.3 に示す.

表 7.3: 文パターン数

鳥バンク	6,196,879 句対
自動作成	203,351,967 句対

表 7.4 に日本語文パターンと日本語参照パターン, 表 7.5 に英語文パターンと英語参照パターンを鳥バンクと自動作成それぞれ示す.

表 7.4: 日本語文パターン例

	例	日本語文パターン	日本語参照パターン
鳥バンク	例 1	N01 で N00 た。	地下鉄で行った。
	例 2	N00 が N01 た。	池が凍った。
	例 3	N01 が N00 てきた。	陸地が見えてきた。
自動作成	例 1	本の N00 を N01。	本の漢字を全部覚えてください。
	例 2	N00 N01 じっと見つめた。	一瞬私をじっと見つめた。
	例 3	N01 が N00 ている。	包みがあいている。

表 7.5: 英語文パターン例

	例	英語文パターン	英語参照パターン
鳥バンク	例 1	We N 00 on the N 01 .	We went on the underground .
	例 2	The N 00 N 01 .	The pond froze .
	例 3	We came in N 00 of N 01 .	We came in sight of land .
自動作成	例 1	N 01 the N 00 in the book .	Please memorize all the kanji in the book .
	例 2	He stared at me hard N 01 N 00 .	He stared at me hard for a moment .
	例 3	The N 01 has been N 00 .	The package has been opened .

7.2.3 フレーズ数

本研究でフレーズコーパスとして用いる鳥バンクと自動作成のフレーズ数を表 7.6 に示す.

表 7.7 に鳥バンクと自動作成したフレーズの例を示す.

表 7.6: フレーズ数

鳥バンク	335,726 句対
自動作成	214,258 句対

表 7.7: フレーズ例

鳥バンク	例 1	N00: さんざんな目にあっ	N01: にわか雨
	例 2	N00: 土手	N01: 切れ
	例 3	N00: 高まっ	N01: 反対の声
自動作成	例 1	N00: 埃 dust	N01: はたく
	例 2	N00: 廊下 corridor	N01: を in the
	例 3	N00: 燃え	N01: 燃え

7.3 翻訳モデルの学習

翻訳モデルの学習には, `train-model.perl`[8] を用いる.

7.4 言語モデルの学習

言語モデルの学習には, SRILM[9] の `ngram-count` を用いる. 本研究では, N -gram モデルは 5-gram とする.

7.5 実験内容

本研究では, パターンに基づく統計翻訳を行い, 翻訳精度の調査を行う. また, 全自動手法と半自動手法との対比較評価を行う.

7.6 評価方法

本研究では, 翻訳システムによって出力した文の評価に対比較評価法を用いる. 対比較評価は, 二つの文を相対的に比較して, どちらがより正しい文であるかを人手で選択する評価方法である. 2つの翻訳システムの出力で優劣を判断する場合に有効である.

第8章 実験結果

8.1 提案手法と半自動手法の比較

提案手法と半自動手法で得られた英語翻訳文 100 文に対して、対比較評価を行った。表 8.1 に提案手法と半自動作成の対比較評価結果を示す。また、表 8.1 中の表記方法について説明する。表中の表記方法を表 8.1 をもとに説明する。

- 提案手法 : 提案手法の翻訳精度が半自動手法の翻訳精度より優れている文
- 半自動手法 : 半自動手法の翻訳精度が提案手法の翻訳精度より優れている文
- 差なし : 2 種類の翻訳精度が同程度である、もしくは両方とも翻訳文として適していない文
- 同一出力 : 2 種類の翻訳文が完全に同一な文

表 8.1: 提案手法と半自動手法の対比較評価結果

提案手法	半自動手法	差なし	同一出力
24	21	38	17

表 8.1 より、提案手法と半自動手法では差はほとんどない結果となった。

8.2 提案手法と全自動手法の比較

提案手法と全自動手法で得られた英語翻訳文 100 文に対して、対比較評価を行った。表 8.2 に提案手法と全自動手法の対比較評価の結果を示す。

表 8.2: 提案手法と全自動手法との対比較評価結果

提案手法	全自動手法	差なし	同一出力
37	14	44	5

表 8.2 より、提案手法と全自動手法では、提案手法の方が優れている結果となった。

8.3 英語翻訳文の例

8.3.1 提案手法 VS 半自動手法

表 8.3 に、表 8.5、表 8.7 に提案手法 の例を示す。表中の表記方法を表 8.3 をもとに説明する。

- “日本語入力文”：入力された日本語文
- “参照文”：日本語入力文と対になっている英語文
- “日本語文パターン”：英語翻訳文を出力する際に選択された日本語文パターン
- “英語文パターン”：日本語文パターンに対応する英語文パターン
- “英語翻訳文”：出力された英語翻訳文

