

概要

近年、機械翻訳において統計翻訳が注目され、盛んに研究が行われている。本論文では統計翻訳に関する3つの研究を扱う。以下にそれぞれの概要について示す。

- **パターン翻訳を用いた学習データ増加手法の検討**

統計翻訳の翻訳品質は、学習データとして用いる対訳データの量に依存する。しかし、人手で翻訳した高品質な対訳データの収集には量的な限界があり、作成コストも大きい。

この問題に対して、本研究では大量の単言語データをパターン翻訳によって翻訳し、既存の対訳学習データに付与した。実験の結果、自動評価および人手評価における提案手法の有効性が認められた。

- **句に基づく統計翻訳における未知語処理の1手法**

統計翻訳において、翻訳されない単語は未知語として出力される。この未知語を減少させるため、様々な試みが行われている。代表的な試みとして、対訳学習データに対訳辞書データを付与する手法がある。しかし、この手法では、単語辞書などの対訳学習データ以外のリソースが必要となる。

そこで本研究では、対訳データのみを利用して、未知語を削減する方法を提案する。実験の結果、単文を用いた実験において、出力文全体の自動評価値がベースラインと比較してわずかに向上した。さらに、提案手法において未知語の改善に大きな効果が認められた。

- **統語的タグを用いた統計的階層句機械翻訳**

従来の統計翻訳では、翻訳課程において統語的情報を用いていない。そのため、日本語-英語といった言語構造が大きく異なる言語間の翻訳において高い翻訳精度が得られない場合が多い。

そこで本研究では、日英統計的階層句機械翻訳(階層型翻訳)において統語的情報を追加する手法を提案し、翻訳精度の向上を目指した。

実験の結果、提案手法において有意な効果は認められなかった。原因として、日本語文へのタグ付与のみでは、統語的情報として不十分であることなどが挙げられる。今後の展開として、生成されたルールのフィルタリングを行う手法などが考えられる。

目次

第1章 序論	1
第2章 日英統計翻訳システム	3
2.1 概要	3
2.2 翻訳モデル	4
2.2.1 IBM 翻訳モデル	4
2.2.1.1 モデル1	5
2.2.1.2 モデル2	7
2.2.1.3 モデル3	8
2.2.1.4 モデル4	9
2.2.1.5 モデル5	9
2.2.2 GIZA++	10
2.2.3 フレーズテーブル作成法	11
2.3 言語モデル	15
2.4 デコーダ	17
2.5 パラメータチューニング	17
第3章 評価手法	19
3.1 自動評価	19
3.1.1 BLEU	19
3.1.2 METEOR	20
3.1.3 RIBES	20
3.2 人手評価	21
第4章 パターン翻訳を用いた学習データ増加手法の検討	22
4.1 背景および目的	22
4.2 パターン翻訳の概要	23

4.2.1	文パターン辞書	24
4.3	提案手法	25
4.3.1	提案手法の手順	25
4.4	実験環境	26
4.4.1	翻訳モデル	26
4.4.2	言語モデル	26
4.4.3	デコーダのパラメータ	26
4.4.4	実験データ	26
4.4.5	評価手法	28
4.5	実験結果	28
4.5.1	自動評価	28
4.5.2	人手評価	29
4.6	分析	33
4.6.1	先行研究との比較	33
4.6.2	パターン翻訳における翻訳文のランク付け	34
4.6.3	文長差を用いた翻訳文の選択	37
4.6.3.1	単文データセットにおける文長差の調査	37
4.6.3.2	文長差によるフィルタリング	37
4.6.4	追加の人手評価	38
4.6.5	市販のルールベース翻訳器を用いた実験	39
4.6.6	使用されたフレーズテーブルの分析	40
4.7	考察	42
4.8	まとめ	42
第5章	句に基づく統計翻訳における未知語処理の1手法	43
5.1	背景および目的	43
5.2	未知語	44
5.3	提案手法	45
5.4	実験環境	46
5.4.1	実験データ	46
5.5	評価手法	46
5.6	実験結果	47

5.6.1	提案手法の効果	47
5.6.2	未知語	48
5.6.2.1	未知語の改善	48
5.6.2.2	未知語改善による翻訳品質への影響	49
5.6.3	システム全体に対する提案手法の影響	51
5.6.3.1	全出力文中の未知語数	51
5.6.3.2	自動評価	51
5.6.4	人手評価結果	52
5.6.4.1	人手評価結果	52
5.6.4.2	対比較評価出力例	53
5.7	解析	55
5.7.1	intersection との比較	55
5.8	考察	57
5.9	まとめ	59
第 6 章	統語的タグを用いた統計的階層句機械翻訳	60
6.1	背景と目的	60
6.2	統計的階層句機械翻訳	61
6.2.1	ルールテーブル	61
6.2.2	glue-grammar	61
6.3	提案手法	62
6.3.1	提案手法の概略	62
6.3.2	タグ付与の手順	62
6.3.3	提案手法の glue-grammar	63
6.4	実験環境	64
6.4.1	実験データ	64
6.5	実験結果	65
6.5.1	自動評価	65
6.5.2	対比較調査	66
6.6	考察	68
6.6.1	生成されたルールテーブルの考察	68
6.6.2	デコーダが考慮するフレーズ長の問題	68

6.6.3	動詞句タグの有効性	69
6.6.4	階層型翻訳におけるタグ付与の効果	69
6.6.5	統語的情報の不足	70
6.7	まとめ	70
第7章	結論	71

目 次

2.1	日英統計翻訳の枠組	3
2.2	日英方向の単語対応	11
2.3	英日方向の単語対応	11
2.4	日英方向の単語対応	11
2.5	英日方向の単語対応	11
2.6	intersection の例	12
2.7	union の例	12
2.8	grow の例	13
2.9	grow-diag の例	13
2.10	grow-diag-final の例	14
2.11	grow-diag-final-and の例	14
2.12	デコーダの動作例	17
4.1	パターン翻訳の枠組	23
4.2	提案手法の概要	25
6.1	階層型翻訳の枠組	61

表 目 次

2.1	フレーズテーブルの例	4
2.2	N -gram の例	15
4.1	文パターン辞書の例	24
4.2	単語辞書の例	24
4.3	単文コーパスの例	27
4.4	重文複文コーパスの例	27
4.5	データの内訳	27
4.6	自動評価結果 (チューニングなし)	28
4.7	自動評価結果 (チューニングあり)	28
4.8	人手評価結果	29
4.9	提案手法○の出力例 (チューニングなし単文)	30
4.10	提案手法×の出力例 (チューニングなし単文)	31
4.11	提案手法○の出力例 (チューニングあり単文)	31
4.12	提案手法×の出力例 (チューニングあり単文)	32
4.13	提案手法○の出力例 (チューニングなし重文複文)	32
4.14	提案手法×の出力例 (チューニングなし重文複文)	32
4.15	提案手法○の出力例 (チューニングあり重文複文)	32
4.16	提案手法×の出力例 (チューニングあり重文複文)	33
4.17	自動評価結果 (チューニングなし)	33
4.18	自動評価結果 (チューニングあり)	34
4.19	人手評価結果	34
4.20	自動評価結果 (チューニングなし)	35
4.21	自動評価結果 (チューニングあり)	36
4.22	文長差が大きいパターン翻訳出力文の例	37
4.23	文長差の統計	38

4.24	自動評価結果	38
4.25	追加人手評価結果	39
4.26	自動評価結果	39
4.27	文例 1	40
4.28	文例 1 で使用されたフレーズテーブル	40
4.29	文例 2	40
4.30	文例 2 で使用されたフレーズテーブル	41
4.31	文例 3	41
4.32	文例 3 で使用されたフレーズテーブル	41
5.1	未知語の出力例	44
5.2	ベースラインのアライメント	44
5.3	フレーズテーブル (grow-dial-final-and) の一部	44
5.4	フレーズテーブル (intersection) の一部	45
5.5	データの内訳	46
5.6	提案手法の出力例	47
5.7	提案手法のアライメント	47
5.8	未知語の改善文数	48
5.9	未知語改善の例	48
5.10	翻訳品質の改善文数	49
5.11	未知語改善による翻訳品質向上の例	50
5.12	未知語数	51
5.13	自動評価結果	51
5.14	人手評価結果	52
5.15	提案手法○ (単文)	53
5.16	提案手法× (単文)	53
5.17	提案手法○ (重文複文)	54
5.18	提案手法× (重文複文)	54
5.19	自動評価結果	55
5.20	人手評価結果	55
5.21	提案手法○ (単文)	56
5.22	提案手法○ (重文複文)	57

5.23	翻訳品質が改善しない例	58
6.1	ルールテーブルの例	61
6.2	glue-grammar	62
6.3	タグ付与の例	63
6.4	提案手法における glue-grammar	63
6.5	データの内訳	64
6.6	タグ付与例	64
6.7	自動評価結果	65
6.8	人手評価結果	66
6.9	単文における出力文例	67
6.10	重文複文における出力文例	67
6.11	生成されたルールの例	68
6.12	MaxPhraseSpan の増加による翻訳精度の変化	68
6.13	動詞句タグの効果	69
6.14	句に基づく統計翻訳におけるタグ付与の有効性	69

第1章 序論

機械翻訳において、人手で翻訳規則を定義し、翻訳を行うルールベース翻訳が一般的であった。しかし、人手で翻訳規則を定義するには、莫大なコストがかかる。また、言語毎に文法規則が異なるため、多言語への拡張が困難であった。そこで近年では、統計翻訳が主流となっている。統計翻訳では、対訳学習データから自動的に翻訳規則を獲得し、翻訳を行う。統計翻訳の利点として、ルールベース翻訳と比較において低コストであること、また、多言語への拡張が容易であることなどが挙げられる。本論文では統計翻訳に関する3つの研究を行い、その効果を示す。

それぞれの概要を以下に示す。

1) パターン翻訳を用いた学習データ増加手法の検討

統計翻訳の翻訳品質は、学習データとして用いる対訳データの量に依存する [1]。しかし、人手で翻訳した高品質な対訳データの収集には量的な限界があり、作成コストも大きい。特に、日英間における統計翻訳では、対訳データの不足が大きな問題となっている。

そこで本研究では、先行研究における単言語データの翻訳に、パターン翻訳を用いる手法を提案する。パターンに適合した場合の翻訳文が高品質であるというパターン翻訳の特徴を活用し、システムにおける翻訳精度の向上を目指す。

2) 句に基づく統計翻訳における未知語処理の1手法

統計翻訳において、翻訳されない単語は未知語として出力される。未知語を減少させるため、様々な試みが行われている。代表的な手法として、単語辞書などの対訳辞書データをパラレルコーパスに追加する手法がある [2]。しかし、この手法では、単語辞書などのパラレルコーパス以外のリソースが必要となる。

