

概要

現在，機械翻訳において統計翻訳が注目され，研究が盛んに行われている．統計翻訳では一般に N -gram モデルを用いる． N -gram モデルは局所的な文法情報であるため，特異な翻訳文が生成される場合がある．

そこで本研究では，まず日英文パターン辞書を用いて日英パターン翻訳を行い，文法構造を英語に近づける．このとき，従来の文パターン辞書は人手で作成するが，本研究では自動作成する．この日英文パターンが持つ大局的な文法情報を用いることで N -gram モデルにおける局所的な構文問題が解消でき，翻訳精度の向上が可能であると考えた．そして出力文に対し，統計翻訳でさらに英英翻訳を行う．この処理により，局所的な修正を行うことで翻訳精度が向上すると考えた．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，その有効性を調査した．

実験の結果，文パターン辞書の変数が少ない場合には文パターンに文法情報が多く残るため翻訳精度の高い文が得られ，提案手法の有効性が示された．この場合での自動評価結果は BLEU 値で 0.5%，METEOR 値では 0.08%，NIST 値で 0.116 の改善が見られた．しかし変数の多い場合には文法情報が損なわれ，翻訳精度が低下することがわかった．この場合での自動評価結果は BLEU 値で 0.4%，NIST 値で 0.095 の低下が見られた．

目次

| | | |
|---------|---------------|----|
| 第1章 | はじめに | 1 |
| 第2章 | 日英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 | パターン翻訳の手順 | 3 |
| 2.3 | 日英文パターン辞書 | 4 |
| 第3章 | 日英統計翻訳システム | 5 |
| 3.1 | 統計翻訳の概要 | 5 |
| 3.2 | 統計翻訳の手順 | 5 |
| 3.3 | 翻訳モデル | 6 |
| 3.3.1 | 翻訳モデルの概要 | 6 |
| 3.3.2 | IBM 翻訳モデル | 6 |
| 3.3.2.1 | モデル1 | 7 |
| 3.3.2.2 | モデル2 | 9 |
| 3.3.2.3 | モデル3 | 10 |
| 3.3.2.4 | モデル4 | 11 |
| 3.3.2.5 | モデル5 | 12 |
| 3.4 | GIZA++ | 12 |
| 3.5 | フレーズテーブル作成法 | 13 |
| 3.6 | 言語モデル | 17 |
| 3.6.1 | 言語モデルの概要 | 17 |
| 3.6.2 | N -gram モデル | 18 |
| 3.7 | デコーダ | 19 |
| 3.8 | パラメータチューニング | 19 |
| 3.9 | 評価方法 | 20 |

| | | |
|--------------|--------------------|-----------|
| 第 4 章 | 翻訳システム | 21 |
| 4.1 | 本研究の翻訳システム | 21 |
| 4.2 | 本研究の翻訳システムの手順 | 22 |
| 4.3 | ベースラインシステム | 25 |
| 第 5 章 | 実験環境 | 26 |
| 5.1 | 実験データ | 26 |
| 5.2 | 翻訳モデルの学習 | 26 |
| 5.3 | 言語モデルの学習 | 26 |
| 5.4 | デコーダのパラメータ | 27 |
| 第 6 章 | 翻訳実験 | 28 |
| 6.1 | 提案手法におけるパターン翻訳部の実験 | 28 |
| 6.2 | パターン翻訳部の実験結果 | 29 |
| 6.3 | 提案手法における統計翻訳部の実験 | 29 |
| 6.4 | ベースラインシステムにおける実験 | 30 |
| 第 7 章 | 実験結果 | 31 |
| 7.1 | 自動評価 | 31 |
| 7.1.1 | 実験条件 1 における評価結果 | 31 |
| 7.1.2 | 実験条件 2 における評価結果 | 32 |
| 7.2 | 人手による評価 | 33 |
| 7.2.1 | 判断基準 | 33 |
| 7.2.2 | 評価結果 | 35 |
| 第 8 章 | 考察 | 36 |
| 8.1 | 日英文パターン辞書における問題点 | 36 |
| 8.2 | パターン翻訳の出力における解析 | 38 |
| 8.2.1 | 実験条件 1 における解析 | 38 |
| 8.2.2 | 実験条件 2 における解析 | 42 |
| 第 9 章 | おわりに | 46 |

目次

| | | |
|-----|----------------------------------|----|
| 2.1 | 日英パターン翻訳の手順 | 3 |
| 3.1 | 日英統計翻訳の手順 | 6 |
| 3.2 | アライメントの例 | 7 |
| 4.1 | 提案手法の翻訳手順 | 21 |
| 4.2 | 提案手法における日英文パターン辞書の作成手順 | 22 |
| 4.3 | 提案手法におけるパターン翻訳の手順 | 23 |
| 4.4 | 提案手法における統計翻訳の手順 | 24 |

表 目 次

| | | |
|------|--------------------------------------|----|
| 2.1 | 日英文パターンの例 | 4 |
| 3.1 | フレーズテーブルの例 | 6 |
| 3.2 | 日英方向の単語対応 | 13 |
| 3.3 | 英日方向の単語対応 | 13 |
| 3.4 | intersection の例 | 14 |
| 3.5 | union の例 | 14 |
| 3.6 | grow の例 | 15 |
| 3.7 | grow-diag の例 | 15 |
| 3.8 | grow-diag-final の例 | 16 |
| 3.9 | grow-diag-final-and の例 | 16 |
| 3.10 | grow-diag-final-and で作成されたフレーズテーブルの例 | 16 |
| 3.11 | 言語モデルの例 | 17 |
| 4.1 | 日英対訳単語辞書の例 | 22 |
| 5.1 | 日英対訳文の例 | 26 |
| 6.1 | 日英対訳単語辞書の登録単語数 | 28 |
| 6.2 | 日本語学習文での適合した文数と出力文数 | 29 |
| 6.3 | 日本語テスト文での適合した文数と出力文数 | 29 |
| 7.1 | 10,000 文の結果 | 31 |
| 7.2 | 362 文の結果 | 31 |
| 7.3 | 10,000 文の結果 | 32 |
| 7.4 | 2,103 文の結果 | 32 |
| 7.5 | 対比較実験の結果 | 35 |
| 8.1 | 作成した日英文パターン辞書の例 | 36 |

| | | |
|-----|-------------------------|----|
| 8.2 | 実験条件 1:出力の成功例 | 38 |
| 8.3 | 実験条件 1:出力の失敗例 | 40 |
| 8.4 | 実験条件 2:出力の成功例 | 42 |
| 8.5 | 実験条件 2:出力の失敗例 | 44 |

第1章 はじめに

機械翻訳の歴史は文法規則や変換規則などを用いて翻訳を行うルールベース翻訳から始まる．そして1960年代半ばに，大量の翻訳対から作成した文パターン辞書を用いて翻訳を行うパターン翻訳が提案される．パターン辞書は人手で作成するので，開発に時間がかかる [1] が，文パターンに適合した場合に翻訳精度の高い翻訳文が得られる．1990年代前半に「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年の初めに「句に基づく統計翻訳 [2]」が提案され「語に基づく統計翻訳」と比べて翻訳精度が高いことから，現在，機械翻訳において統計翻訳が主流となっている．統計翻訳では，一般に N -gram モデルを文法情報として用いる． N -gram モデルは局所的な文法情報であるため，文法構造の似た言語間では翻訳精度が高い．しかし，日本語と英語のように文法構造が異なる言語間では特異な文が生成される可能性がある．過去に，日本語の動詞を主語の後ろに移動し文法構造を変換する研究 [3] がされているが翻訳精度はあまり向上していない．

そこで本研究では，まず日英文パターン辞書を用いて日英パターン翻訳を行い，文法構造を英語に近づける．この日英文パターンが持つ大局的な文法情報を用いることで N -gram モデルにおける局所的な構文問題が解消でき，翻訳精度の向上が可能であると考えた．そして出力文に対し，統計翻訳でさらに英英翻訳を行う．この処理により，局所的な修正を行うことで翻訳精度が向上すると考えた．実験の結果，従来の日英統計翻訳システムと比べて提案手法のシステムでは，文パターンの文法情報が多く残されている場合に翻訳精度が高く，有効性が確認出来た．この場合での自動評価結果は BLEU 値で 0.5%，METEOR 値では 0.08%，NIST 値で 0.116 の翻訳精度の向上が見られた．しかし文パターンの文法情報が損なわれている場合には翻訳精度が低下することがわかった．この場合での自動評価結果は BLEU 値で 0.4%，NIST 値で 0.095 の翻訳精度の低下が見られた．

本論文の構成は以下の通りである．第2章で従来の日英パターン翻訳システムについて説明し，第3章で従来の日英統計翻訳システムについて説明し，第4章では本研究で提案する翻訳システムについて説明する．第5章で実験環境を示し，第6章で翻訳実験

を述べ，第7章で実験結果を示し，第8章で本研究の考察を述べる．

第2章 日英パターン翻訳システム

2.1 パターン翻訳の概要

パターン翻訳とは、文パターン辞書と単語辞書を用いて、日本語文パターンに適合した情報を対応する英語文パターンに用いて適切な翻訳文を生成する、機械翻訳手法の1つである。パターン翻訳には文パターンが有する大局的な文法情報を用いることで英語文全体の構造を保持した翻訳精度の高い翻訳文を生成出来るという利点がある。しかし、従来の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という欠点がある。また、文パターン辞書に適合しない入力文については出力を行うことが出来ないため、入力文に対するカバー率が低い問題もある。

2.2 パターン翻訳の手順

日英パターン翻訳は、まず日本語入力文 j が与えられたとき、日英文パターン辞書と日英単語辞書を参照する。次に、日本語入力文に適合した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単語に置き換えることで翻訳を行う。図 2.1 に日英パターン翻訳の手順を示す。