表 8.3: 提案手法 の例 1

	日本語入力文	その土地はわれわれの村の共有である。
	参照文	The land belongs to our village .
提案手法	日本語文パターン	その N00 は N02 N01 N03 である。
	英語文パターン	The N 00 is N 02 N 03 N 01 .
	英語翻訳文	The land is our share of the village .
半自動手法	日本語文パターン	N00 は N02 の N01 N03 である。
	英語文パターン	N 00 is N 02 N 03 N 01 .
	英語翻訳文	The land is We share the village .

表 8.4 に変数部に対応するフレーズを示す。

表 8.4: 変数対応

変数名	提案手法	半自動手法
N00	土地 : land	その土地 : The land
N01	村の : of the village	村の : the village
N02	われわれの : our	われわれ : We
N03	共有 : share	共有 : share

表 8.3, 表 8.4 より, 半自動手法は変数部に対応するフレーズ (“われわれ : We) が英語文法的に不自然である。よって, 提案手法 と判断した。

表 8.5: 提案手法 の例 2

	日本語入力文	反対の声が高まってきた。
	参照文	Opposition to it became more clamorous .
提案手法	日本語文パターン	N00 が高まってきた。
	英語文パターン	N00 is increasing nowadays.
	英語翻訳文	opposition is increasing nowadays .
半自動手法	日本語文パターン	N01 が N00 てきた。
	英語文パターン	We came in N00 of N01 .
	英語翻訳文	We came in 高まつ of opposition .

表 8.6 に変数部に対応するフレーズを示す。

表 8.6: 変数対応

変数名	提案手法	半自動手法
N00	反対の声 : opposition	高まっ : 高まっ
N01		反対の声 : opposition

表 8.6 より, 半自動手法では, 変数に対応するフレーズはそれほど不自然ではない。しかし, 表 8.5 より, 文パターンに問題がある。よって, 提案手法 と判断した。

表 8.7: 提案手法 の例 3

	日本語入力文	2 人の走者はほぼ同時にゴールインした。
	参照文	The two runners reached the goal line almost at the same time .
提案手法	日本語文パターン	2 人の N03 は N02 N00 N01 た。
	英語文パターン	Their N 03 N 01 N 00 N 02 .
	英語翻訳文	Their runner reached the goal Simultaneously roughly .
半自動手法	日本語文パターン	N01 の N02 は N00 N03 N04 た。
	英語文パターン	N 01 N 02 N 00 N 03 N 04 .
	英語翻訳文	They runner almost Simultaneously reached the goal .

表 8.8 に変数部に対応するフレーズを示す。

表 8.8: 変数対応

変数名	提案手法	自動作成
N00	同時に : Simultaneously	ほぼ : almost
N01	ゴールイン し : reached the goal	2 人 : They
N02	ほぼ : roughly	走者 : runner
N03	走者 : runner	同時に : Simultaneously
N04		ゴールイン し : reached the goal

表 8.7, 表 8.8 より, 半自動手法は変数部に対応するフレーズ (“2 人 : They) が英語文法的に不自然である。よって, 提案手法 と判断した。

表 8.9, 表 8.11, 表 8.13 に半自動手法 の例を示す.

表 8.9: 半自動手法 の例 1

	日本語入力文	それは一頃 流行したものだ。
	正解文	It was once in fashion .
提案手法	日本語文パターン	N00 は N03 N02 N01 た もの だ 。
	英語文パターン	N 00 N 01 N 02 N 03 .
	英語翻訳文	It has fashion once .
半自動手法	日本語文パターン	N00 は N01 た もの だ 。
	英語文パターン	N 00 N 01 .
	英語翻訳文	It was once in fashion .

表 8.10 に変数に対応するフレーズを示す .