そこで本研究では、パラレルコーパスのみを利用して、未知語を削減する方法を提案する。統計翻訳では、単語列から単語列の翻訳確率をフレーズテーブルと呼ばれる表で管理する。フレーズテーブルは、単語対応からヒューリスティックスを用

いて作成される。このフレーズテーブル作成の際に用いるヒューリスティックスにはいくつかの種類が存在するが、あるヒューリスティックスで作成されたフレーズテーブルに存在しないフレーズが、別のヒューリスティックスで作成したフレーズテーブルに存在する場合がある。したがって、異なるヒューリスティックスを用いたフレーズテーブルを併用することで、未知語が削減できる可能性がある。

本研究では、ヒューリスティックスとして“grow-diag-final-and”と“intersection”を併用することで、未知語の削減と翻訳精度の改善を目指す。

3) 統語的タグを用いた統計的階層句機械翻訳

従来の統計翻訳では、翻訳過程において統語的情報を用いていない。そのため、日本語と英語のような言語構造が大きく異なる言語間の翻訳において高い翻訳精度が得られない場合が多い。この問題に対して、様々な研究が行われている。

大西らは、文書レベルの文脈情報を用いてフレーズの並び替えを制限する手法を提案し、有意な成果を得た [3]。また、Zollmann らは統語的情報を用いた機械翻訳システムを提案し、公開している [4]。しかし、統語的情報を用いた機械翻訳の問題点として、統語ラベルの爆発的増加による、解析および翻訳時間の増加が挙げられる。

そこで本研究では、日英統計的階層句機械翻訳において、‘浅い’統語的情報を用いる手法を提案する。浅い統語的情報として、日本語文の名詞句と動詞句にタグを付与する。タグを付与した日本語文を用いて学習することで、統語的情報を含む文法規則が生成され、翻訳精度が向上すると考えられる。

本論文の構成を以下に示す。第2章で従来の日英統計翻訳システムの概要について述べる。第3章でシステムの評価手法について説明する。そして、第4章、第5章、第6章で3つの研究それぞれについて述べる。最後に、第7章で3つの提案手法についてまとめる。

第2章 日英統計翻訳システム

2.1 概要

統計翻訳において、「単語に基づく統計翻訳」と、「句に基づく統計翻訳」がある。初期の統計翻訳は、単語に基づく統計翻訳であった。しかし、近年提案された句に基づく統計翻訳は、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、単語に基づく統計翻訳よりも優れている。よって、現在は句に基づく統計翻訳が主流となっている。そのため、本研究で扱う統計翻訳システムにおいても、句に基づく統計翻訳を用いる。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。日英統計翻訳の枠組みを図2.1に示す。図2.1において、 J は日本語コーパス、 E は英語コーパスを表している。

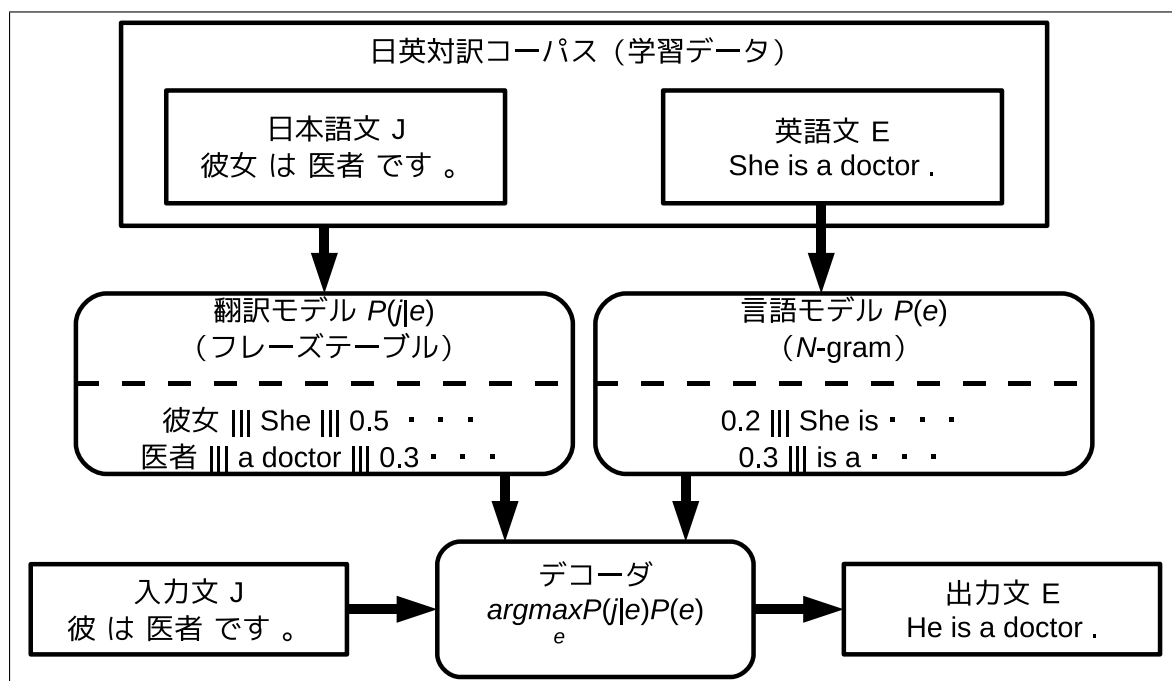


図 2.1: 日英統計翻訳の枠組

日英統計翻訳では、日本語の1文 j を入力文とした場合に、翻訳モデル $P(j|e)$ と言語モデル $P(e)$ の全ての組み合わせから、確率が最大となる英語の1文 \hat{e} を出力文とする。 \hat{e} を探索するシステムをデコーダと呼ぶ。以下に基本的なモデルを示す。

$$\hat{e} = \arg \max_e P(e|j) \quad (2.1)$$

$$\simeq \arg \max_e P(j|e)P(e) \quad (2.2)$$

2.2 翻訳モデル

翻訳モデルは、日本語から英語の単語列へ、確率的に翻訳を行うためのモデルである。統計翻訳において、句に基づく翻訳モデルとして、一般的に、フレーズテーブルが用いられている。フレーズテーブルは以下の手順によって作成される。

手順1 後述する IBM モデルを用いて、単語の対応を得る

手順2 ヒューリスティックなルールを用いて句に基づく対応を得る

手順3 手順2で求めた句対応から、フレーズテーブルを作成する

詳しい作成手順については、後述する。また、表 2.1 にフレーズテーブルの例を示す。

表 2.1: フレーズテーブルの例

日フレーズ	英フレーズ	$P(j e)$	$\prod P(j e)$	$P(e j)$	$\prod P(e j)$
季節	the season	0.3	3.6	0.1	1.8
新緑の季節	the season of green leaves	0.3	3.6	0.1	2.6
雨の後	after the rain	0.3	1.2	0.1	1.8

左から順に、日本語フレーズ、英語フレーズ、日英方向の翻訳確率 $P(j|e)$ 、日英方向の単語の翻訳確率の積、英日方向の翻訳確率 $P(e|j)$ 、英日方向の単語の翻訳確率の積である。

2.2.1 IBM 翻訳モデル

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 [5] がある。この翻訳モデルは、提案者の Brown らが全員 IBM 社員で

あったため、IBM 翻訳モデルと呼ばれている。IBM 翻訳モデルは、原言語側から目的言語側への翻訳確率の近似方法が異なる、モデル 1 からモデル 5 までの 5 つのモデルで構成されている。各モデルの概要を以下に示す。

モデル 1 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使用

モデル 2 モデル 1 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) を使用

モデル 3 モデル 2 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使用

モデル 4 モデル 3 における permutation 確率を改良 (モデル 2 の絶対位置に対して、相対位置)

モデル 5 モデル 4 における permutation 確率を更に改良

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、日英翻訳を前提に説明する。

原言語の日本語文を J 、目的言語の英語文を E として定義する。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 モデルの基本的な計算式を示す。

$$P(J|E) = \sum_a P(J, a|E) \quad (2.3)$$

ここで、アライメント a は、日本語単語 j と英単語 e の対応を意味している。IBM 翻訳モデルにおいて、各日単語に対応する英単語は 1 つであるのに対して、各英単語に対応する日単語は 0 から n 個あると仮定する。また、日単語と適切な英単語が対応しない場合、英語文の先頭に e_0 という空単語があると仮定し、日単語と対応させる。

2.2.1.1 モデル 1

式 2.3 は以下の式に置き換えられる。

$$P(J, a|E) = P(m|E) \prod_{k=1}^m P(a_k | a_1^{k-1}, j_1^{k-1}, m, E) P(j_k | a_1^k, j_1^{k-1}, m, E) \quad (2.4)$$

m は日本語文の文長を示す．また， a_1^{k-1} は日本語文の 1 単語目から $k-1$ 単語目までのアライメントである．そして j_1^{k-1} は日本語文の 1 番目から $k-1$ 番目までの単語を示す．ここで，モデル 1 では以下を仮定している．

- 日本語文の長さの確率 ϵ は， m と E に依存しない
 $\epsilon \equiv P(m|E)$
- アライメントの確率は英語文の長さ l にのみ依存する
 $P(a_k|a_1^{k-1}, j_1^{k-1}, m, E) \equiv (l+1)^{-1}$
- 日本語の翻訳確率 $t(j_k|e_{a_k})$ は，日単語に対応する英単語にのみ依存する
 $P(j_k|a_1^k, j_1^{k-1}, m, E) \equiv t(j_k|e_{a_k})$

以上の仮定を用いて， $P(J, a|E)$ と $P(J, E)$ は以下のように表される．

$$P(J, a|E) = \frac{\epsilon}{(l+1)^m} \prod_{k=1}^m t(j_k|e_{a_k}) \quad (2.5)$$

$$P(J|E) = \frac{\epsilon}{(l+1)^m} \sum_{a_1=0}^l \cdots \sum_{a_m=0}^l \prod_{k=1}^m t(j_k|e_{a_k}) \quad (2.6)$$

$$= \frac{\epsilon}{(l+1)^m} \prod_{k=1}^m \sum_{i=0}^l t(j_k|e_i) \quad (2.7)$$

モデル 1 において，翻訳確率 $t(j|e)$ の初期値が 0 でない場合には，EM アルゴリズム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．ここで，日英の対訳文の文数を S 文とする．