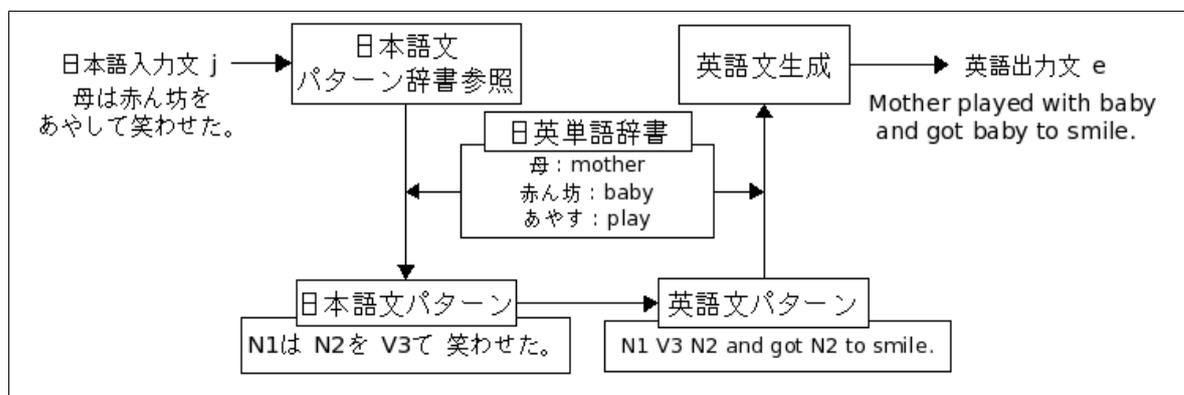


図 2.1 日英パターン翻訳の手順

2.3 日英文パターン辞書

日英文パターン辞書とは，大量の対訳コーパス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データベースである．

過去の研究に，単文での文パターン辞書作成が行われている [4]．作成方法としては，大量の対訳コーパスと対訳単語辞書を用いて，対となる名詞などを変数化する．また翻訳時の変数部を単語に置き換える際，置き換える単語の候補をしぼるために，品詞や字面などの意味属性制約を変数に付与する．この過程により，文パターン辞書を作成する．表 2.1 に [4] で作成された文パターンの例を示す．

表 2.1 日英文パターンの例

| | |
|----------|---|
| 日本語原文 | 彼はその戦闘で傷を受けた。 |
| 英語原文 | He was wounded in that battle. |
| 日本語文パターン | <i>PRO1</i> は <i>REN2N3</i> で傷を受けた。 |
| 英語文パターン | <i>PRO1</i> was wounded in <i>REN2 N3</i> . |

表 2.1 の例では，*PRO* は代名詞，*REN* は連体詞，*N* は名詞を表しており，翻訳時に置き換える単語が限定される．しかし，第??章で詳しく説明する，本研究で作成する文パターンでは意味属性制約などは使用せず，適合する単語の全ての組合せの出力を行う．

第3章 日英統計翻訳システム

3.1 統計翻訳の概要

統計翻訳とは、翻訳する言語と目的言語の対訳文を大量に収集した対訳データを用いて、自動的に翻訳規則を獲得し翻訳を行う、機械翻訳手法の1つである。統計翻訳以前のシステムとしてルールベース翻訳がある。この手法では、日本語と英語のルールを定義する必要がある。例えば、日本語のルール $[S \rightarrow NP \text{ は } VP]$ がある場合には、英語のルール $[S \rightarrow NP VP]$ といった、ルールの対が大量に必要となる。したがって、翻訳するためのルールを作成するには、文パターン辞書と同様に多言語化が困難であり、開発に時間がかかる。一方、統計翻訳では、統計的モデルを学習するアルゴリズムは言語に依存しない。したがって、大量の対訳データがあれば多言語化が容易であり、開発時間も抑えることが可能である。統計翻訳には、最初に提案された「語に基づく統計翻訳」と近年提案された「句に基づく統計翻訳」があり、現在は「句に基づく統計翻訳」が主流となっている。これは、語順の並び替えや文脈における訳語の選択や翻訳精度において、「句に基づく統計翻訳」の方が「語に基づく統計翻訳」よりも優れているためである。また統計翻訳の特徴として、文法構造が似ている言語間では翻訳精度が高い傾向があり、文法構造の異なる言語間では翻訳精度が低い傾向がある。

3.2 統計翻訳の手順

日英統計翻訳は、日本語入力文 j が与えられたとき、全ての組合せから確率が最大となる英語文 \hat{e} を探索して翻訳を行う。

$$\hat{e} = \arg \max_e P(e|j) \quad (3.1)$$

$$\approx \arg \max_e P(j|e)P(e) \quad (3.2)$$

$P(j|e)$ は翻訳モデル、 $P(e)$ は言語モデルと呼ぶ。図 3.1 に統計翻訳の手順を示す。

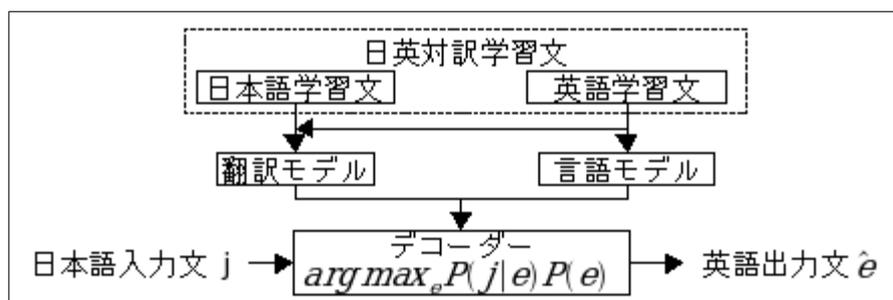


図 3.1 日英統計翻訳の手順

3.3 翻訳モデル

3.3.1 翻訳モデルの概要

翻訳モデルは英語から日本語の単語列へ確率的に翻訳を行うためのモデルである。翻訳モデルはフレーズテーブルで管理されている。フレーズテーブルの例を表 3.1 に示す。

表 3.1 フレーズテーブルの例

| | | | | | | | | | | | | | | | | | |
|-----|---|-------|--|---------------|--|-------|-----|-----|--|-----|-------|-----|--|-----|------|------|-------|
| 1 | 0 | パーセント | | a ten percent | | (0,1) | (1) | (2) | | (0) | (0,1) | (2) | | 1 | 0.02 | 0.25 | 0.005 |
| あなた | と | 私 | | you and I | | (0) | (1) | (2) | | (0) | (1) | (2) | | 0.5 | 0.01 | 0.5 | 0.03 |

左から、日本語フレーズ、英語フレーズ、日英フレーズ内単語対応、英日のフレーズ内単語対応、フレーズの日英方向の翻訳確率、日英の単語翻訳確率の積、フレーズの英日方向の翻訳確率、英日の単語翻訳確率の積である。

3.3.2 IBM 翻訳モデル

統計翻訳の翻訳モデルとして代表的なものに、Brown らが提案した IBM の仏英翻訳モデル [5] がある。この翻訳モデルは提案者である Brown らが全員 IBM の社員であったため、IBM 翻訳モデルと呼ばれている。この翻訳モデルは順に複雑な計算を行うモデル 1 からモデル 5 の 5 つのモデルから成る。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、フランス語文を F 、英語文を E として説明を行う。

IBM モデルでは英語文 E とフランス語文 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 モデルの基本的な計算式を示す。

$$P(F|E) = \sum_a P(F, a|E) \quad (3.3)$$

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．図 3.2 にアライメントの例を示す．

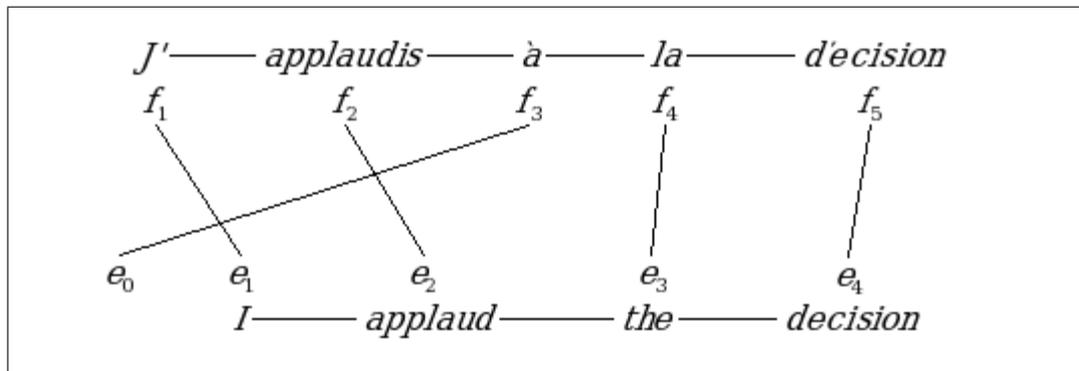


図 3.2 アライメントの例

3.3.2.1 モデル 1

(3.3) 式は以下の式に分解することが出来る． 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．

$$P(F, a|E) = P(m|E) \prod_{j=1}^m P(a_j|a_1^{j-1}, f_1^{j-1}, m, E) P(f_j|a_1^j, f_1^{j-1}, m, E) \quad (3.4)$$

(3.4) 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 ϵ 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。

$$P(F, a | E) = \frac{\epsilon}{(l + 1)^m} \prod_{j=1}^m t(f_j | e_{a_j}) \quad (3.5)$$

$$P(F | E) = \frac{\epsilon}{(l + 1)^m} \sum_{a_1=0}^l \cdots \sum_{a_m=0}^l \prod_{j=1}^m t(f_j | e_{a_j}) \quad (3.6)$$

$$= \frac{\epsilon}{(l + 1)^m} \prod_{j=1}^m \sum_{i=0}^l t(f_j | e_{a_j}) \quad (3.7)$$

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。

$$c(f|e; F, E) = \frac{t(f|e)}{t(f|e_0) + \cdots + t(f|e_l)} \sum_{j=1}^m \delta(f, f_j) \sum_{i=0}^l \delta(e, e_i) \quad (3.8)$$

$\delta(f, f_j)$ 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 λ_e を以下の式により計算する.

$$\lambda_e = \sum_f \sum_{s=1}^S c(f|e; F^{(s)}, E^{(s)}) \quad (3.9)$$

2. (3.9) 式より求めた λ_e を用いて, 翻訳確率 $t(f|e)$ を再計算する.

$$\begin{aligned} t(f|e) &= \lambda_e^{-1} \sum_{s=1}^S c(f|e; F^{(s)}, E^{(s)}) \\ &= \frac{\sum_{s=1}^S c(f|e; F^{(s)}, E^{(s)})}{\sum_f \sum_{s=1}^S c(f|e; F^{(s)}, E^{(s)})} \end{aligned} \quad (3.10)$$

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.