表 8.10: 変数対応

変数名	提案手法	半自動手法
N00	それ : It	それ : It
N01	し : has	一頃 流行 し : was once in fashion
N02	流行 : fashion	
N03	一頃 : once	

表 8.10 より, 半自動手法の英語翻訳文が正解文と同一であった. よって半自動手法 と判断した.

表 8.11: 半自動手法 の例 2

	日本語入力文	その土地はわれわれの村の共有である。
	正解文	The land belongs to our village .
提案手法	日本語文パターン	その N00 は N02 N01 N03 である。
	英語文パターン	The N 00 is N 02 N 03 N 01 .
	英語翻訳文	The land is our share of the village .
半自動手法	日本語文パターン	N00 は N02 の N01 N03 である。
	英語文パターン	N 00 is N 02 N 03 N 01 .
	英語翻訳文	The land is We share the village .

表 8.12 に変数に対応するフレーズを示す。

表 8.12: 変数対応

変数名	提案手法	半自動手法
N00	土地 : land	その土地 : The land
N01	村の : of the village	村の : the village
N02	われわれの : our	われわれ : We
N03	共有 : share	共有 : share

表 8.11, 表 8.12 より, 提案手法は変数部に対応するフレーズ (“われわれ : We) が英語文法的に不自然である。よって, 半自動手法 と判断した。

表 8.13: 半自動手法 の例 3

	日本語入力文	きわめて因習的な考えを持っている。
	正解文	She has very conventional beliefs .
提案手法	日本語文パターン	N02 N01 N00 を持っている。
	英語文パターン	N 02 N 01 N 00 .
	英語翻訳文	Very conventional idea .
半自動手法	日本語文パターン	N03 N01 的な N02 を N00 ている。
	英語文パターン	She N 00 a N 03 N 01 N 02 .
	英語翻訳文	She has a very 因習 idea .

表 8.14 に変数に対応するフレーズを示す .

表 8.14: 変数対応

変数名	提案手法	半自動手法
N00	考え : idea	持つ : has
N01	因習的な : conventional	因習 : 因習
N02	きわめて : Very	考え : idea
N03		きわめて : very

表 8.14 より, 提案手法は, 変数に対応するフレーズはそれほど不自然ではない. しかし, 表 8.13 より, 半自動手法に比べて“持っている”という日本語を翻訳できていない. よって半自動手法 と判断した.

8.3.2 提案手法 VS 全自動手法

表 8.15 に, 表 8.17, 表 8.19, 表 8.21 に提案手法 の例を示す.

表 8.15: 提案手法 の例 1

	日本語入力文	本の埃をはたく。
	参照文	Dust off a book .
提案手法	日本語文パターン	本の N00 を N01 。
	英語文パターン	N 01 the N 00 in the book .
	英語翻訳文	beat the dust in the book .
半自動手法	日本語文パターン	本の N00 を N01 。
	英語文パターン	N 01 the N 00 in the book . .
	英語翻訳文	gives the dust in the book .

表 8.16 に変数部に対応するフレーズを示す .

表 8.16: 変数対応

変数名	提案手法	全自動手法
N00	埃 : dust	埃 : dust
N01	はたく : beat	はたく : gives

表 8.16 より, 全自動手法は変数部に対応するフレーズ (“はたく : gives) が不自然である. よって, 提案手法 と判断した.

表 8.17: 提案手法 の例 2

	日本語入力文	家が燃えている。
	参照文	The house is burning .
提案手法	日本語文パターン	N01 が N00 ている。
	英語文パターン	The N 01 is N 00 .
	英語翻訳文	The house is burning .
全自動手法	日本語文パターン	N01 が N00 ている。
	英語文パターン	The N 01 has been N 00 .
	英語翻訳文	The house has been filled .

表 8.18 に変数部に対応するフレーズを示す。

表 8.18: 変数対応

変数名	提案手法	全自動手法
N00	燃え : burning	燃え : filled
N01	家 : house	家 : house

表 8.18 より, 全自動手法は変数部に対応するフレーズ (“燃え : filled) が不自然である。また提案手法の英語翻訳文が正解文と同一だったので, 提案手法 と判断した。

表 8.19: 提案手法 の例 3

	日本語入力文	コペルニクスの考えは人々を驚かせた。
	参照文	The thinking of Copernicus surprised the people .
提案手法	日本語文パターン	N01 の N02 は N00 を驚かせた。
	英語文パターン	N 01 N 02 surprised N 00 .
	英語翻訳文	コペルニクス idea surprised people .
全自動手法	日本語文パターン	N02 の N03 は N01 を N00 た。
	英語文パターン	N 01 N 03 N 00 N 02 .
	英語翻訳文	People think surprised コペルニクス .