手順 1 $t(j|e)$ に初期値を設定する

手順 2 日本語と英語の対訳文 $(J^{(s)}, E^{(s)}) (1 \leq s \leq S)$ において，日単語 j と英単語 e が対応付けられる回数の期待値を求める．ここで $\delta(j, j_k)$ は日本語文 J において日単語 j が出現する回数を表す．そして $\delta(e, e_i)$ は英語文 E において英単語 e が出現する回数を表す．

$$c(j|e; J, E) = \frac{t(j|e)}{t(j|e_0) + \cdots + t(j|e_l)} \sum_{k=1}^m \delta(j, j_k) \sum_{i=0}^l \delta(e, e_i) \quad (2.8)$$

手順 3 英語文 $E^{(s)}$ において，1 回以上出現する英単語 e に対して，翻訳確率 $t(j|e)$ を計算する．

- 定数 λ_e を以下の式で計算する

$$\lambda_e = \sum_j \sum_{s=1}^S c(j|e; J^{(s)}, E^{(s)}) \quad (2.9)$$

- 上式で求めた定数 λ_e を用いて $t(j|e)$ を以下の式で再計算する

$$t(j|e) = \lambda_e^{-1} \sum_{s=1}^S c(j|e; J^{(s)}, E^{(s)}) \quad (2.10)$$

$$= \frac{\sum_{s=1}^S c(j|e; J^{(s)}, E^{(s)})}{\sum_k \sum_{s=1}^S c(j|e; J^{(s)}, E^{(s)})} \quad (2.11)$$

手順4 $t(j|e)$ が収束するまで、手順2と手順3を繰り返す

2.2.1.2 モデル2

モデル1において、アライメントの確率は英語文の長さ l にのみ依存する。そこでモデル2では、英語文の長さ l に加え、 k 単語目のアライメント a_k 、日本語文の長さ m に依存するとし、以下の式で表す。

$$a(a_k|k, m, l) \equiv P(a_k|a_1^{k-1}, j_1^{k-1}, m, l) \quad (2.12)$$

よって、モデル1の式2.6は以下のように置き換えられる。

$$P(J|E) = \epsilon \sum_{a_1=0}^l \cdots \sum_{a_m=0}^l \prod_{k=1}^m t(j_k|e_{a_k}) a(a_k|k, m, l) \quad (2.13)$$

$$= \epsilon \prod_{k=1}^m \sum_{i=0}^l t(j_k|e_i) a(i|k, m, l) \quad (2.14)$$

モデル2において、対訳文中の英単語 e と日単語 j が対応付けされる回数の期待値である $c(j|e; J^{(s)}, E^{(s)})$ と、日単語の位置 j と英単語の位置 i が対応付けられる回数の期待値 $c(i|k, m, l; J^{(s)}, E^{(s)})$ が存在する。以下に、期待値 $c(j|e; J^{(s)}, E^{(s)})$ と $c(i|k, m, l; J^{(s)}, E^{(s)})$ を求める式を示す。

$$c(j|e; J^{(s)}, E^{(s)}) = \sum_{k=1}^m \sum_{i=0}^l \frac{t(j|e) a(i|k, m, l) \delta(j, j_k) \delta(e, e_i)}{t(j|e_0) a(0|k, m, l) + \cdots + t(j|e_l) a(l|k, m, l)} \quad (2.15)$$

$$c(i|k, m, l; J^{(s)}, E^{(s)}) = \frac{t(j_k|e_i) a(i|k, m, l)}{t(j_j|e_0) a(0|k, m, l) + \cdots + t(j_k|e_l) a(l|k, m, l)} \quad (2.16)$$

モデル2においても、最適解を推定するためにEMアルゴリズムを用いる。しかし、計算によって複数の極大値が算出され、最適解が獲得できない場合が存在する。モデル2の特殊な場合に、 $a(i|k, m, l) = (l + 1)^{-1}$ が挙げられるが、これはモデル1として考えることができる。また、最適解が保証されているモデル1で求められた値を初期値として用いることで、最適解を求めることができる。

2.2.1.3 モデル3

モデル1およびモデル2において、日単語と英単語の対応は1対1の場合のみを考慮していた。しかし、モデル3では、1つの単語が複数の単語に対応する場合や、単語の翻訳位置の距離についても考慮する。また、モデル3では単語の位置を絶対位置として考えている。モデル3では以下のパラメータを用いる。

- $P(j|e)$
英単語 e が日単語 j に翻訳される確率
- $n(\phi|e)$
英単語 e が ϕ 個の日単語と対応する確率
- $d(k|i, m, l)$
英語文の長さ l 、日本語文の長さ m のとき、 i 番目の英単語 e_i が k 番目の日単語 j_k に翻訳される確率

さらに、英単語に翻訳されない日本語の単語数を ϕ_0 として、そのような単語が発生する確率 p_0 を以下の式に表す。

$$P(\phi_0|\phi_1^l, e) = \binom{\phi_1 + \dots + \phi_l}{\phi_0} p_0^{\phi_1 + \dots + \phi_l - \phi_0} p_1^{\phi_0} \quad (2.17)$$

したがって、モデル3は以下の式によって表される。

$$P(J|E) = \sum_{a_1=0}^l \cdots \sum_{a_m=0}^l P(J, a|E) \quad (2.18)$$

$$= \sum_{a_1=0}^l \cdots \sum_{a_m=0}^l \binom{m - \phi_0}{\phi_0} p_0^{m - 2\phi_0} p_1^{\phi_0} \prod_{i=1}^l \phi_i! n(\phi_i|e_i) \times \prod_{k=1}^m t(j_k|e_{a_k}) d(k|a_k, m, l) \quad (2.19)$$

モデル3では、全ての単語対応を考慮して計算するため、計算量が膨大となる。そのため、期待値は近似によって求められる。

2.2.1.4 モデル4

モデル3とモデル4の違いは、単語の位置の考慮の仕方である。モデル3において、単語の位置は絶対位置で考慮していた。それに対して、モデル4では単語の位置を相対位置で考慮する。また、各単語ごとの位置も考慮している。モデル4では、単語位置の歪みの確率である $d(k|i, m, l)$ を以下の2通りで考慮する。

- 英単語に対応する日単語が1以上あるときに、その中で最も文頭に近い場合

$$P(\Pi_{[i]1} = k | \pi_1^{[i]-1}, \tau_0^l, \phi_0^l, E) = d_1(k - \odot_{i-1} | \mathcal{A}(e_{[i-1]}), \mathcal{B}(j_k)) \quad (2.20)$$

- それ以外の場合

$$P(\Pi_{[i]x} = k | \pi_{[i]1}^{x-1}, \pi_1^{[i]-1}, \tau_0^l, \phi_0^l, E) = d_{>1}(k - \pi_{[i]x-1} | \mathcal{B}(j_k)) \quad (2.21)$$

2.2.1.5 モデル5

モデル4では、単語の位置に関して直前の単語のみを考慮している。そのため、複数の単語が同じ位置に生じたり、単語が存在しない位置に生成されるという問題がある。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に制約が施されている。

- 繁殖数が1以上の英単語に対応する日単語中, 最も文頭に近い場合

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

v_k は k 番目までの空白数, \mathcal{A} は英単語のクラス \mathcal{B} は日単語のクラスを表している.

- それ以外の場合

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

2.2.2 GIZA++

GIZA++[6] とは, 統計翻訳に用いるための単語の確率値の計算を行うツールである. IBM 翻訳モデルのモデル1からモデル5に基づいて, 単語の対応関係の確率値を計算する. GIZA++を用いた場合, 以下のファイルが出力される.

1. **T TABLE (Translation Table)** T TABLE は, モデル1からモデル3により作成された翻訳確率 $P(f|e)$ のデータである. f は翻訳する言語で, e は目的言語である. T TABLE は各行が, 目的言語の単語 ID(e_id), 翻訳する言語の単語 ID(f_id), 翻訳する言語の単語から目的言語の単語へ翻訳する確率 ($P(f_id|e_id)$) で構成される.
2. **N TABLE (Fertility Table)** N TABLE は, 目的言語の単語における繁殖数を表したデータである. N TABLE は各行が, 目的言語の単語 ID(e_id), 繁殖数が0である確率 (p_0), 繁殖数が1である確率 (p_1), ..., 繁殖数が n である確率 (p_n) で構成される.

2.2.3 フレーズテーブル作成法

IBM モデルは、方向のある 1 対多の単語アライメントである。しかし、句に基づく統計翻訳においては単語アライメントのみでは翻訳を行うことができない。したがって、句レベルのフレーズテーブルとして両方向の 1 対多のアライメントを得る必要がある。以下にフレーズテーブル作成の手順を示す。

a). まず、GIZA++を用いて、学習文から日英、英日方向の最尤な単語アライメントを得る。日本語文“風でろうそくが消えた”と、その対訳英語文“The wind blew out the candle”を例に挙げ、図 2.2 と図 2.4 に日英方向の単語対応を示す。また、図 2.3 と図 2.5 に英日方向の単語対応を示す。なお、図 2.4 と図 2.5 において、●は対応点を示す。

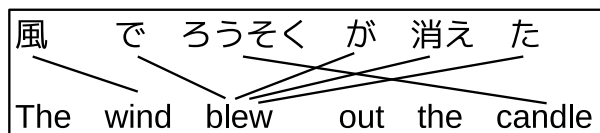


図 2.2: 日英方向の単語対応

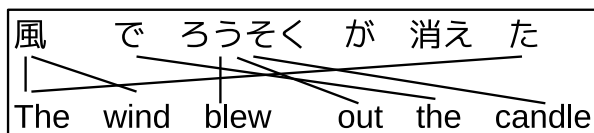


図 2.3: 英日方向の単語対応

	The	wind	blew	out	the	candle
風		●				
で			●			
ろうそく						●
が			●			
消え			●			
た			●			

図 2.4: 日英方向の単語対応

	The	wind	blew	out	the	candle
風	●	●				
で					●	
ろうそく			●	●		●
が						
消え						
た	●					

図 2.5: 英日方向の単語対応

b). 次に、両方向のアライメントから、両方向に1対多の対応を認めた単語アライメントをヒューリスティックなルールにより計算する．ここで、ヒューリスティックとは、人間の日々の意思決定に類似した直感的かつ発見的な思考方法である．基本のヒューリスティックとして、“intersection(積)”と、“union(和)”，“grow(成長)”，そして“grow-diag”がある．intersection は、両方向共に存在する対応点のみを用いる．また、union は、両方向の対応点を全て用いる．intersection の例を図 2.6 に、union の例を図 2.8 に示す．

	The	wind	blew	out	the	candle
風		●				
で						
ろうそく						●
が						
消え						
た						

図 2.6: intersection の例

	The	wind	blew	out	the	candle
風	●	●				
で			●		●	
ろうそく			●	●		●
が			●			
消え			●			
た	●		●			

図 2.7: union の例

c). そして, grow, grow-diag は intersection と union の中間である. intersection からスタートし, 既に採用した対応点の周りに union の対応点を加えていく. grow では縦と横の方向に, grow-diag では縦と横と対角に union の対応点がある場合に, その対応点を用いる. 図 2.8 に grow の例を, 図 2.9 に grow-diag の例を示す. なお, 図 2.8 と 2.9 において, ○は, intersection から追加された対応点を示す.