3.3.2.2 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.

$$a(a_j|j, m, l) \equiv P(a_j|a_1^{j-1}, f_1^{j-1}, m, l) \quad (3.11)$$

この関係からモデル 1 における (3.6) 式は, 以下の式に変換できる.

$$P(F|E) = \epsilon \sum_{a_1=0}^l \cdots \sum_{a_m=0}^l \prod_{j=1}^m t(f_j|e_{a_j}) a(a_j|j, m, l) \quad (3.12)$$

$$= \epsilon \prod_{j=1}^m \sum_{i=0}^l t(f_j|e_{a_j}) a(a_j|j, m, l) \quad (3.13)$$

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．

$$c(f|e; F, E) = \frac{t(f|e)}{t(f|e_0) + \cdots + t(f|e_l)} \sum_{j=1}^m \delta(f, f_j) \sum_{i=1}^l \delta(e, e_i) \quad (3.14)$$

$$= \sum_{j=1}^m \sum_{i=0}^l \frac{t(f|e) a(i|j, m, l) \delta(f, f_j) \delta(e, e_i)}{t(f|e_0) a(0|j, m, l) + \cdots + t(f|e_l) a(l|j, m, l)} \quad (3.15)$$

$$c(i|j, m, l; F, E) = \sum_a P(a|E, F) \delta(i, a_j) \quad (3.16)$$

$$= \frac{t(f_j|e_i) a(i|j, m, l)}{t(f_j|e_0) a(0|j, m, l) + \cdots + t(f_j|e_l) a(l|j, m, l)} \quad (3.17)$$

$c(f|e; F, E)$ 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．

3.3.2.3 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 ϕ 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 ϕ_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．

$$P(\phi_0|\phi_1^l, E) = \binom{\phi_1 + \dots + \phi_l}{\phi_0} p_0^{\phi_1 + \dots + \phi_l - \phi_0} p_1^{\phi_0} \quad (3.18)$$

したがって、モデル3は以下の式で求められる。

$$P(F|E) = \sum_{a_1=0}^l \dots \sum_{a_m=0}^l P(F, a|E) \quad (3.19)$$

$$= \sum_{a_1=0}^l \dots \sum_{a_m=0}^l \binom{m - \phi_0}{\phi_0} p_0^{m-2\phi_0} p_1^{\phi_0} \prod_{i=1}^l \phi_i! n(\phi_i|e_i) \\ \times \prod_{j=1}^m t(f_j|e_{a_j}) d(j|a_j, m, l) \quad (3.20)$$

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。

3.3.2.4 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

$$P(\Pi_{[i]1} = j | \pi_1^{[i]-1}, \tau_0^l, \phi_0^l, E) = d_1(j - \odot_{i-1} | \mathcal{A}(e_{[i-1]}), \mathcal{B}(f_j)) \quad (3.21)$$

\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

$$P(\Pi_{[i]k} = j | \pi_{[i]1}^{k-1}, \pi_1^{[i]-1}, \tau_0^l, \phi_0^l, E) = d_{>1}(j - \pi_{[i]k-1} | \mathcal{B}(f_j)) \quad (3.22)$$

$\pi_{[i]k-1}$ 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。

3.3.2.5 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

$$\begin{aligned} P(\Pi_{[i]1} = j | \pi_1^{[i]-1}, \tau_0^l, \phi_o^l, E) \\ = d_1(v_j | \mathcal{B}(f_j), v_{\odot_{i-1}}, v_m - \phi_{[i]} + 1)(1 - \delta(v_j, v_{j-1})) \end{aligned} \quad (3.23)$$

v_j は j 番目までの空白数， A は英語の単語クラス 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

$$\begin{aligned} P(\Pi_{[i]k} = j | \pi_{[i]1}^{k-1}, \pi_1^{[i]-1}, \tau_0^l, \phi_0^l, E) \\ = d_{>1}(v_j - v_{\pi_{[i]k-1}} | \mathcal{B}(f_j), v_m - v_{\pi_{[i]k-1}} - \phi_{[i]} + k)(1 - \delta(v_j, v_{j-1})) \end{aligned} \quad (3.24)$$

3.4 GIZA++

GIZA++[6] とは、統計翻訳で用いることを前提に作られた単語対応のアライメントを行うツールである。IBM モデル1~5を学習し、単語の対応関係の確率値を計算する。

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日英，英日方向の双方向で最尤な単語アライメントを得る．日英方向の単語対応の例を表 3.2，英日方向の単語対応の例を表 3.3 に示す．尚，表中の●は得られた単語アライメントを示す．

表 3.2 日英方向の単語対応

| | I | was | dazzled | by | the | headlights | of | a | car |
|--------|---|-----|---------|----|-----|------------|----|---|-----|
| 自動車 | | | ● | | | | | | |
| の | | | | | | | ● | | |
| ヘッドライト | | | | | | ● | | | |
| に | | | | | | | ● | | |
| 目 | | | | ● | | | | | |
| が | | | | | | | | ● | |
| くらん | | | | | | | | | ● |
| だ | | | | | | | | | ● |

表 3.3 英日方向の単語対応

| | I | was | dazzled | by | the | headlights | of | a | car |
|--------|---|-----|---------|----|-----|------------|----|---|-----|
| 自動車 | ● | | | | | | | | |
| の | | | | | | | ● | | |
| ヘッドライト | | ● | | | | ● | | | |
| に | | | | | | | | | |
| 目 | | | | | ● | | | | |
| が | | | | | | | | ● | |
| くらん | | | | ● | | | | | ● |
| だ | | | ● | | | | | | |

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 3.2 と表 3.3 を用いた，“intersection”での例を表 3.4，“union”での例を表 3.5 に示す．

表 3.4 intersection の例

| | I | was | dazzled | by | the | headlights | of | a | car |
|--------|---|-----|---------|----|-----|------------|----|---|-----|
| 自動車 | | | | | | | | | |
| の | | | | | | | • | | |
| ヘッドライト | | | | | | • | | | |
| に | | | | | | | | | |
| 目 | | | | | | | | | |
| が | | | | | | | | • | |
| くらん | | | | | | | | | • |
| だ | | | | | | | | | |

表 3.5 union の例

| | I | was | dazzled | by | the | headlights | of | a | car |
|--------|---|-----|---------|----|-----|------------|----|---|-----|
| 自動車 | • | | | | | | | | |
| の | | | | | | | • | | |
| ヘッドライト | | • | | | | • | | | |
| に | | | | | | | • | | |
| 目 | | | | • | • | | | | |
| が | | | | | | | | • | |
| くらん | | | | • | | | | | • |
| だ | | | • | | | | | | • |

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 3.6 , “grow-diag” の例を表 3.7 に示す .

表 3.6 grow の例

| | I | was | dazzled | by | the | headlights | of | a | car |
|--------|---|-----|---------|----|-----|------------|----|---|-----|
| 自動車 | | | | | | | | | |
| の | | | | | | | • | | |
| ヘッドライト | | | | | | • | | | |
| に | | | | | | | | | |
| 目 | | | | | | | | | |
| が | | | | | | | | • | |
| くらん | | | | | | | | | • |
| だ | | | | | | | | | • |

表 3.7 grow-diag の例

| | I | was | dazzled | by | the | headlights | of | a | car |
|--------|---|-----|---------|----|-----|------------|----|---|-----|
| 自動車 | | | | | | | | | |
| の | | | | | | | • | | |
| ヘッドライト | | | | | | • | | | |
| に | | | | | | | • | | |
| 目 | | | | | | | | | |
| が | | | | | | | | • | |
| くらん | | | | | | | | | • |
| だ | | | | | | | | | • |

また、この“grow”と“grow-diag”の最後に行う処理として“final”と“final-and”がある。“final”は“union”の単語対応があれば用いる。“final-and”では、“final”に加えて、双方向ともに単語対応がないアライメントも用いる。“grow-diag-final”の例を表3.8，“grow-diag-final-and”の例を表3.9に示す。

表 3.8 grow-diag-final の例

| | I | was | dazzled | by | the | headlights | of | a | car |
|--------|---|-----|---------|----|-----|------------|----|---|-----|
| 自動車 | • | | | | | | | | |
| の | | | | | | | • | | |
| ヘッドライト | | • | | | | • | | | |
| に | | | | | | | • | | |
| 目 | | | | | • | | | | |
| が | | | | | | | | • | |
| くらん | | | | • | | | | | • |
| だ | | | • | | | | | | • |

表 3.9 grow-diag-final-and の例

| | I | was | dazzled | by | the | headlights | of | a | car |
|--------|---|-----|---------|----|-----|------------|----|---|-----|
| 自動車 | • | | | | | | | | |
| の | | | | | | | • | | |
| ヘッドライト | | | | | | • | | | |
| に | | | | | | | • | | |
| 目 | | | | | • | | | | |
| が | | | | | | | | • | |
| くらん | | | | | | | | | • |
| だ | | | | | | | | | • |