表 8.20 に変数部に対応するフレーズを示す。

表 8.20: 変数対応

変数名	提案手法	自動作成
N00	人々 : people	驚か せ : surprised
N01	コペルニクス : コペルニクス	人々 : People
N02	考え : idea	コペルニクス : コペルニクス
N03		考え : think

表 8.20 より, 全自動手法は, 変数に対応するフレーズはそれほど不自然ではない。しかし 8.19 より, 英語翻訳文が “人々の考えはコペルニクスを驚かせた” となり翻訳文として相応しくない。よって, 提案手法 と判断した。

表 8.21, 表 8.25 に全自動手法 の例を示す.

表 8.21: 提案手法 の例 4

	日本語入力文	3つ目の角を右に曲がりなさい。
	正解文	Turn right at the third corner .
提案手法	日本語文パターン	N00の角を右に曲がりなさい。
	英語文パターン	Turn right at the N 00 corner .
	英語翻訳文	Turn right at the third corner .
全自動手法	日本語文パターン	N01 N00角を右に曲がりなさい。
	英語文パターン	Turn right N 00 N 01 corner .
	英語翻訳文	Turn right as the eye three corner .

表 8.22 に変数に対応するフレーズを示す .

表 8.22: 変数対応

変数名	提案手法	全自動手法
N00	3つ目 : third	目の : as the eye
N01		3つ : three

表 8.22 より, 全自動手法は変数部に対応するフレーズ (“目の : as the eye) が英語翻訳文的に不自然である. また, 提案手法の英語翻訳文が正解文と同一であった. よって提案手法 と判断した.

表 8.23: 全自動手法 の例 1

	日本語入力文	彼は車の方へ向かった。
	正解文	He made for the car .
提案手法	日本語文パターン	彼は N01 N00 へ N02 た。
	英語文パターン	He N 02 N 00 N 01 .
	英語翻訳文	He left side of the car .
全自動手法	日本語文パターン	彼は N01 の方 N00 N02 た。
	英語文パターン	He N 02 N 00 the N 01 .
	英語翻訳文	He headed off to the car .

表 8.24 に変数に対応するフレーズを示す。

表 8.24: 変数対応

変数名	提案手法	全自動手法
N00	方 : side	へ : off to
N01	車 の : of the car	車 : car
N02	向かっ : left	向かっ : headed

表 8.24 より, 提案手法の英語翻訳文は“車の側を去る”となり, 翻訳文として相応しくない。よって, 全自動手法 と判断した。

表 8.25: 全自動手法 の例 2

	日本語入力文	彼はちょっと考え込んだ。
	正解文	He bethought himself a moment .
提案手法	日本語文パターン	N00 はちょっと考え込んだ。
	英語文パターン	N 00 was thoughtful for a moment .
	英語翻訳文	She was thoughtful for a moment .
全自動手法	日本語文パターン	N03 N01 的な N02 を N00 ている。
	英語文パターン	N 01 N 00 thoughtful for a moment .
	英語翻訳文	He is thoughtful for a moment .

表 8.26 に変数に対応するフレーズを示す。

表 8.26: 変数対応

変数名	提案手法	全自動手法
N00	は : is	彼: She
N01	彼 : He	因習 : 因習

表 8.26 より, 提案手法は変数部に対応するフレーズ (“彼 : She) が不自然である。よって, 全自動手法 と判断した。

8.4 実験結果のまとめ

8.2 節より, 提案手法と全自動手法では提案手法のほうが翻訳精度が優れているということが分かった. この結果より, 人手作成したフレーズ辞書を使用することが有効であるということが分かった. しかし, 8.3.1 節より, 提案手法と半自動手法では翻訳精度に大きな差がないことがわかる. この結果から, 文パターン数をこれ以上増やしても, 翻訳精度が大きく向上する可能性は低いということが分かった.