	The	wind	blew	out	the	candle
風	○	●				
で						
ろうそく						●
が						
消え						
た						

図 2.8: grow の例

	The	wind	blew	out	the	candle
風	○	●				
で			○		○	
ろうそく						●
が						
消え						
た						

図 2.9: grow-diag の例

d). 最後に、最終処理のヒューリスティクスとして、“final”と、“final-and”を用いる。finalは、少なくとも片方の言語の単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、両側言語の単語の単語対応がない場合に、unionの候補対応点を追加する。図2.10に grow-diag-finalの例を、図2.11に grow-diag-final-andの例を示す。ここでも、図2.10と2.11において、○は、grow-diagから追加された対応点を示す。

	The	wind	blew	out	the	candle
風	●	●				
で			●		●	
ろうそく				○		●
が			○			
消え			○			
た	○		○			

図 2.10: grow-diag-final の例

	The	wind	blew	out	the	candle
風	●	●				
で			●		●	
ろうそく				○		●
が						
消え						
た						

図 2.11: grow-diag-final-and の例

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。

2.3 言語モデル

言語モデルは、単語列の生成確率を付与するモデルである。日英翻訳では、翻訳モデルを用いて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ために用いる。統計翻訳では一般的に、 N -gram モデルを用いる。表 2.2 に N -gram モデルにおける 2-gram の例を示す。

表 2.2: N -gram の例

$\log_{10}P(w_2 w_1)$	英単語列 ($w_1 w_2$)	$\log_{10}P(w_2 w_1)$ (スムージング推定)
-4.191673	the socket	-0.3293359
-3.661356	the sofa	-0.2541532
-3.70543	the software	-0.08667657
-3.343311	the soil	-0.3595161
-4.37106	the solar	-0.09943552

表の 1 行目の例では、左の数値は “the” の後に “socket” がくる確率の常用対数 \log_{10} を求めた値である。中央は、2-gram で表された単語列、そしてバックオフスムージングで推定された確率値である。バックオフスムージングについては後述する。N-gram モデルは、“単語列 $w_1^n = w_1, w_2, \dots, w_n$ の i 番目の単語 w_i の生起確率 $P(w_i)$ は直前の $(n-1)$ 単語に依存する”，という仮説に基づくモデルである。計算式を以下に示す。

$$P(w_1^n) = P(c_1) \times P(c_2|c_1) \times P(c_3|c_1c_2) \cdots \quad (2.22)$$

$$= \prod_{i=1}^n P(w_i|c_1c_2c_3 \cdots c_{i-1}) \quad (2.23)$$

$$\simeq \prod_{i=1}^n P(w_i|w_{i-N+1}, w_{i-N+2}, \cdots, w_{i-1}) \quad (2.24)$$

例えば，“He is a man” という文字列に対する 2-gram モデルは以下のようなになる。

$$P(e = \text{“He is a man”}) \approx P(He) \times P(is | He) \times P(a | is) \times P(man | a) \quad (2.25)$$

3-gram の場合を考えると，“He is” という単語列の次に “a” 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。

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ことが必要である。そこで、出現数の少ない単語列をモデルの学習から削除する手法や、確率が0となるのを防ぐための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としてバックオフスムージング (back-off smoothing) が挙げられる。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 の値をより低次の N -gram の値から推定する。trigram の場合の例を以下に示す。

$$P(w_n|w_{n-2}^{n-1}) = \begin{cases} \lambda(w_{n-2}^n) \frac{C(w_{n-2}^n)}{C(w_{n-2}^{n-1})} & \text{if } C(w_{n-2}^n) > 0 \\ (1 - \lambda_0(w_{n-2}^{n-1}))\alpha P(w_n|w_{n-1}) & \text{else if } C(w_{n-2}^{n-1}) > 0 \\ P(w_n|w_{n-1}) & \text{otherwise} \end{cases} \quad (2.26)$$

λ はディスカウントと呼ばれ、出現しない N -gram に対して確率を付与するための係数である。また、 α は確率の和を1にするための正規化係数である。

2.4 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から、確率が最大となる翻訳候補を探索し、出力する。入力文として、「彼は医者です。」が入力されたときの翻訳例を図 2.12 に示す。

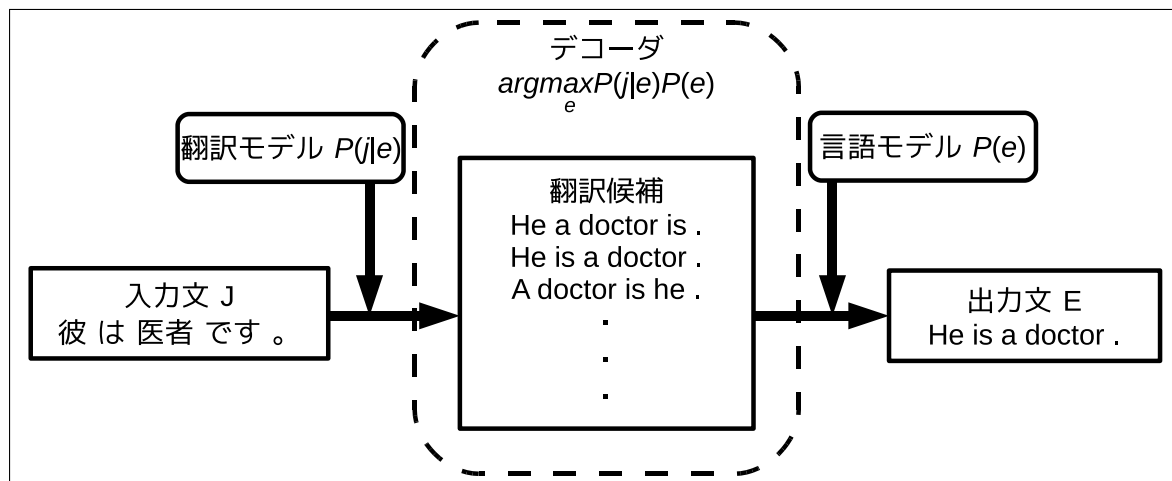


図 2.12: デコーダの動作例

デコーダは、日英統計翻訳において、 $\operatorname{argmax}_e P(j|e)P(e)$ の確率が最大となる英語文を出力するために、適切な順序で日本語と英語の単語対応を選択する必要がある。しかし、適切な英語文を決定するためには、莫大な計算量が必要となる。そこで莫大な計算量を削減するための手法として、ビームサーチ法やマルチスタック法が存在する。

2.5 パラメータチューニング

デコーダは、言語モデルや翻訳モデルに対して重みを与えることができる。例えば、言語モデルに対して高い重みを与えると、デコーダは言語モデルの確率 $P(e)$ を重視した出力を行う。各モデルに与える重みをパラメータと呼ぶ。このパラメータを最適化するため、MERT (Minimum Error Rate Training) [7] という手法を用いる。MERT は、後述する自動評価法 BLEU のスコアが最大となる翻訳結果を出力するようにパラメータ $\hat{\lambda}_1^n$ の調整を行う。 n 個のパラメータ $\hat{\lambda}_1^n$ の最適化に用いる式を以下に示す。

$$\hat{\lambda}_1^n = \operatorname{argmax}_{\lambda_1^n} \operatorname{BLEU}(\operatorname{smt}(\lambda_1^n), e_{\text{ref}}) \quad (2.27)$$

ここで、 $smt(\lambda)$ はパラメータ λ が与えられたときの、デコーダの出力文である。また、 $BLEU()$ は BLEU のスコアであり、デコーダの出力文と、入力文に対してあらかじめ用意された正解文 e_{ref} から計算される。なお、パラメータチューニングにおける入力文として、ディベロップメント文と呼ばれるデータを用いる。ディベロップメント文を試し翻訳し、各文に対して上位 N 個の翻訳候補を出力する。そして N 個の中から、より自動評価値が高い翻訳候補が上位に来るようにパラメータ $\hat{\lambda}_1^n$ を最適化する。試し翻訳とパラメータの調整を繰り返すことで、パラメータチューニングを行う。

第3章 評価手法

3.1 自動評価

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、用意された正解文と、機械翻訳システムが出力した出力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手法が提案されているが、本論文では、BLEU(BiLingual Evaluation Understudy)[8], METEOR(Metric for Evaluation of Translation with Explicit ORdering)[9], RIBES(Rank based Intuitive Bilingual Evaluation Measure)[10]を用いる。

3.1.1 BLEU

BLEU[8]は、機械翻訳の分野において、最も一般的な自動評価基準である。BLEUは、n-gram マッチ率に基づく手法を用いている。以下に計算式を示す。

$$\text{BLEU} = \text{BP}_{\text{BLEU}} * \exp\left(\frac{1}{N} \sum_{n=1}^N \log p_n\right) \quad (3.1)$$

出力される値の領域は0~1であり、高いほど翻訳文が正解文に近いことを示す。(3.1)式の N には、一般的に $N = 4$ が用いられる。 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。このペナルティは、翻訳文が正解文よりも短い場合に適合率が高くなる問題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。

$$\text{BP}_{\text{BLEU}} = \begin{cases} 1 & (c > r) \\ e^{1-\frac{r}{c}} & (c \leq r) \end{cases} \quad (3.2)$$

翻訳文の文字数が正解文よりも長い場合には、 $\text{BP}_{\text{BLEU}} = 1$ となり、影響を及ぼさない。翻訳文の文字数が正解文よりも短い場合には、その短さに応じて、 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

文は1文ではないため、各文の長さの自由度は残されている。また、(3.1)式の指数関数 p_n は以下の式で計算される。

$$p_n = \frac{\sum_{\text{全翻訳文}} \sum_{\text{正解文}} \text{正解文と一致する } N\text{-gram 数}}{\sum_{\text{全翻訳文}} \sum_{\text{正解文}} \text{全 } N\text{-gram 数}} \quad (3.3)$$

p_n は正解文と一致する N -gram 数をカウントするとき、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である。また、正解文に対して、翻訳文の方が長い場合には、 p_n の値は小さくなるため、BLEU 値も小さくなる。