そして、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。“grow-diag-final-and” で作成されたフレーズテーブルの例を表 3.10 に示す。

表 3.10 grow-diag-final-and で作成されたフレーズテーブルの例

| | | | | | | |
|--------|------------------|----------|----------|------|-------------|-------|
| 自動車 | I was dazzled by | 1 | 0.333333 | 0.25 | 0.000751315 | 2.718 |
| 自動車 | I was dazzled | 1 | 0.333333 | 0.25 | 0.00826446 | 2.718 |
| 自動車 | I was | 1 | 0.333333 | 0.25 | 0.0909091 | 2.718 |
| 自動車 | I | 0.333333 | 0.333333 | 0.25 | 1 | 2.718 |
| がくらんだ | a car | 1 | 0.166667 | 1 | 0.666667 | 2.718 |
| がくらんだ。 | a car . | 1 | 0.166667 | 1 | 0.666667 | 2.718 |

3.6 言語モデル

3.6.1 言語モデルの概要

言語モデルは単語列の生じる確率を与えるモデルである。日英翻訳では、翻訳モデルで生成された翻訳候補から英語として自然な文を選出する。統計翻訳では一般に、 N -gramモデルを用いる。表 3.11 に言語モデル (2-gram) の例を示す。

表 3.11 言語モデルの例

| | | |
|---------------|---------------|------------|
| | ... | |
| -0.8844527 | 20 percent | 0.03774649 |
| -0.0001015063 | According to | -0.575068 |
| -0.3670357 | Give me | -0.365383 |
| -2.819615 | a baby | -0.2337404 |
| -2.940145 | by bicycle | -0.3376624 |
| -1.589816 | don't believe | -0.0742262 |
| -0.9539242 | exhibition of | -0.3110435 |
| -0.9351615 | higher than | -0.2300005 |
| -2.251086 | other country | 0.1438151 |
| -1.019305 | zero degrees | -0.1247963 |
| | ... | |

表の1行目の例では、左の数値は20の後に *percent* が来る確率 $P(\textit{percent}|\textit{20})$ を常用対数 \log_{10} でとった値 $\log_{10} P(\textit{percent}|\textit{20})$ である。次に、2-gram で表された単語列 *20 percent*、そしてバックオフスムージングで推定された、20の後に *percent* が来る確率 $P(\textit{percent}|\textit{20})$ を常用対数 \log_{10} でとった値 $\log_{10} P(\textit{percent}|\textit{20})$ である。バックオフスムージングに関しては、第 3.6.2 章で述べる。

3.6.2 N -gram モデル

言語モデルの代表的なモデルとして N -gram モデルがある。この N -gram モデルとは 1 次元の単語列 $w_1, w_2, w_3, \dots, w_n = w_1^n$ における i 番目の単語 w_i の生起確率 $P(w_i)$ は w_i の直前の $(N - 1)$ 単語に依存するモデルであり、以下の式で表現される。

$$P(w_1^n) = P(c_1) \times P(c_2|c_1) \times P(c_3|c_1c_2) \cdots \quad (3.25)$$

$$= \prod_{i=1}^n P(w_i|c_1c_2c_3 \cdots c_{i-1}) \quad (3.26)$$

$$\simeq \prod_{i=1}^n P(w_i|w_{i-N+1}, w_{i-N+2}, \dots, w_{i-1}) \quad (3.27)$$

“She is wise .” という英語文 e に対する $N = 2$ とした 2-gram モデルの例を以下に示す。

$$P(e = \text{She is wise .}) \simeq P(\text{She})P(\text{is}|\text{She})P(\text{wise}|\text{is})P(.|\text{wise}) \quad (3.28)$$

3-gram モデルであれば $P(\text{wise}|\text{She is})$, 4-gram モデルであれば $P(.|\text{She is wise})$ となる。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 と特有の呼びかたをする。 $N = 4$ 以上は 4-gram など数値を用いて呼ぶ。一般的に trigram の精度が高いとされている。(3.9) 式を最尤推定により求めると以下の式となる。

$$P(w_n|w_{n-N+1}^{n-1}) = \frac{C(w_{n-N+1}^n)}{C(w_{n-N+1}^{n-1})} \quad (3.29)$$

この (3.29) 式から信頼性の高い値を推定するためには単語列 w_1^n が多く出現している必要がある。しかし実際には多くの単語列は出現数が 0 となることが多いため信頼できる値を推定できない場合が多い。したがって、スムージングを行う。スムージングとは生起確率が 0 となるのを防ぐために、大きい確率値を小さく、小さい確率値を大きくし、確率値の平滑化する手法であり、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では学習データに出現しない N -gram を $(N - 1)$ -gram の値から推定を行う。tri-gram の場合には確率が以下の式で推定される。

$$P(w_n|w_{n-2}^{n-1}) = \begin{cases} \lambda(w_{n-2}^n) \frac{C(w_{n-2}^n)}{C(w_{n-2}^{n-1})} & \text{if } C(w_{n-2}^n) > 0 \\ (1 - \lambda_0(w_{n-2}^{n-1}))\alpha P(w_n|w_{n-1}) & \text{else if } C(w_{n-2}^{n-1}) > 0 \\ P(w_n|w_{n-1}) & \text{otherwise} \end{cases} \quad (3.30)$$

λ はディスカウントと呼ばれる係数であり，出現しない N -gram に対して確率を付与するために， $(N - 1)$ -gram から確率を推定する役割を持つ． α は確率の和を 1 にするための正規化係数である．ディスカウント係数を求める方法は多数存在する．例として，“-ndiscount” や “-kndiscount” や “-ukndiscount” などがある．

N -gram モデルは局所的な情報であり，少ない文法情報しか保持できない．したがって，文法構造が似ている言語間では，部分的な翻訳で十分な翻訳文が生成でき，翻訳精度が高い傾向がある．また，日本語と英語のような文法構造の異なる言語間では特異な文が生成される場合があり翻訳精度が低下する傾向がある．

3.7 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合せから確率が最大となる出力文を探索して翻訳を行う．この探索には莫大な計算量が必要となるが，ビームサーチ法を用いて候補をしばることで計算量を減らす．代表的なデコーダとして “Moses[7]” や “pharaoh[8]” がある．

3.8 パラメータチューニング

デコーダには翻訳確率の重みなどのパラメータが複数存在する．このパラメータを最適化する手法として，一般に Minimum Error Rate Training(MERT)[9] が用いられる．MERT は目的の評価関数（一般に BLEU）を最大にするための翻訳結果を選択するためにパラメータのチューニングを行う．このとき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使用し，各文に対し上位 100 文程度の翻訳候補を出力する．その候補の中で重みを変えることでよりよい翻訳候補が上位にくるように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、あらかじめ用意した正解文と、翻訳システムが出力した文とを比較する手法が一般的である。自動評価法には多くの方法がある。本研究では、BLEU[10]とNIST[11]とMETEOR[12]を用いる。BLEUは語順(4-gram)が正しい場合に高いスコアを出す。NISTではBLEUと同様に語順の正しさで比較を行うが、5-gramを用いる。METEORは単語属性(3人称単数など)が正しい場合に高いスコアを出す。BLEUとMETEORでは0から1までの間で評価され、NISTでは0から までの間で評価される。いずれの評価方法でも、評価方法が高いほど翻訳精度が高いことを表す。尚、本研究では入力文1文に対して正解文1文を用いて評価を行う。

第4章 翻訳システム

4.1 本研究の翻訳システム

統計翻訳では文法情報として、一般に N -gram モデルを用いる。しかし N -gram モデルは局所的な文法情報をモデル化するので、大局的な文法情報がない。したがって、文法構造が大きく異なる言語間では特異な文章が生成される可能性がある。

そこで本研究では、まず統計翻訳を行う前にパターン翻訳を行う。この処理により、日英文パターンが有する大局的な文法情報を用いるので N -gram モデルの局所的な構文問題が解消出来ると考えている。次に、パターン翻訳の出力文に対し統計翻訳を行う。この処理により、局所的な修正を行うことで翻訳精度が向上すると考えている。

また本研究では、日英文パターン辞書を人手で作成するのではなく、プログラムで自動的に作成する。この手法により、日英文パターン辞書の開発時間が短縮出来る。図 4.1 に提案手法の翻訳手順を示す。