第9章 考察

9.1 提案手法と moses の比較

提案手法と moses[8] で得られた英語翻訳文 100 文に対して, 対比較評価を行った. 表 9.1 に提案手法と全自動手法の対比較評価の結果を示す.

表 9.1: 提案手法と moses との対比較評価結果

提案手法	moses	差なし	同一出力
35	17	45	3

表 9.1 より, 提案手法と moses では, 提案手法の方が優れている結果となった.

9.1.1 英語翻訳文の例

表 9.2 に, 表 9.4, に提案手法 と moses の例を示す.

表 9.2: 提案手法 の例

	日本語入力文	昨夜 停電 が あった。
	参照文	We had a blackout last night .
提案手法	日本語文パターン	N01 N00 が あった。
	英語文パターン	There was a N 00 N 01 .
	英語翻訳文	There was a blackout last night .
moses	英語翻訳文	The last night .

表 9.3: 変数対応

変数名	提案手法
N00	停電 : blackout
N01	昨夜 : last night

表 9.2 より, moses の英語翻訳文は, ”停電” の部分の翻訳が出来ていない, 提案手法の英語翻訳文は日本語入力文として問題なかったので, 提案手法 と判断した

表 9.4: moses の例

	日本語入力文	彼は私の注意を引いた。
	参照文	He attracted my attention .
提案手法	日本語文パターン	彼は N02 の N00 を N01 た。
	英語文パターン	He N 01 his N 00 to N 02 .
	英語翻訳文	He drew his attention to me .
moses	英語翻訳文	He took my attention .

表 9.5: 変数対応

変数名	提案手法
N00	注意 : attention
N01	引い : drew
N02	私 : me

表 9.4 より, 提案手法の英語翻訳文, ”his attention to me .” 私の注意という翻訳文に適していないと考え, moses と判断した.

9.2 翻訳精度

本研究において、提案手法と全自動手法では、提案手法の方が優れていた。この結果より、人手作成したフレーズ辞書の有効性が示せた。しかし、提案手法と半自動手法では、翻訳精度に大きな差がなかった。この結果より、文パターンを増やしても翻訳精度に差がないという事が分かった。今後の課題として、文パターンの数を増やす以外の方法で翻訳精度を向上する方法を考える必要がある。そこで、文パターンの数を増やす以外に、フレーズ辞書の数を増やす方法が考えられる。具体的には、半自動手法のフレーズ辞書に全自動手法のフレーズ辞書を加える。この方法を、今後の課題として研究していきたい。

第10章 おわりに

本研究では、文パターンを増やすことで、翻訳文全体の翻訳精度を向上させる方法を提案し、その方法として、全自動手法で得られた文パターン辞書(全自動)と、半自動手法で得られたフレーズ辞書(手動)を用いて、パターンに基づく統計翻訳を行った。そして、全自動作成、半自動手法との対比較評価により、翻訳精度の調査を行った。実験の結果、提案手法が半自動手法と比べて24文(半自動手法 21文)、全自動手法と比べて37文(全自動手法 14文)であり、提案手法と半自動手法ではほぼ差がない結果となった。しかし、提案手法と全自動手法では、提案手法のほうが優れているという結果になった。

謝辞

最後に，1年間に渡りご指導いただきました鳥取大学工学部知能情報工学科計算機工学講座C研究室の村上仁一准教授，村田真樹教授そして，坂田純さんをはじめ，計算機工学講座C研究室の方々に厚く御礼申し上げます．

また，参考にさせていただいた論文の著者の方々に対して深く感謝申し上げます．

参考文献

- [1] 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.
- [2] 江木孝史：“句に基づく文パターンを用いた英日翻訳”，2014年修論
- [3] 坂東俊樹：“人手作成による対訳句を用いたパターンに基づく統計翻訳”，2015年卒業論文
- [4] 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.
- [5] GIZA++ : <http://www.fjoch.com/GIZA++>
- [6] 村上仁一, 藤波進 “日本語と英語の対訳文対の収集と著作権の考察”，第一回コーパス日本語学ワークショップ, pp.119-130. 2012.
- [7] 鳥バンク : <http://unicorn.ike.tottori-u.ac.jp/toribank/>
- [8] Moses: 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.
- [9] SRILM: Andreas Stolcke, “SRILM - an Extensible Language Modeling Toolkit”, 7th International Conference on Spoken Language Processing, pp.901-904, 2002.