3.1.2 METEOR

METEOR[9] は、再現率 R と適合率 P に基づく F 値に対して単語の非連続性に対するペナルティ関数 Pen を利用した評価基準である。以下に計算式を示す。

$$F = \frac{P \times R}{\alpha \times P + (1 - \alpha) \times R} \quad (3.4)$$

$$Pen = \gamma \times (c/m)^\beta \quad (3.5)$$

$$\text{METEOR} = F \times (1 - Pen) \quad (3.6)$$

式 3.5 のペナルティ関数 Pen にある m は機械翻訳の出力文と正解文との間で一致した単語数であり、 c は一致した各単語を対象として語順が同じものを1つのまとまりとして統合した場合のまとまりの数である。したがって、機械翻訳の出力文と正解文が完全一致の場合には $c = 1$ となり、語順が全て逆の場合には $c = m$ となる。 α , β , γ の値はパラメータである。METEOR は、0 から 1 の値をスコアとして出力し、スコアが高いほど評価が良い。

3.1.3 RIBES

RIBES[10] とは、順位相関係数を用いて共通単語の出現順序を評価する自動評価法である。順位相関係数を用いることで、文全体の語順に着目することが可能である。自動

評価と人手評価の相関について述べた松本らの研究 [11] において、日英間翻訳の場合に比較的人手評価と相関が高い自動評価法であると報告されている。以下に RIBES の計算式を示す。

$$\text{RIBES}(S) = \text{NSR} \times P^\alpha \quad (3.7)$$

$$\text{RIBES}(K) = \text{NKR} \times P^\alpha \quad (3.8)$$

式 3.7 において、NSR とはスピアマンの順位相関係数である。また、式 3.8 において NKT はケンドールの順位相関係数である。そして、 P はペナルティを表しており、評価システムの翻訳が参照文に含まれる単語を含む割合である。 α はペナルティに対する重みであり、 $0 \leq \alpha \leq 1$ の値である。RIBES は、0 から 1 の値をスコアとして出力し、スコアが高いほど評価が良い。

3.2 人手評価

本論文では、人手による評価として、対比較評価を用いる。対比較評価では、“入力文”、“正解文”、“ベースライン出力文”、“提案手法出力文”が与えられ、ベースライン出力文と提案手法出力文の比較を行う。自動評価では、正解文に近い文の評価が高い。一方、人手評価では、入力文の翻訳文として確からしい文の評価を行う。そのため、自動評価と比較して評価の精度が高い [11]。しかし、人手評価には大きなコストがかかる。

第4章 パターン翻訳を用いた学習データ 増加手法の検討

4.1 背景および目的

統計翻訳の翻訳品質は、学習データとして用いる対訳データの量に依存する [1]。しかし、人手で翻訳した高品質な対訳データの収集には量的な限界があり、作成コストも大きい。特に、日英間における統計翻訳では、対訳データの不足が大きな問題となっている。

Xiaoguang らは、中英翻訳において、単言語データをルールベース翻訳システムによって翻訳し、既存の対訳データに加えることで翻訳精度の向上を試みた [12]。また、Schwenk は、仏英翻訳において、大量の単言語データを統計翻訳によって翻訳し、対訳データに追加することで、対訳コーパスを増加させた [13]。しかし、いずれも翻訳精度の向上はほとんど認められなかった。原因として、対訳学習データへ付与した翻訳データの精度に問題があったと考えられる。

そこで本研究では、先行研究における単言語データの翻訳に、パターン翻訳を用いる手法を提案する。パターン翻訳の特徴として、パターンに照合した翻訳文の翻訳品質が高い点が挙げられる [14]。高品質な単言語データの翻訳対を対訳学習データに付与することで、既存の統計翻訳システムにおける翻訳品質の向上を目指す。

4.2 パターン翻訳の概要

パターン翻訳とは、機械翻訳手法の1つであり、大局的な文法情報を含む文パターンを用いて翻訳を行うことで、構文構造を保持した精度の高い翻訳が可能である。入力文が与えられると、文パターン辞書と入力文を照合しパターンに適合するかを調べる。文パターンに適合した場合、単語辞書を参照し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。しかし、従来のパターン翻訳は文パターン辞書を人手で作成するため、開発に時間を要する。したがって本研究では、西村らの手法を用いて文パターンの自動生成を行う [14]。また、パターン翻訳において、文パターン辞書に適合しない入力文については出力文を得られない。そのため、入力文に対するカバー率が比較的低いという問題がある。図 4.1 にパターン翻訳の概要を示す。

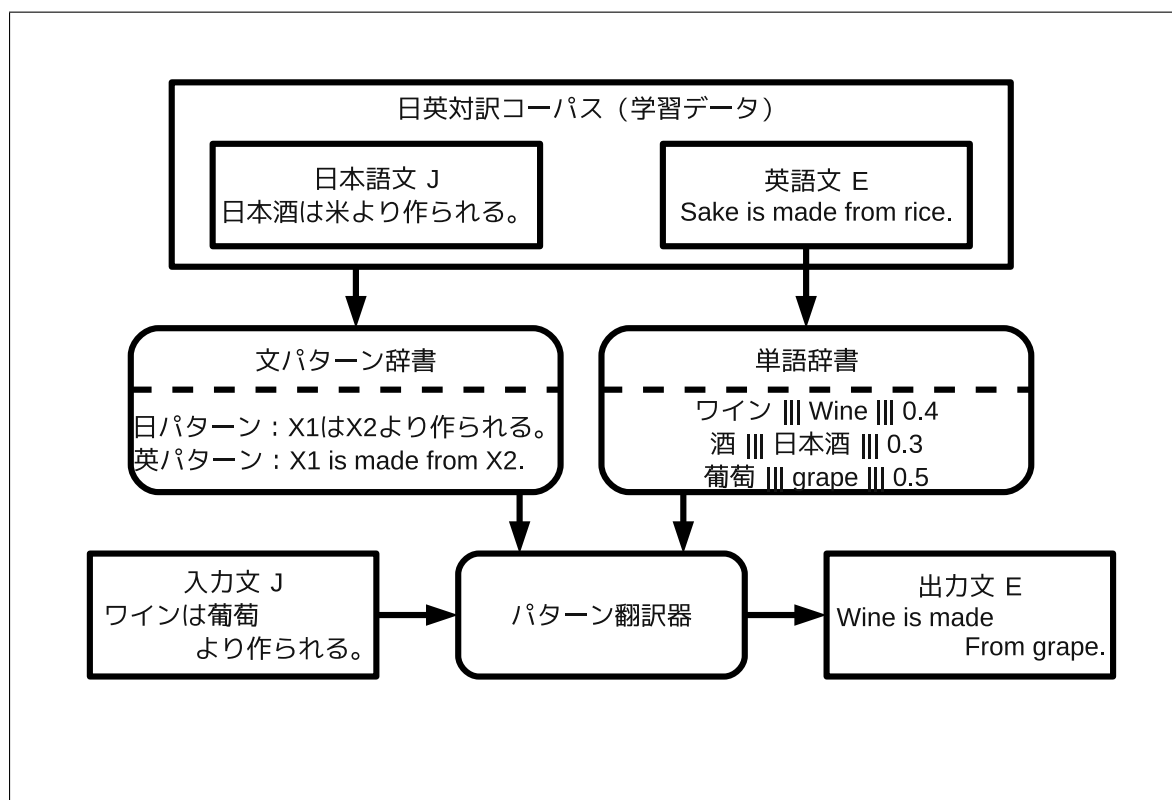


図 4.1: パターン翻訳の枠組

4.2.1 文パターン辞書

文パターン辞書は，対訳データにおいて対応する単語を変数で置き換えることで生成される．従来の手法では，変数化の際に人手で名詞や字面などの意味属性制約を変数に付与する．しかし，本研究で用いる西村らの手法では意味属性制約などを用いず，自動的に対応する単語を変数に置き換える．表 4.1 に文パターン辞書の例を示す．また，表 4.2 に単語辞書の例を示す．

表 4.1: 文パターン辞書の例

日文パターン	N00 は N01 より 作られる。
英文パターン	N00 is made from N01 .
日文パターン	我々の N00 は 異なっている。
英文パターン	Our N00 differ .
日文パターン	N00 は N01 に 身を入れた。
英文パターン	N00 put his heart into N01 .

表 4.2: 単語辞書の例

日単語	英単語	翻訳確率
祈り	prayer	0.19
祈り	prayers	0.38
祈る	pray	0.16
季節	season	0.15
季節	seasons	0.35
季節はずれ	Unseasonable	1.00

4.3 提案手法

本研究では、単言語データに対してパターン翻訳を行い、既存の対訳学習データに付与する手法を提案する。パターン翻訳の特徴である、パターンに照合した場合の高品質な翻訳文を既存の対訳学習データに付与することで、システムの翻訳精度が向上すると考えられる。図 4.2 に提案手法の概要を示す。

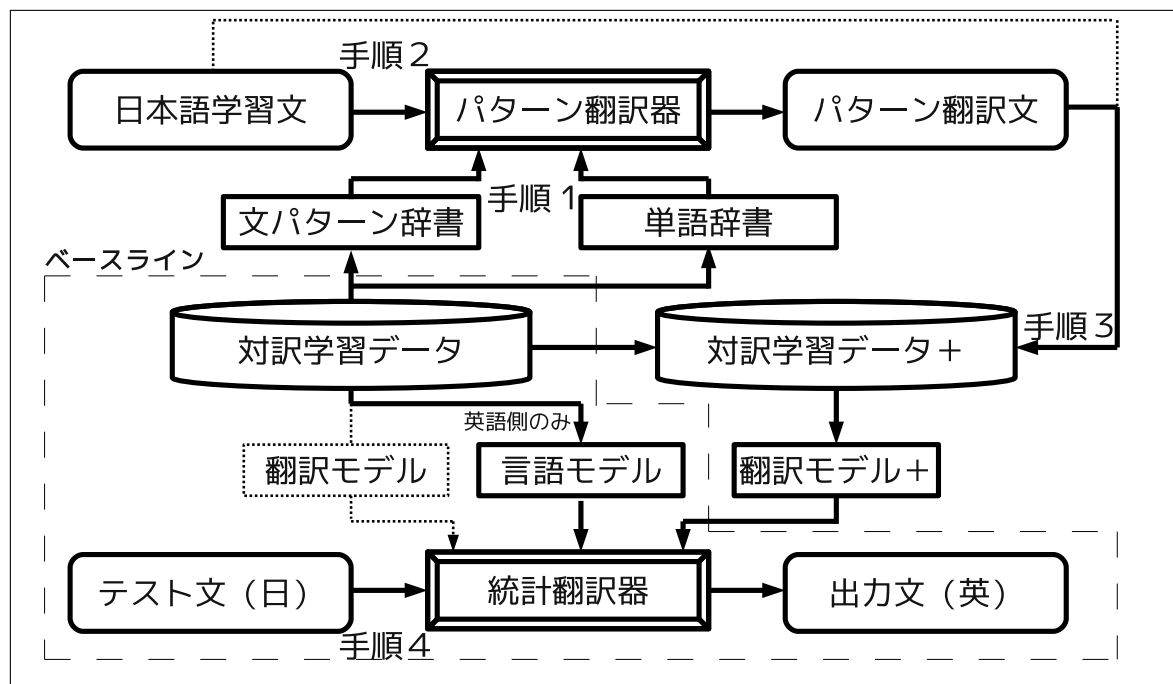


図 4.2: 提案手法の概要

4.3.1 提案手法の手順

提案手法の手順を以下に示す。

手順 1 既存の対訳学習データから文パターン辞書および単語辞書を作成する。

手順 2 手順 1 で作成した辞書を用いて、日本語学習文に対してパターン翻訳を行う。

手順 3 翻訳した文と対応する日本語学習文を既存の日英対訳学習データに付与する。

手順 4 手順 3 で作成した日英対訳学習データを用いて学習を行った統計翻訳システムを用いて日本語テスト文を翻訳する。また、既存の対訳コーパスのみを用いて統計翻訳を行う手法をベースラインとする。