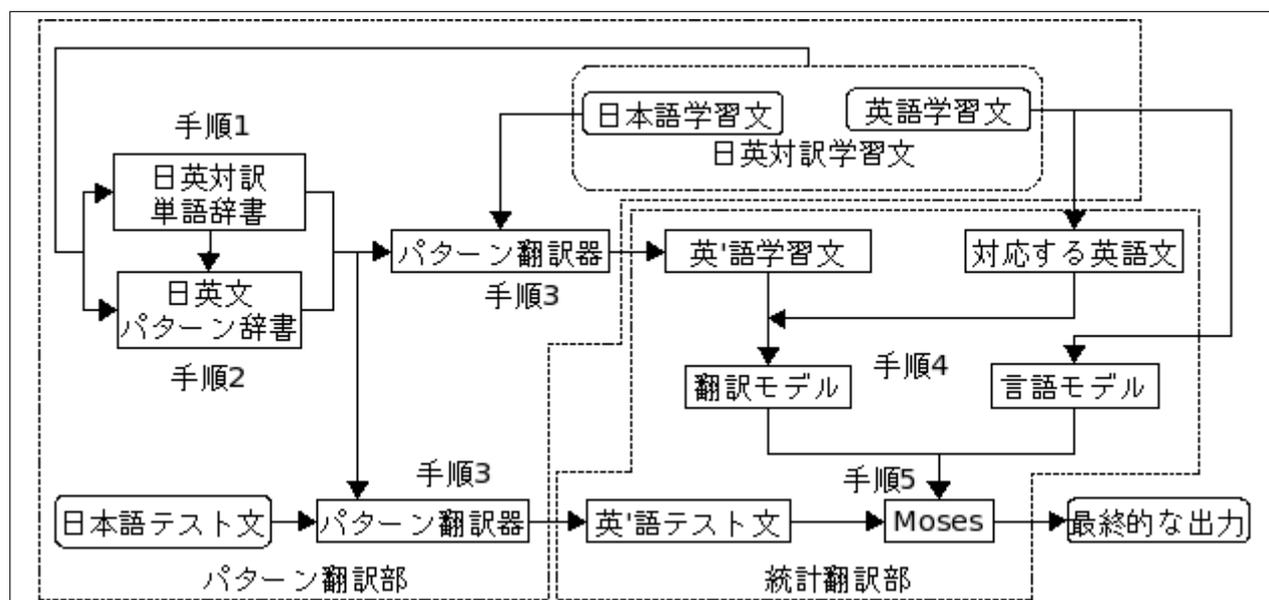


図 4.1 提案手法の翻訳手順

4.2 本研究の翻訳システムの手順

本研究の翻訳システムの手順を以下に示す。

手順 1 日英対訳単語辞書の作成

まず日英対訳学習文から GIZA++ を用いて日英単語辞書と英日単語辞書を作成する。次に両辞書における各単語の確率を掛け合わせる。そして閾値以上の確率を持つ単語を用いて日英対訳単語辞書を作成する。表 4.1 に日英対訳単語辞書の例を示す。

表 4.1 日英対訳単語辞書の例

| 日本語 | 英語 | 確率 |
|-----|--------|-----|
| 彼 | He | 0.4 |
| 彼女 | She | 0.5 |
| ハワイ | Hawaii | 0.4 |

手順 2 日英文パターン辞書の作成

日英対訳単語辞書を用いて日英対訳学習文から日英文パターン辞書を自動的に作成する。図 4.2 に日英文パターン辞書の作成手順を示す。

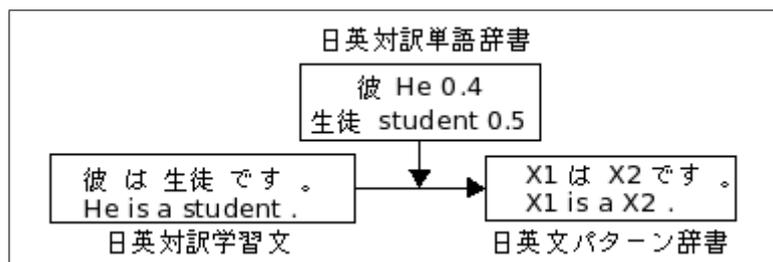


図 4.2 提案手法における日英文パターン辞書の作成手順

日英文パターン辞書は日英対訳単語辞書を参照し、日英対訳学習文中で適合する単語を変数化して作成する。図 4.2 において、日英対訳学習文中にある単語“彼”と“He”が日英対訳単語辞書中にある。したがって、両者を変数“X1”に置換する。同様に“生徒”と“student”も変数“X2”として置換する。以上の処理を日英対訳学習文全てに対して行い、日英文パターン辞書を作成する。

手順3 パターン翻訳

手順1の日英対訳単語辞書と手順2の日英文パターン辞書を用いて、日本語テスト文と日本語学習文に対してパターン翻訳を行う。パターン翻訳を行う際、日本語入力文中の単語と日英対訳単語辞書中の単語で、対応する単語が複数ある場合には、全ての組合せを翻訳候補として出力する。次に各翻訳候補で使用した単語の確率を掛け合わせ、翻訳候補の中で確率が最も高い候補文を選択する。以後、選択した文を英'語出力文とする。また入力文1文に対して複数のパターンに適合する場合、各文パターンにつき1文を出力する。図4.3にパターン翻訳の例を示す。

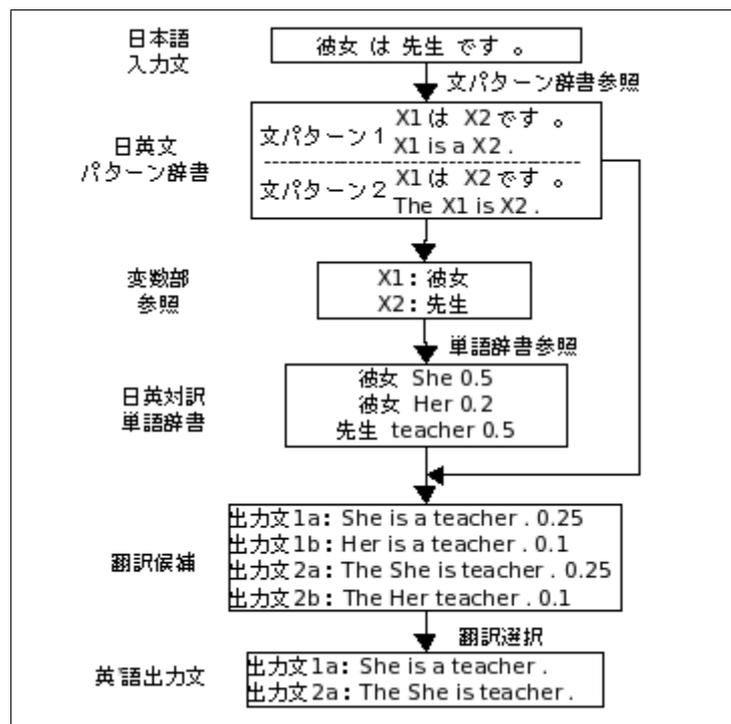


図 4.3 提案手法におけるパターン翻訳の手順

図4.3の例では、日本語入力文に対して2つの日英文パターンが適合する。まず、日本語入力文の“彼女”に対応する単語が日英対訳単語辞書に2つあり、“先生”に対応する単語が1つあるので、文パターン1、文パターン2においてそれぞれ2文が翻訳候補として出力される。次に、翻訳候補が使用した単語の確率を掛け合わせる。そして掛け合わせた確率を使用し、最も確率が高い翻訳候補を選択する。文パターン1では“出力文1a”を、文パターン2では“出力文2a”を選択し、2つの出力文を英'語出力文とする。以上の処理と同様にして、入力文全てに対してパターン

翻訳を行う。

尚、提案手法のパターン翻訳において、次の場合にはパターン翻訳の出力をしない。

- 日本語入力文が日英文パターンに適合しない
- 日英対訳単語辞書を参照する時に適合する単語が日英対訳単語辞書に登録されていない

また日英文パターンに適合しない日本語テスト文に関しては、4.3章で述べるベースラインシステムと同様の翻訳を行う。

手順 4 統計翻訳の翻訳モデルと言語モデルの学習

学習データには、日本語学習文のパターン翻訳で出力された英'語学習文と、その英'語学習文に対応する英語学習文を用いる。この英'語学習文と英語学習文を用いて翻訳モデルを、英語学習文を用いて言語モデルを学習する。

手順 5 統計翻訳における英語文生成

本研究における統計翻訳のデコーダには Moses を用いる。手順 4 で学習した翻訳モデルと言語モデルを用いて、手順 3 の英'語テスト文に英'英統計翻訳を行う。図 4.4 に統計翻訳の例を示す。

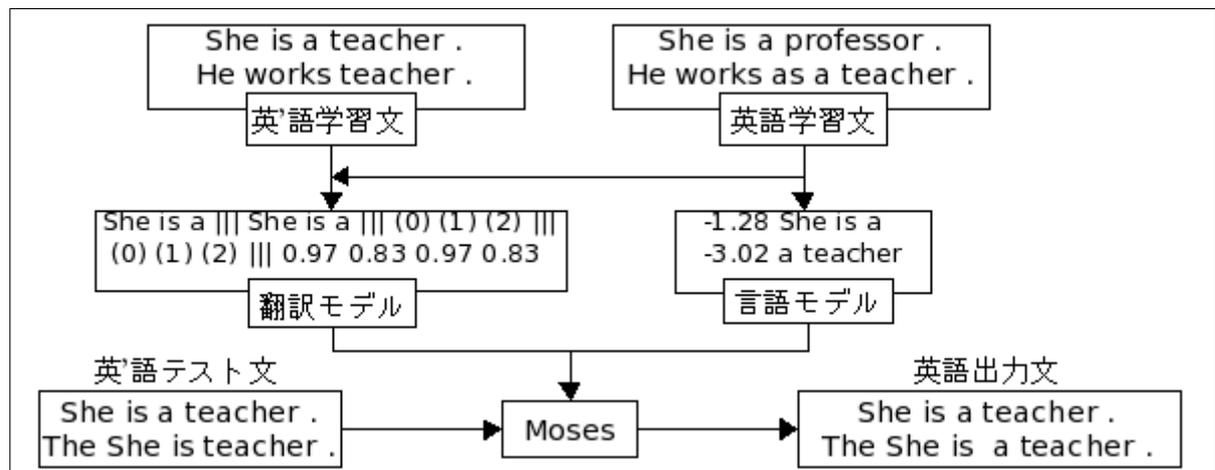


図 4.4 提案手法における統計翻訳の手順

手順 6 翻訳候補の選択

手順 3 で得られた英'語テスト文には、入力文 1 文に対して複数の翻訳候補があ

る．そこで，複数の翻訳候補の中から手順5の統計翻訳の出力時における確率が最も高い文を最終的な出力として1文選択する．

4.3 ベースラインシステム

Moses を用いた日本語テスト文から英語出力文への翻訳を行うシステムをベースラインシステムと呼ぶ．翻訳モデルの学習には日英対訳学習文を，言語モデルの学習には英語学習文を用いる．

第5章 実験環境

5.1 実験データ

実験データには、辞書の例文から抽出した日英対訳文の単文のみを用いる。学習文には100,000文、テスト文には10,000文を用いる。また前処理として、日本語文に対してはchasen[13]を用いて形態素解析を行い、形態素と句読点の間にスペースを入れる。英語文に対してはコンマ、ピリオドの前後にスペースを入れる。表5.1に日英対訳文の例を示す。

表 5.1 日英対訳文の例

| | |
|------|--------------------------------------|
| 日本語文 | もっと右へ寄ってください。 |
| 英語文 | Please move over more to the right. |
| 日本語文 | この傷は至急手当をせねばならない。 |
| 英語文 | The wound requires prompt treatment. |

5.2 翻訳モデルの学習

翻訳モデルの学習には、Mosesに付属している“train-factored-phrase-model.perl”を用いて行う。このプログラムはGIZA++を利用し、IBMモデルの計算を行う。本実験では、フレーズテーブルを作成するヒューリスティクスとして“grow-diag-final-and”を用い、max-phrase-lengthはデフォルト値の20とする。

5.3 言語モデルの学習

言語モデルにはN-gramモデルを用いる。本研究では、SRILM[14]の“ngram-count”を用いて5-gramの言語モデルを学習する。尚、スムージングに“-ndiscount”を用いる。

5.4 デコーダのパラメータ

過去の研究 [15] から，翻訳モデルには日英翻訳確率と英日翻訳確率の共起確率を用いる．したがって，フレーズテーブルの各種の重み “weight-t” を “0.5 0.0 0.5 0.0 0.0” とする．また翻訳時にフレーズの位置の変化を柔軟に対応するために “distortion-weight” を “0.2” とする．尚，本研究ではパラメータチューニングは行わない．

第6章 翻訳実験

6.1 提案手法におけるパターン翻訳部の実験

日本語テスト文 10,000 文と日本語学習文 100,000 文のパターン翻訳には、同じ日英対訳単語辞書と日英文パターン辞書を用いて行う。提案手法で用いる日英対訳単語辞書と日英文パターン辞書の作成手順を以下に示す。

手順 1

GIZA++を用いて日英対訳学習文 100,000 文から日英単語辞書と英日単語辞書を作成する。

手順 2

手順 1 の日英単語辞書と英日単語辞書の各単語の確率を掛け合わせ、閾値以上の確率を持つ単語を用いて日英対訳単語辞書を作成する。

手順 3

日英対訳学習文 100,000 文と手順 2 の日英対訳単語辞書を用いて日英文パターン辞書を作成する。

日英対訳単語辞書を作成する時の閾値について、今回の実験では 0.25, 0.02, 0.01 の 3 つの場合で日英対訳単語辞書を作成する。表 6.1 に各閾値における日英対訳単語辞書の登録単語数を示す。

表 6.1 日英対訳単語辞書の登録単語数

| 閾値 | 登録単語数 |
|------|--------|
| 0.25 | 3,033 |
| 0.02 | 34,587 |
| 0.01 | 53,185 |

6.2 パターン翻訳部の実験結果

実験は以下の2つの実験条件で行う。

実験条件1

閾値が0.25以上の日英対訳単語辞書を用いて日英文パターン辞書を作成し、その日英文パターン辞書と閾値が0.01以上の日英対訳単語辞書を用いてパターン翻訳を行う

実験条件2

閾値が0.01以上の日英対訳単語辞書を用いて日英文パターン辞書を作成し、その日英文パターン辞書と閾値が0.02以上の日英対訳単語辞書を用いてパターン翻訳を行う

日本語学習文での出力文数を表6.2に、日本語テスト文での出力文数を表6.3に示す。尚、表中における“適合した文数”は10,000文中でパターン翻訳が行われた日本語入力文の数を示し、“出力文数”は適合した文数から得られた英’語出力文の数を表している。

表 6.2 日本語学習文での適合した文数と出力文数

| | 適合した文数 | 出力文数 |
|-------|--------|---------|
| 実験条件1 | 47,053 | 51,441 |
| 実験条件2 | 80,988 | 728,819 |

表 6.3 日本語テスト文での適合した文数と出力文数

| | 適合した文数 | 出力文数 |
|-------|--------|--------|
| 実験条件1 | 362 | 566 |
| 実験条件2 | 2,103 | 61,467 |

表6.3の結果から、実験条件1では日本語テスト文の適合した文数が362文であるのに対し、実験条件2では2,103文と大きく増加している。

6.3 提案手法における統計翻訳部の実験

翻訳モデルの学習には表6.2の英’語学習文と英’語学習文に対応する英語学習文を学習する。言語モデルの学習には英語学習文100,000文を用いる。この言語モデルと翻訳

モデルを用いて、表 6.3 の英' 語テスト文に対して英' 英統計翻訳を行う。

6.4 ベースラインシステムにおける実験

ベースラインシステムの翻訳モデルの学習には日英対訳学習文 100,000 文を、言語モデルの学習には英語学習文 100,000 文を用いる。この翻訳モデルと言語モデルを用いて、日本語テスト文 10,000 文に対して日英統計翻訳を行う。

第7章 実験結果

日本語テスト文 10,000 文での結果と，その 10,000 文における表 6.3 の適合した文数での結果を各実験条件ごとに示す．尚，実験条件 1 と実験条件 2 では実験環境が異なるので，ベースラインの BLEU と NIST の値も異なっている．

7.1 自動評価

出力文の評価は自動評価法の BLEU と NIST と METEOR を用いる．

7.1.1 実験条件 1 における評価結果

表 7.1 に 10,000 文での評価結果を示す．また表 7.2 に，表 6.3 における適合した文数 362 文での評価結果を示す．

表 7.1 10,000 文の結果

| | BLEU | NIST | METEOR |
|--------|-------|-------|--------|
| ベースライン | 0.102 | 4.034 | 0.347 |
| 提案手法 | 0.107 | 4.150 | 0.355 |

表 7.2 362 文の結果

| | BLEU | NIST | METEOR |
|--------|-------|-------|--------|
| ベースライン | 0.355 | 5.450 | 0.564 |
| 提案手法 | 0.376 | 5.520 | 0.568 |

表 7.1 において，ベースラインと比較して提案手法では BLEU 値が 0.5% 向上していることから，提案手法の有効性が分かる．また表 7.2 において，ベースラインと比較して提案手法では 2.1% 向上していることから，適合した文での翻訳精度の向上が表 7.1 の結果に影響し，提案手法での翻訳精度が向上した．

7.1.2 実験条件2における評価結果

表 7.3 に 10,000 文での評価結果を示す．また表 7.4 に，表 6.3 における適合した文数 2,103 文での評価結果を示す．

表 7.3 10,000 文の結果

| | BLEU | NIST |
|--------|-------|-------|
| ベースライン | 0.107 | 4.157 |
| 提案手法 | 0.103 | 4.062 |

表 7.4 2,103 文の結果

| | BLEU | NIST |
|--------|-------|-------|
| ベースライン | 0.206 | 5.133 |
| 提案手法 | 0.171 | 4.499 |

表 7.3 において，ベースラインと比較して提案手法では BLEU 値が 0.4% 低下していることから，提案手法が有効でないことがわかる．また表 7.4 において，ベースラインと比べて提案手法では 3.5% 低下していることから，適合した文での翻訳精度の低下が表 7.3 の結果に影響し，提案手法での翻訳精度が低下した．

7.2 人手による評価

表 7.2 と表 7.4 のそれぞれの出力に対してベースラインとの対比較実験を行う。

7.2.1 判断基準

人手による 4 つの判断基準に基づいて評価を行う。評価基準と評価例を以下に示す。尚、未知語はローマ字変換して評価する。

評価 1(>) 提案手法の翻訳結果がベースラインよりも優れている

| 例 1 | |
|--------|--|
| 入力文 | 私は彼女に結婚を申し込んだ。 |
| 正解文 | I proposed to her . |
| ベースライン | I He asked her for her hand . |
| 提案手法 | I made a proposal of marriage to her . |

| 例 2 | |
|--------|---|
| 入力文 | 彼はスライドをスクリーンに映した。 |
| 正解文 | He projected the slides on the screen . |
| ベースライン | He looked at herself in the screen . |
| 提案手法 | He projected a slide The picture on the TV screen . |

評価 2(<) 提案手法の翻訳結果がベースラインよりも劣っている

| 例 1 | |
|--------|--|
| 入力文 | 仕事は山場に入った。 |
| 正解文 | Work has reached the critical point . |
| ベースライン | The work is appear to have entered the final stage . |
| 提案手法 | work went into the labor-management . |

| 例 2 | |
|--------|---------------------------------------|
| 入力文 | 台所の空気が汚れている。 |
| 正解文 | The air in the kitchen is foul . |
| ベースライン | The air is polluted the kitchen . |
| 提案手法 | The kitchen is the dirt off the air . |

評価 3(≈) どちらも似たような文である，またはどちらも入力文で伝えたい情報が理解できない

| 例 1 | |
|--------|------------------------------------|
| 入力文 | 豊作 になり そう だ。 |
| 正解文 | The harvest looks promising . |
| ベースライン | Hopes looks like . |
| 提案手法 | It looks like rejoicing . |
| 例 2 | |
| 入力文 | それは 彼の 彫刻 した もの だ。 |
| 正解文 | It is a production of his chisel . |
| ベースライン | It will be a is engraved . |
| 提案手法 | It is engraved . |