4.4 実験環境

4.4.1 翻訳モデル

翻訳モデルの学習には，“GIZA++[6]”に基づいてフレーズテーブルを作成するプログラム“train-factored-phrase-model.perl[15]”を用いる．なお，本研究では，ヒューリスティックスとして，“grow-diag-final-and”を用いる．

4.4.2 言語モデル

言語モデルの学習には，“SRILM[16]”の“ngram-count”を用いる．本研究では， N -gramモデルに5-gramを用いる．

4.4.3 デコーダのパラメータ

デコーダには，“moses[15]”を用いる．また，本研究において，mosesの各パラメータを最適化しない場合を“チューニングなし”，“mert-moses.pl[15]”を用いて最適化した実験を“チューニングあり”と表記する．デコーダのパラメータの一種である“ttable-limit”と“distortion-limit”についてはパラメータチューニングでは変更されない．“ttable-limit”とは，1つの日本語のフレーズに対して考慮する，目的言語のフレーズ数の制限である．また，“distortion-limit”とは，フレーズの並び替えの範囲の制限である．よって，本研究では，“ttable-limit”の値を60，また“distortion-limit”の値を-1(無制限)としている．

4.4.4 実験データ

本研究における実験データには，辞書の例文より抽出した日英対訳文である単文コーパス，重文複文コーパスを用いる[17]．統計翻訳の前処理として，各コーパスの日本語文に対して，“MeCab[18]”を用いて形態素解析を行う．また，英語文に対して“tokenizer.perl[15]”を用いて分かち書きを行う．単文コーパスの例を表4.3に，重文複文コーパスの例を表4.4に示す．また，単文翻訳実験および重文複文翻訳実験それぞれにおけるデータの内訳を表4.5に示す．

表 4.3: 単文コーパスの例

日本語	誰だって1人ではできない。
英語	No one man can do it .
日本語	彼女は音楽の先生をしている。
英語	She is a music teacher .
日本語	それはできない相談だ。
英語	That's an impossible proposition .

表 4.4: 重文複文コーパスの例

日本語	ああいう人と付き合ってはだめだ。
英語	You should not associate with that kind of person .
日本語	流れの急な川で泳ぐのは危ない。
英語	Swimming in a rapid river is dangerous .
日本語	原っぱへ行ってキャッチボールをしよう。
英語	Let's go to the open field and play catch .

表 4.5: データの内訳

使用データ	単文	重文複文
対訳学習データ	60,000 文	50,000 文
テスト文	10,000 文	10,000 文
ディベロップメント文	2,000 文	1,200 文
日本語学習文	100,000 文	60,000 文

4.4.5 評価手法

出力文の評価において，自動評価と人手評価を行う．自動評価法には BLEU[8]，METEOR[9]，RIBES[10] を用いる．また，人手評価として対比較評価を行う．

4.5 実験結果

4.5.1 自動評価

表 4.6 にチューニングなしの場合における実験結果を示す．また，表 4.5.1 にチューニングあり翻訳実験の結果を示す．表中のベースラインは，既存の対訳学習データのみを用いて統計翻訳を行った結果である．提案手法において，パターンに適合し翻訳が行われた日本語学習文は，単文実験で 100,000 文中 55,299 文，重文複文実験で 60,000 文中 1,922 文であった．

表 4.6: 自動評価結果 (チューニングなし)

(a) 単文			
翻訳手法	BLEU	METEOR	RIBES
ベースライン	0.0936	0.4395	0.6689
提案手法	0.0980	0.4482	0.6731

(b) 重文複文			
翻訳手法	BLEU	METEOR	RIBES
ベースライン	0.0715	0.3825	0.6269
提案手法	0.0718	0.3822	0.6223

表 4.7: 自動評価結果 (チューニングあり)

(a) 単文			
翻訳手法	BLEU	METEOR	RIBES
ベースライン	0.1187	0.4805	0.6963
提案手法	0.1166	0.4841	0.6930

(b) 重文複文

翻訳手法	BLEU	METEOR	RIBES
ベースライン	0.0925	0.4205	0.6457
提案手法	0.0951	0.4303	0.6509

結果より，ベースラインと比較して単文データセットを用いたチューニングなしの実験および重文複文データセットを用いたチューニングありの実験において，提案手法の評価値が向上した．

4.5.2 人手評価

提案手法の出力文からランダムに抽出した 50 文に対して，ベースラインとの対比較評価を行う．判断基準を以下に示す．

提案手法○ 提案手法の出力結果がベースラインの出力結果よりも優れている

提案手法× 提案手法の出力結果がベースラインの出力結果よりも劣っている

差なし 文質に明確な差がない

同一出力 完全に同一の出力

表 4.8 に人手による対比較評価の結果を示す．

表 4.8: 人手評価結果

(a) 単文 (チューニングなし)			
提案手法○	提案手法×	差なし	同一出力
12	4	31	3

(b) 単文 (チューニングあり)			
提案手法○	提案手法×	差なし	同一出力
2	5	36	7

(c) 重文複文 (チューニングなし)

提案手法○	提案手法×	差なし	同一出力
1	1	36	12

(d) 重文複文 (チューニングあり)

提案手法○	提案手法×	差なし	同一出力
1	2	43	4

結果より，単文データセットを用いたチューニングなし実験において，提案手法の有効性が認められる．しかし，その他の実験ではベースラインと比較して提案手法の有効性は認められなかった．

それぞれの実験における提案手法○および提案手法×の出力例を，表 4.9～表 4.16 に示す．

表 4.9: 提案手法○の出力例 (チューニングなし単文)

入力文	わが社は神戸から大阪へ移転した。
参照文	Our office removed from Kobe to Osaka .
ベースライン	Our company from Kobe to Osaka .
提案手法	Our company relocated from Kobe to Osaka .
入力文	彼は大きな犬を飼っている。
参照文	He has a big dog .
ベースライン	He has a large have a dog .
提案手法	He has a big dog .
入力文	力が全部抜けてしまった。
参照文	All my strength is gone .
ベースライン	The all .
提案手法	All the power has gone out .

表 4.10: 提案手法×の出力例 (チューニングなし単文)

入力文	一刻を争う問題です。
参照文	It is a burning question .
ベースライン	It is a burning question .
提案手法	a burning question .
入力文	フクロウは主に夜飛ぶ。
参照文	Owls fly chiefly at night .
ベースライン	The owl flies at night .
提案手法	The night owl .
入力文	明日横浜へ参ります。
参照文	I am going to Yokohama tomorrow .
ベースライン	I'll come to Yokohama tomorrow .
提案手法	I will go to leaving Yokohama .

表 4.11: 提案手法○の出力例 (チューニングあり単文)

入力文	労働力はまだ定員に達していない。
参照文	The workforce is not yet up to full strength .
ベースライン	I have not yet reached the limit immigrant labor .
提案手法	The labor has not yet reached the limit .
入力文	火が町中に広がった。
参照文	Fire swept the town .
ベースライン	The fire propagated across the town .
提案手法	The fire spread all over town .

表 4.12: 提案手法×の出力例 (チューニングあり単文)

入力文	交通規則がこの春改正された。
参照文	The traffic regulations have been revised this spring .
ベースライン	The traffic rules revision of this spring .
提案手法	The traffic rules have been revised this spring .
入力文	それには料金がかかります。
参照文	There is a fee for that .
ベースライン	It takes a fee .
提案手法	The for it .
入力文	彼は時計を3分進めた。
参照文	He put his watch on three minutes .
ベースライン	He put his watch three minutes .
提案手法	He advanced three minutes a watch .

表 4.13: 提案手法○の出力例 (チューニングなし重文複文)

入力文	これがそのアルバムでいちばんいい歌だ。
参照文	It is the best song on the album .
ベースライン	This is the best exquisite album .
提案手法	This is the best in the album .

表 4.14: 提案手法×の出力例 (チューニングなし重文複文)

入力文	折からの強風にあおられて火はたちまち広がった。
参照文	Fanned by the strong wind , the fire spread quickly .
ベースライン	Fanned by the strong wind that prevailed at the time , the fire spread quickly .
提案手法	The , the on fire .

表 4.15: 提案手法○の出力例 (チューニングあり重文複文)

入力文	彼は私を見るとすぐ逃げた。
参照文	He ran away the moment he saw me .
ベースライン	He is to see me he ran away .
提案手法	He ran away to see me .

表 4.16: 提案手法×の出力例 (チューニングあり重文複文)

入力文	一番に間に合うように早く起きた。
参照文	I got up early , so as to be in time for the first train .
ベースライン	I get up early to be in time for the best .
提案手法	The best in time to get up early .
入力文	私は夕方の混雑を避けて早目に出かけた。
参照文	I left early to avoid the evening congestion .
ベースライン	I went to a early avoiding the crowds of the evening .
提案手法	I went to the crowds of our in the evening .