評価 4(=) 提案手法とベースラインの翻訳結果が全く同じである

| 例 1 | |
|--------|---|
| 入力文 | 彼は 故郷 を 恋し がっ て いる 。 |
| 正解文 | He is sick for home . |
| ベースライン | He is homesick . |
| 提案手法 | He is homesick . |
| 例 2 | |
| 入力文 | 患者 は 盲腸炎 の 手術 を 受け た 。 |
| 正解文 | The patient was operated on for appendicitis . |
| ベースライン | The patient had an operation for appendicitis . |
| 提案手法 | The patient had an operation for appendicitis . |

7.2.2 評価結果

各翻訳結果からランダムに抽出した 100 文を対象に評価を行う．評価結果を表 7.5 に示す．

表 7.5 対比較実験の結果

| | 実験条件 1 | 実験条件 2 |
|-------------------|----------|----------|
| 評価 1(>) | 24 / 100 | 26 / 100 |
| 評価 2(<) | 18 / 100 | 24 / 100 |
| 評価 3(\approx) | 18 / 100 | 40 / 100 |
| 評価 4(=) | 40 / 100 | 10 / 100 |

実験条件 1 では提案手法の出力文がベースラインよりも優れている文が多い．しかし英' 語テスト文の多い実験条件 2 では実験条件 1 に比べて提案手法がベースラインよりも劣っている文が多い．

第8章 考察

8.1 日英文パターン辞書における問題点

表 6.3 の結果から実験条件 1 と比べて実験条件 2 では日本語テスト文の適合数が多い。しかし表 7.2 の結果，実験条件 1 では提案手法が有効であるが，表 7.4 の結果から，実験条件 2 では提案手法が翻訳精度の低下を引き起こしている。この原因として，日英文パターン辞書を作成するときに用いた日英対訳単語辞書の閾値が影響していると考えられる。実験条件 1 の日英文パターン辞書の作成には，閾値が 0.25 以上の登録単語数が少ない日英対訳単語辞書を使用している。したがって，日英文パターン中の変数が少なく，文法情報が多く残され，翻訳精度の高い文が出力されたと考えられる。しかし実験条件 2 の日英文パターン辞書では閾値が 0.01 以上の登録単語数が多い単語を使用している。したがって，日本語テスト文の適合数は増加するが日英文パターン中の変数が多く，複雑になり文法情報が損なわれ，翻訳精度の低い文が出力されたと考えられる。表 8.1 に作成した日英文パターンの例を示す。

表 8.1 作成した日英文パターン辞書の例

| 例 1 | |
|-----------------|---|
| 実験条件 1 の日英文パターン | 私は X1 を見に行く。 I go to see a X1 . |
| 実験条件 2 の日英文パターン | X1 X2 X3 を見 X4 X5 。 X1 X5 X4 see X2 X3 . |
| 例 2 | |
| 実験条件 1 の日英文パターン | 電車が X1 を抜ける。 The train goes through the X1 . |
| 実験条件 2 の日英文パターン | X1 X2 X3 X4 抜ける。 X2 X1 goes through X4 X3 . |
| 例 3 | |
| 実験条件 1 の日英文パターン | X1 に X2 を X3 。 X3 X2 on a X1 . |
| 実験条件 2 の日英文パターン | X1 X2 X3 X4 X5 。 X5 X3 X2 X4 X1 . |

例1の実験条件1の英語文パターンでは“I”が主語であり“go to see”が複合動詞であることが分かる。しかし実験条件2では“see”が動詞であると考えられる。したがって、実験条件1では翻訳精度の高い翻訳文が出力されるが実験条件2では翻訳精度の低い文が出力されることが考えられる。

例2の実験条件1の文パターンでは“電車が”となっており主語が電車であると分かり、目的語として1単語があることが分かる。しかし実験条件2の文パターンでは、動詞の“抜ける”と“go through”が対応していることが分かるが、それ以外の部分では全く制限がないため、適合する文数は増えるが本来適合すると考えられる入力文以外の文が適合する可能性があり、翻訳精度が低下する原因となる可能性があると考えられる。

例3では、例2と同様に、適合文に関する問題があるが、実験条件2の文パターン中の全てが変数であるため、例2よりも適合文数が大幅に増加し、最適な翻訳が得られなくなる可能性がより高いと考えられる。

この問題に対し、今後日英文パターンが文法情報を十分に保持し、なおかつ日本語テスト文の適合数を増やすための方法を考える必要がある。

8.2 パターン翻訳の出力における解析

8.2.1 実験条件1における解析

実験条件1での提案手法における出力の成功例を表8.2に示す。尚、表中の“パターン”はパターン翻訳のみの出力文を表す。

表 8.2 実験条件 1:出力の成功例

| 例 1 | |
|----------|---|
| 入力文 | 彼はじっと横になっていた。 |
| ベースライン | He intently at the side . |
| パターン | His lay still . |
| 提案手法 | He lay still . |
| 日文パターン | X1 はじっと横になっていた。 |
| 英文パターン | X1 lay still . |
| 例 2 | |
| 入力文 | 私は彼女に結婚を申し込んだ。 |
| ベースライン | I He asked her for her hand . |
| パターン 1 | my made a proposal of marriage to her . |
| パターン 2 | my proposed marriage to Her . |
| 提案手法 | I made a proposal of marriage to her . |
| 日文パターン 1 | X1 は彼女に結婚を申し込んだ。 |
| 英文パターン 1 | X1 made a proposal of marriage to her . |
| 日文パターン 2 | X1 は X2 に結婚を申し込んだ。 |
| 英文パターン 2 | X1 proposed marriage to X2 . |
| 例 3 | |
| 入力文 | 台所の空気が汚れている。 |
| ベースライン | The air in the kitchen . |
| パターン | The air of this kitchen is foul . |
| 提案手法 | The air of this kitchen is foul . |
| 日文パターン | X1 の空気が汚れている。 |
| 英文パターン | The air of this X1 is foul . |

表 8.2 の例 1 では、ベースラインの出力文には動詞がなく入力文の意味が理解出来ない。パターンのみの出力では適合したパターンから“X1”が“His”に翻訳されていることが分かる。そして提案手法では、“His”を“He”に翻訳され入力文の意味が理解できる文となっている。

例2では、ベースラインの出力文では入力文の意味が理解できないが、パターン翻訳のみの結果のどちらでもおおまかに意味を理解できる。そして提案手法ではさらに“my”が“I”に翻訳され、精度の高い翻訳文が得られている。

例3では、ベースラインの出力文では動詞がなく、“汚れている”に相当する単語が存在しないため入力文の意味が理解できない。一方、パターン翻訳のみでの出力文では入力文を理解できるほど十分に翻訳され、提案手法でも結果が変わっていない。

以上の結果から、入力文が適合したパターンには、文法情報が十分に残っており、パターン翻訳の細かい部分の誤りを統計翻訳で修正できていることが分かる。この結果、パターン翻訳の精度が高く、提案手法の翻訳精度が向上したと考えられる。

次に、提案手法における出力の失敗例を表 8.3 に示す。

表 8.3 実験条件 1:出力の失敗例

| 例 1 | |
|----------|--------------------------------------|
| 入力文 | 目が痛む。 |
| ベースライン | I have a pain in my eyes . |
| パターン 1 | The eyes hurts . |
| パターン 2 | The eyes hurts . |
| パターン 3 | The eyes hurts over . |
| 提案手法 | I am in my joints . |
| 日文パターン 1 | X1 が X2 。 |
| 英文パターン 1 | The X1 X2 . |
| 日文パターン 2 | X1 が X2 。 |
| 英文パターン 2 | The X1 X2 . |
| 日文パターン 3 | X1 が X2 。 |
| 英文パターン 3 | The X1 X2 over . |
| 例 2 | |
| 入力文 | トウモロコシが市場に出た。 |
| ベースライン | The corn on the market . |
| パターン | The productive answered the market . |
| 提案手法 | The productive answered the market . |
| 日文パターン | X1 が X2 に出た。 |
| 英文パターン | The X1 answered the X2 . |
| 例 3 | |
| 入力文 | 彼は電話に出た。 |
| ベースライン | He answered the telephone . |
| パターン | His came out at a telephone . |
| 提案手法 | He went out into the telephone . |
| 日文パターン | X1 は X2 に出た。 |
| 英文パターン | X1 came out at a X2 . |

表 8.3 の例 1 では、“パターン 1” の出力と “パターン 2” の出力が全く同じとなっており、意味が分かる文となっている。さらに適合したパターンを見ると全く同じの文パターンとなっていることがわかる。“パターン 3” では、その出力に英単語 “over” が出力されており、意味にあまり変化がない。しかし、提案手法では、パターン翻訳のみの出力が大きく変化し翻訳精度を低下させていることが分かる。

例2では、パターン翻訳のみの出力と提案手法の出力が全く同じであり、入力文の主語である“トウモロコシ”という単語が“productive”となっており入力文の意味が理解できなくなっている。また適合したパターンの動詞が“answered”となっていることからパターンの適合に問題があると考えられる。

例3では、ベースラインの出力文から入力文の意味が理解できる。一方、パターン翻訳のみの出力文では入力文と全く意味が異なる出力文となっている。そして提案手法では、意味的に問題のある文になってしまう結果となっている。適合したパターンを見ると、“出た”という日本語の動詞に対応する英語が“came out”となっていることから、例2と同様にパターンの適合に問題があると考えられる。

以上の結果から、本来適合すべきパターンが適合していないことがわかる。したがって、パターン翻訳における翻訳精度の改善を行う必要があると考えられる。この問題に対し、品詞タグを用いることで動詞の入る場所には動詞のみを置換するルールを用いるなどの方法が考えられる。また統計翻訳が翻訳低下を引き起こしている場合がある。この問題については今後調査し、対処を行う必要があると考えている。