4.6 分析

4.6.1 先行研究との比較

Schwenk は単言語データを統計翻訳によって翻訳し，既存の対訳学習データに付与した．本研究では Schwenk らの手法を先行研究とし，提案手法との比較を行った．表 4.17 にチューニングなし実験の結果を示す．また，表 4.18 にチューニングあり実験の結果を示す．

表 4.17: 自動評価結果 (チューニングなし)

(a) 単文

翻訳手法	BLEU	METEOR	RIBES
先行研究	0.1086	0.4653	0.6891
提案手法	0.0980	0.4482	0.6731

(b) 重文複文

翻訳手法	BLEU	METEOR	RIBES
先行研究	0.0794	0.3992	0.6387
提案手法	0.0718	0.3822	0.6223

表 4.18: 自動評価結果 (チューニングあり)

(a) 単文

翻訳手法	BLEU	METEOR	RIBES
先行研究	0.1242	0.4903	0.7033
提案手法	0.1166	0.4841	0.6930

(b) 重文複文

翻訳手法	BLEU	METEOR	RIBES
先行研究	0.0976	0.4171	0.6505
提案手法	0.0951	0.4303	0.6509

結果より，自動評価において提案手法は先行研究よりも評価値が低下した．次に，ベースラインとの比較において有効性が認められた単文のチューニングなし実験において，提案手法と先行研究との人手による対比較評価を行った．結果を 4.6.1 に示す．

表 4.19: 人手評価結果

提案手法○	提案手法×	差なし	同一出力
8	2	35	5

結果より，単文を用いたチューニングなし実験では，人手評価における提案手法の有効性が認められる．

4.6.2 パターン翻訳における翻訳文のランク付け

江木らは，パターン翻訳出力文に対して tri-gram スコアを用いた絞り込みを行うことで，精度の高い文を選出している [19]．したがって，本研究においても同様の手法を用いてパターン翻訳出力文のランク付けを行い，各ランクを用いて実験を行った場合の結果を調査する．ランク付けは以下のように行う．

まず，対訳データの英語側における tri-gram データを用いて，パターン翻訳出力文の tri-gram スコアを計算する．スコアの計算において，該当する tri-gram が存在しない場合はペナルティ-1000 をスコアに付与する．そして，tri-gram スコアに閾値を設定し，出力文をスコア別にランク付けする．以下に本研究で用いる各ランクの定義を示す．

- A ランク : $-1000 < \text{tri-gram スコア} \leq 0$
- B ランク : $-2000 < \text{tri-gram スコア} \leq -1000$
- C ランク : $-3000 < \text{tri-gram スコア} \leq -2000$
- D ランク : $\text{tri-gram スコア} \leq -3000$

各ランクを用いた実験における自動評価結果を表 4.20 と表 4.21 に示す. 表において, 追加文数は各手法において既存の対訳データに追加した文数である. A ランク~D ランクは各ランクを用いた結果を表している. そして, A+B ランクなどはそれぞれのランクで得られた出力文を足し合わせ, 既存の対訳データに追加した結果である.

表 4.20: 自動評価結果 (チューニングなし)

(a) 単文				
翻訳手法	追加文数	BLEU	METEOR	RIBES
ベースライン	0	0.0936	0.4395	0.6689
A ランク	1,464	0.0951	0.4394	0.6675
B ランク	1,868	0.0935	0.4392	0.6688
C ランク	2,829	0.0925	0.4390	0.6668
D ランク	49,138	0.0962	0.4466	0.6731
A+B ランク	3,332	0.0935	0.4397	0.6720
A+B+C ランク	6,161	0.0949	0.4403	0.6694
提案手法	55,299	0.0980	0.4482	0.6731

(b) 重文複文				
翻訳手法	追加文数	BLEU	METEOR	RIBES
ベースライン	0	0.0715	0.3825	0.6269
A ランク	884	0.0728	0.3838	0.6228
B ランク	265	0.0717	0.3820	0.6240
C ランク	77	0.0708	0.3822	0.6232
D ランク	693	0.0712	0.3824	0.6242
A+B ランク	1,152	0.0719	0.3829	0.6240
A+B+C ランク	1,229	0.0715	0.3825	0.6229
提案手法	1,922	0.0718	0.3822	0.6223

表 4.21: 自動評価結果 (チューニングあり)

(a) 単文

翻訳手法	追加文数	BLEU	METEOR	RIBES
ベースライン	0	0.1187	0.4805	0.6963
A ランク	1,464	0.1190	0.4799	0.6938
B ランク	1,868	0.1162	0.4812	0.6965
C ランク	2,829	0.1156	0.4812	0.6939
D ランク	49,138	0.1170	0.4821	0.6925
A+B ランク	3,332	0.1178	0.4815	0.6975
A+B+C ランク	6,161	0.1182	0.4775	0.6937
提案手法	5,5299	0.1166	0.4841	0.6930

(b) 重文複文

翻訳手法	追加文数	BLEU	METEOR	RIBES
ベースライン	0	0.0925	0.4205	0.6457
A ランク	884	0.0912	0.4246	0.6478
B ランク	265	0.0926	0.4221	0.6445
C ランク	77	0.0941	0.4273	0.6463
D ランク	693	0.0930	0.4243	0.6460
A+B ランク	1,152	0.0896	0.4210	0.6350
A+B+C ランク	1,229	0.0942	0.4262	0.6454
提案手法	1,922	0.0951	0.4303	0.6509

結果より、データセットによってランク付けの影響が異なる。より精密な分析を行うため、今後は各ランクの文数を同数にして、再度実験を行う必要があると思われる。

4.6.3 文長差を用いた翻訳文の選択

パターン翻訳文の解析を行ったところ、“X1はX2である。”のような簡単な文パターンに適合して翻訳が行われた場合、入力文と出力文の文長に大きな差が出る場合があることが分かった。表 4.22 に例を示す。

表 4.22: 文長差が大きいパターン翻訳出力文の例

入力文	親切は美德である。
出力文	What is occurring is nothing less than the virtues of kindness .
入力文	彼と私のいずれかが出席する。
出力文	us has something the elements are be by those administration of him with .
入力文	会社は開発要員を工場に配置する。
出力文	Development the factories participants for are be by Hands administration of company a .

このような例では、パターン翻訳出力文の翻訳精度が悪い場合が多い。したがって、本節ではパターン翻訳出力文を文長差でフィルタリングする手法を試みる。

4.6.3.1 単文データセットにおける文長差の調査

まず、日英の文長差がいくらであれば適切であるかを調査する。本研究で用いる単文コーパス 181,988 文において各文長差の文数をカウントした。結果を表 4.6.3.1 に示す。表において、単語数の差は日本語の単語数を基準としている (+2 → 日本語の方が 2 単語多い)。

結果より、単文においては 0~+4 の文長差が比較的多数であることが確認できた。

4.6.3.2 文長差によるフィルタリング

4.6.3.1 節において、多数であった文長差を正しい文長差であると仮定し、パターン翻訳対から 0~4 の文長差の文対を抽出し、実験を行なった。実験は単文データセットにおいて行った。実験結果を表 4.24 に示す。表中の、“文長差フィルタ”が本節の実験を示し

表 4.23: 文長差の統計

文長差	文数
...	...
-3	4,400
-2	8,308
-1	15,027
0	23,711
1	29,694
2	30,464
3	24,974
4	17,150
5	9,885
6	5,339
7	2,682
...	...

ている。また、文長差フィルタにおいて、抽出が行われたのはパターン翻訳出力文 55,299 文中 44,862 文であった。

表 4.24: 自動評価結果

翻訳手法	BLEU	METEOR	RIBES
ベースライン	0.0891	0.4351	0.6632
提案手法	0.0942	0.4436	0.6680
文長差フィルタ	0.0974	0.4461	0.6716

結果より、文長差フィルタを用いた実験は、ベースラインおよび提案手法よりも評価値が向上している。したがって、文長差によるフィルタが、パターン翻訳における良い翻訳文対の選出に有効である可能性がある。今後、重文・複文などの他のデータセットにおいても調査が必要である。

4.6.4 追加の人手評価

提案手法の効果が認められた単文チューニングなしの実験において、追加の人手評価を行った。本節では、提案手法の出力文からランダムに抽出した 200 文に対して、ベースラインとの対比較評価を行う。表 4.25 に結果を示す。

表 4.25: 追加人手評価結果

提案手法○	提案手法×	差なし	同一出力
22	1	153	24

結果より，このデータセットにおいては，提案手法における明らかな有効性が確認できる．

4.6.5 市販のルールベース翻訳器を用いた実験

日本語学習文の翻訳に，市販のルールベース翻訳器を用いた場合の実験を行った．ルールベース翻訳器には東芝の“TAURAS”[20]を用いた．実験は単文コーパスデータセットを用いて，チューニングありの条件で行った．結果を表 4.26 に示す．表中の TAURAS がルールベースの翻訳文を付与した結果を示している．なお，TAURAS を用いた手法では日本語学習文 100,000 文が全て翻訳され，既存の対訳コーパスへ付与されている．

表 4.26: 自動評価結果

翻訳手法	BLEU	METEOR	RIBES
ベースライン	0.1187	0.4805	0.6963
提案手法	0.1166	0.4841	0.6930
TAURAS	0.1404	0.5183	0.7195

実験の結果，TAURAS を用いた実験では評価値が大幅に向上している．これは，TAURAS に既存の対訳コーパス以外のリソースが大量に含まれていることが原因であると考えられる．

4.6.6 使用されたフレーズテーブルの分析

自動評価および人手評価において、提案手法がベースラインよりも優れていた単文データセットを用いたチューニングなし実験における解析を行った。人手評価において、提案手法○と判断された文 12 文のうち、3 文の解析結果を以下に示す。

表 4.27: 文例 1

入力文	事態の深刻さは私たちが恐れていた以上だ。
参照文	The seriousness of the situation is greater than we had feared .
ベースライン	The state of this was terrified of more than .
提案手法	I was terrified of more than the seriousness of the situation .

表 4.28: 文例 1 で使用されたフレーズテーブル

ベースライン		提案手法	
日本語句	英語句	日本語句	英語句
事態	The state	は私	I
の深刻さ	of	恐れていた	was terrified of
は私たちが	this	以上	more than
恐れていた	was terrified of	たちが	the
以上	more than	事態の深刻さ	seriousness of the situation
だ。	.	だ。	.

文例 1 では、提案手法において“事態の深刻さ”に対して“seriousness of the situation”の選択されたことにより、よりよい翻訳が導かれている。

表 4.29: 文例 2

入力文	彼女は恥ずかしさで顔を赤らめた。
参照文	She blushed with shame .
ベースライン	Her face with shame .
提案手法	he blushed with embarrassment .

表 4.30: 文例 2 で使用されたフレーズテーブル

ベースライン		提案手法	
日本語句	英語句	日本語句	英語句
彼女は	Her	彼女は	She
で 顔	face with	顔を 赤らめた	blushed
恥ずかし さ	shame	恥ずかし さで	with embarrassment
を 赤らめた。 .	.	。	.