8.2.2 実験条件2における解析

実験条件2での提案手法における出力の成功例を表8.4を示す。

表 8.4 実験条件2:出力の成功例

| 例 1 | |
|----------|---|
| 入力文 | ワインはブドウで作る。 |
| ベースライン | The wine grapes . |
| パターン | Wine is made in grapes . |
| 提案手法 | Wine is made from grapes . |
| 日文パターン | X1 は X2 X3 作る。 |
| 英文パターン | X1 is made X3 X2 . |
| 例 2 | |
| 入力文 | 彼に不満を抱いている。 |
| ベースライン | I asked him to his dissatisfaction with . |
| パターン 1 | He has His in dissatisfied . |
| パターン 2 | in His is sitting on her dissatisfied . |
| 提案手法 | He has dissatisfied with him . |
| 日文パターン 1 | X1 X2 X3 を抱いている。 |
| 英文パターン 1 | He has X1 X2 X3 . |
| 日文パターン 2 | X1 X2 X3 を抱い X4 いる。 |
| 英文パターン 2 | X2 X1 X4 sitting on her X3 . |
| 例 3 | |
| 入力文 | このセーターは体に合わない。 |
| ベースライン | This sweater does not agree with . |
| パターン 1 | The sweater do not match with in body . |
| パターン 2 | This sweater does not match in The body . |
| 提案手法 | The sweater do not match with the body . |
| 日文パターン 1 | この X1 X2 X3 X4 合わ X5 。 |
| 英文パターン 1 | X2 X1 do X5 match with X4 X3 . |
| 日文パターン 2 | X1 X2 X3 X4 X5 合わ X6 。 |
| 英文パターン 2 | X1 X2 does X6 match X5 X3 X4 . |

表8.4の例1では、適合した文パターンに十分な文法情報が残されていることがわかる。パターン翻訳のみの出力では“in”となっている部分が、提案手法では“from”となり、精度の高い文が得られている。ベースラインでは動詞がないので、意味が理解できない文となっている。

例2では、ベースラインの出力では“asked”が存在することで文全体の意味が入力文と大きく異なる文となっている。パターン翻訳のみの出力では語順に問題があることがわかる。提案手法ではその語順の問題を修正し、意味が理解できる文となっている。

例3では、ベースラインの出力の動詞には“agree”が使用されており、英語文として意味的に問題がある。一方、パターン翻訳のみの出力では、“パターン1”の出力と“パターン2”の出力の差はほとんどなく、どちらの出力も入力文をある程度理解することが出来る。そして提案手法では、文の途中の大文字などの部分を修正し、精度の高い文となっている。

以上の結果から、文パターンの変数が少し増加し、パターン翻訳の精度が低下しても、統計翻訳の局所的な修正を行うことで翻訳精度が高い文が得られることが分かる。

次に、提案手法における出力の失敗例を表 8.5 を示す。

表 8.5 実験条件 2:出力の失敗例

| 例 1 | |
|----------|--|
| 入力文 | 母は悲しみに沈んでいます。 |
| ベースライン | My mother has been submerged in grief . |
| パターン 1 | My Sorrow The Mother are in in sank . |
| パターン 2 | Mother sank in Sorrow city . |
| パターン 3 | in are Sorrow sank up The the Mother . |
| パターン 4 | The Mother in sank at in front Sorrow . |
| 提案手法 | My mother fell in the sadness in the sea . |
| 日文パターン 1 | X1 X2 X3 X4 X5 X6 います。 |
| 英文パターン 1 | My X3 X2 X1 are X6 X4 X5 . |
| 日文パターン 2 | X1 は X2 X3 X4 でいます。 |
| 英文パターン 2 | X1 X4 X3 X2 city . |
| 日文パターン 3 | X1 X2 X3 X4 X5 でいます。 |
| 英文パターン 3 | X4 are X3 X5 up X2 the X1 . |
| 日文パターン 4 | X1 X2 X3 X4 X5 X6 います。 |
| 英文パターン 4 | X2 X1 X6 X5 at X4 front X3 . |
| 例 2 | |
| 入力文 | 彼女は家庭の事情で高校を中退した。 |
| ベースライン | She 中退 a high school because of family circumstances . |
| パターン | The Her midyear family a of high in his circumstances . |
| 提案手法 | She midyear of the high in his family circumstances of that time . |
| 日文パターン | X1 X2 X3 X4 X5 X6 X7 X8 X9 した。 |
| 英文パターン | X2 X1 X9 X3 X8 X4 X7 X6 his X5 . |
| 例 3 | |
| 入力文 | キャンプファイヤーの火が赤々と燃えている。 |
| ベースライン | The fire in the camp fire is burning brightly . |
| パターン | and of fire The Fields is burning juddered red refugee juddered . |
| 提案手法 | The fire is burning red with the Falabella camp fire . |
| 日文パターン | X1 X2 X3 X4 X5 X6 X7 X8 X9 X10 X11 いる。 |
| 英文パターン | X9 X4 X5 X6 X8 X11 X10 X3 X7 X1 X2 . |

表 8.3 の例 1 では、ベースラインの出力文からは入力文の意味をおおまかに理解できる。一方、パターン翻訳のみの出力では、不要であると考えられる単語が存在するため、

入力文の意味が理解できない．提案手法では，修正されており意味を出来る文となっ
ているが，“in the sea”という不要なフレーズが存在している．

例2では，ベースラインの出力文には未知語が存在しているが，文全体として入力文
の意味が理解できる文となっている．しかしパターン翻訳のみの出力文では動詞が存在
せず，提案手法の出力文でも動詞が存在しないので，精度の低い文となっている．

例3では，ベースラインの出力では入力文の意味が理解できる文となっている．パター
ン翻訳時に適合した英文パターンは全て変数化されており，出力文では全く意味が理解
できない文となっている．提案手法では大幅に修正され入力文の意味を少し理解できる
文となっている．しかし，“Falabella”という不要な単語が存在している．

以上の結果から，実験条件2では，文パターン中の変数が増加し，文パターンが持つ
文法情報が大幅に損なわれていることがわかる．しかし，第7.1.2章の結果と，第8.1章
の結果から，実験条件1と比較して入力文の適合数は大きく増加している．したがって，
文パターン作成時に使用する日英対訳単語辞書の閾値を考慮することで，適合数を増加
し，文法情報を少しでも多く残すことが出来ると考えられる．

第9章 おわりに

本研究では、日本語-英語間における大きく異なる文法構造に着目し、まずパターン翻訳を行うことで文法構造を英語に近づけ、次に統計翻訳を行うことで文法構造に対し、局所的な修正を行う手法を提案した。実験の結果、文パターン辞書における変数が少ない場合は文パターンに文法情報が多く残されるため翻訳精度の高い文が得られ、提案手法の有効性が示された。この場合での自動評価の結果、BLEU 値で 0.5%、METEOR 値では 0.08%、NIST 値で 0.116 の翻訳精度の向上が見られた。しかし変数の多い場合には文パターンの文法情報が損なわれ、この場合での自動評価の結果、BLEU 値で 0.4%、NIST 値で 0.095 の低下が見られた。

文パターンにおける文法情報の減少に対し、今後日英対訳単語辞書の閾値を考慮する必要がある。またパターン翻訳における翻訳精度を向上するために、パターン翻訳時に品詞を考慮したルールを用いるなどの処理を行い、翻訳精度の高い翻訳候補を得る必要がある。

謝辞

最後に，本研究を遂行するにあたり，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C研究室の池原悟教授，菅原一孔教授，村上仁一准教授，徳久雅人助教授そして計算機工学C研究室の方々に深く感謝致します．また，参考にさせていただいた文献の著者の方々に対して感謝致します．

参考文献

- [1] 池原悟, 宮崎正弘, 白井諭, 横尾昭男, 中岩浩巳, 小倉健太郎, 大山芳史, 林良彦, “日本語語彙大系”, 岩波書店, 1997
- [2] Richard Zens, Franz Josef Och, Hermann Ney “Phrase-based Statistical Machine Translation”, KI 2002, pp35-56, 2002
- [3] 岡崎弘樹, 村上仁一, 徳久雅人, 池原悟, “日本語文法構造変換による日英統計翻訳”, 言語処理学会第 15 回年次大会, pp240-243, 2009
- [4] 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
- [5] Peter F. Brown, Stephen A. Della Pietra, Vincent j. Della Pietra, Robert L. Mercer, “The Mathematics of Statistical Machine Translation: Parameter Estimation”, Association for Computational Linguistics 1993, pp263-311, 1993
- [6] GIZA++, Training of statistical translation models
<http://www.fjoch.com/GIZA++.html>
- [7] Moses, A Beam Search Decoder for Factored Phrase-Based Statistical Machine Translation Models
<http://www.statmt.org/moses/>
- [8] Pharaoh, A Beam Search Decoder for Phrase-Based Statistical Machine Translation Models
<http://www.isi.edu/licensed-sw/pharaoh/>
- [9] Franz Josef och, “Minimum Error Rate Training in Statistical Machine Translation”, Association for Computational Linguistics 2003, pp160-167, 2003

- [10] BLEU, NIST Open MT Scoring
<http://www.itl.nist.gov/iad/894.01/tests/mt/2008/scoring.html>
- [11] NIST, Automatic Evaluation of Machine Translation Quality Using N-gram Co-Occurrence Statistics
<http://www.itl.nist.gov/iad/mig/tests/mt/2008/scoring.html>
- [12] METEOR, The METEOR Automatic Machine Translation Evaluation System
<http://www-2.cs.cmu.edu/~alavie/METEOR/>
- [13] chasen, 日本語形態素解析器
<http://chasen-legacy.sourceforge.jp/>
- [14] SRILM, The SRI Language Modeling Toolkit
<http://www.speech.sri.com/projects/srilm/>
- [15] 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
- [16] 西村拓哉, 村上仁一, 徳久雅人, 池原悟, "文単位の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 (発表予定)