文例 2 も同様に，“恥ずかし さ で”に対して “with embarrassment”が選択されることでベースラインよりも正確な訳が出力されている。

表 4.31: 文例 3

入力文	彼女の脈を見つけられなかった。
参照文	He could not find her pulse .
ベースライン	Her a pulse .
提案手法	She could not find the pulse .

表 4.32: 文例 3 で使用されたフレーズテーブル

ベースライン		提案手法	
日本語句	英語句	日本語句	英語句
彼女の	Her	彼女	She
を見つけ	a	見つけられなかった	could not find
脈	pulse	を	the
られなかった。 .	.	の脈	pulse
		。	.

文例 3 の例では，“見つけられなかった”と “could not find”のフレーズが選択されている。この他の例も文例 1， 2， 3 と同様にフレーズの対応が改善された場合が多く見受けられた。

4.7 考察

結果より、提案手法において、単文データセットにおけるチューニングなし実験で提案手法の効果が確認できた。しかし、その他の実験において提案手法の効果は確認できなかった。原因として、パターン翻訳の学習が既存の対訳データを用いて行われていることが挙げられる。つまり、パターン翻訳を用いて日本語データを翻訳しても、学習データが既存の対訳データであるため、パターン翻訳出力文には有用なデータが少量しか含まれていなかったと考えられる。単文データセットにおけるチューニングなし実験で提案手法に効果が認められた原因としては、ベースライン翻訳における翻訳精度の低さが考えられる。ベースラインの翻訳精度が低かったために、パターン翻訳出力文から獲得できる情報が他のデータセットと比較して多かったのではないかと思われる。今後、使用するデータによる翻訳結果の差異の、より詳細な調査が必要である。また、提案手法において、英日方向のパターン翻訳を用いることで、日英パターン翻訳と異なる翻訳対が得られる可能性がある。

4.8 まとめ

本研究では、単言語データに対してパターン翻訳を行い、既存の対訳学習データに付与することで、対訳学習データを増加させる手法を提案した。実験の結果、自動評価では、ベースラインと比較して単文データセットを用いたチューニングなしの実験、および重文複文データセットを用いたチューニングありの実験において、提案手法の評価値が向上した。また、人手評価においては単文データセットを用いたチューニングなしの実験で提案手法の有効性が確認できた。しかし、その他の実験では提案手法の有効性は認められなかった。原因として、パターン翻訳の学習が既存の対訳データを用いて行われていることが挙げられる。今後は、使用するデータによる翻訳結果の差異の、より詳細な調査が必要である。また、提案手法において、英日方向のパターン翻訳を用いる実験が考えられる。

第5章 句に基づく統計翻訳における未知語処理の1手法

5.1 背景および目的

統計翻訳において、翻訳されない単語が出力される場合がある。翻訳されない単語の中でも、学習コーパス中には対応する訳語が存在する場合がある。本研究ではそれらの未知語と定義する。そのような単語を減少させるため、様々な試みが行われている。代表的な手法として、単語辞書などの対訳辞書データをパラレルコーパスに追加する手法がある [2]。しかし、この手法では、単語辞書などのパラレルコーパス以外のリソースが必要となる。

そこで本研究では、パラレルコーパスのみを利用して、翻訳されない単語を削減する方法を提案する。統計翻訳では、翻訳の確からしさを表すモデルとして翻訳モデルを用いている。翻訳モデルは、フレーズテーブルと呼ばれる表で管理される。フレーズテーブルは単語対応からヒューリスティックスを用いて作成される。一般的に用いられるヒューリスティックス “grow-diag-final-and” では長いフレーズが作成され、短いフレーズが作成されない傾向にある。そして、翻訳において長いフレーズが優先的に利用されるため、短い単語列、特に1単語が未知語として出力される場合がある。

一方、ヒューリスティックスの一つである “intersection(単語対応の積集合)” を用いたフレーズテーブルには、未知語として出力された単語に対応するフレーズが存在する場合がある。しかし、“intersection” を用いた翻訳では、フレーズの候補が膨大になるため、枝刈り探索の問題から翻訳効率および翻訳精度が低下する。そこで、“grow-diag-final-and” のフレーズテーブルと、未知語として出力される単語に対応する “intersection” のフレーズテーブルを併用することで、未知語が削減できる可能性がある。

本研究では、“grow-diag-final-and” と “intersection” を併用することで、未知語の削減と翻訳精度の改善を目指す。

5.2 未知語

統計翻訳において、学習データが十分であっても、未知語が出力される場合がある。フレーズテーブル作成に grow-diag-final-and を利用した翻訳において、未知語が出力された例を表 5.1 に示す。フレーズテーブルの作成方法および各ヒューリスティックの詳細については、2.5 節に記述している。また、翻訳の際に利用されたアライメントを表 5.2 に示す。

表 5.1: 未知語の出力例

入力文	ライオンが調教師に歯向かった。
参照文	The lion rose up against its trainer .
出力文	The lion 調教 turned against his teacher .

表 5.2: ベースラインのアライメント

日本語句	英語句
ライオンが	The lion
調教	調教
に歯向かつ	turned against
師	his teacher
た。	.

表 5.1 において、“調教”が翻訳されず、未知語として出力されている。表 5.1 の翻訳で用いたフレーズテーブルの一部を表 5.3 に示す。

表 5.3: フレーズテーブル (grow-dial-final-and) の一部

日本語句	ライオンは調教師にかみついた。
英語句	The lion bit his trainer the arm .
日本語句	調教師にかみついた
英語句	bit his trainer the arm
日本語句	調教師は
英語句	The trainer
日本語句	調教師はライオンを飼い馴らした
英語句	The trainer tamed the lion .

表 5.3 より、“調教”のアライメントは存在するが、grow-diag-final-and のヒューリスティックを利用しているため、このフレーズテーブルに“調教”1 単語に対応するフレーズ

は存在しない．一方，intersection を利用して作成したフレーズテーブルでは，“調教”の 1 単語に対応するフレーズが存在する．intersection を用いたフレーズテーブルの例を表 5.4 に示す．

表 5.4: フレーズテーブル (intersection) の一部

日本語句	調教
英語句	The trainer
日本語句	調教
英語句	bit his trainer
日本語句	調教
英語句	his trainer
日本語句	調教
英語句	trainer

5.3 提案手法

前節で示した通り，grow-diag-final-and で作成したフレーズテーブルに存在せず，intersection で作成したフレーズテーブルに存在するフレーズ対がある．そこで本研究では，grow-diag-final-and を用いた翻訳において，未知語として出力された単語 (1 単語) を日本語側に含む intersection のフレーズテーブル (表 5.4) を，grow-diag-final-and のフレーズテーブル (表 5.3) に追加する手法を提案する．

5.4 実験環境

5.4.1 実験データ

提案手法において，単文および重文複文 [17] を用いた実験を行う．使用したデータの内訳を表 5.5 に示す．

表 5.5: データの内訳

使用データ	単文実験	重文複文実験
対訳学習データ	単文 150,000 文	単文 100,000 文 + 重文複文 100,000 文
テスト文	単文 10,000 文	重文複文 10,000 文
ディベロップメント文	単文 1,000 文	重文複文 1,000 文

5.5 評価手法

出力文の評価において，自動評価と人手評価を行う．自動評価法は BLEU [8]，METEOR [9]，RIBES [10] を用いる．また，人手による評価として対比較評価を行う．

5.6 実験結果

5.6.1 提案手法の効果

表 5.1 の例文における提案手法の出力例を表 5.6 に示す。表中のベースラインはフレーズテーブル作成に grow-diag-final-and を用いた場合の結果である。また、表 5.6 における提案手法のアライメントを表 5.7 に示す。

表 5.6: 提案手法の出力例

入力文	ライオンが調教師に歯向かった。
参照文	The lion rose up against its trainer .
ベースライン	The lion 調教 turned against his teacher .
提案手法	The lion turned against his trainer .

表 5.7: 提案手法のアライメント

日本語単語列	英語単語列
ライオンが	The lion
に歯向かつ	turned against
師	his
調教	trainer
。	.

表 5.6 より、intersection のフレーズ対の追加による未知語の改善が確認できる。

5.6.2 未知語

5.6.2.1 未知語の改善

本節では、未知語が改善されている文を調査する。ベースライン (grow-diag-final-and) の出力において、未知語を含む文からランダムに 100 文を抽出し、提案手法の出力との比較を行った。そして、提案手法において未知語が改善されている文数を調査した。結果を表 5.8 に示す。また、例を表 5.9 に例を示す。

表 5.8: 未知語の改善文数

データセット	未知語の改善文数
単文	61/100
重文複文	68/100

表 5.9: 未知語改善の例

(a) 単文

入力文	桃太郎は鬼どもを退治した。
参照文	Momotaro subdued the ogres .
ベースライン	We finish off the 桃太郎 is so .
提案手法	We finish off the Momotaro is so .
入力文	ジェームス・ディーンは 24 年の生涯を駆け抜けた。
参照文	James Dean ran through his life of 24 years .
ベースライン	The Dean and his life of ジェームス 24 years .
提案手法	James Dean and his life at the age of 24 .
入力文	この坂は 12 度のこう配がある。
参照文	This slope has a grade of 12 degrees .
ベースライン	There are twelve degrees こう配 of this slope .
提案手法	There is a steep slope of this twelve degrees .

(b) 重文複文

入力文	彼は純情な男でだまされ易い。
参照文	He is naive and easily cheated .
ベースライン	He was a man 純情 .
提案手法	He is a man was innocent .
入力文	この海岸に立つと、晴れた日には佐渡島を望むことができる。
参照文	You can see Sado Island from this beach when the weather is clear .
ベースライン	This on the beach 佐渡 on a clear day , we hope to the island .
提案手法	We Sadogashima island on a clear day , this on the beach .
入力文	その名を聞いて彼の心に思い当たることがあった。
参照文	The name rang a bell in his mind .
ベースライン	On hearing his name 思い当たる in his mind .
提案手法	On hearing his name reminds me of him to the heart .

表 5.9 の例では“桃太郎”や“純情”の未知語が改善されている。しかし、全体の翻訳品質に影響はない。

5.6.2.2 未知語改善による翻訳品質への影響

未知語が改善した文において、翻訳品質が向上した文数の調査を行った。結果を表 5.6.2.2 に示す。また、表 5.11 に例を示す。

表 5.10: 翻訳品質の改善文数

データセット	翻訳品質が改善した文数
単文	21/61
重文複文	12/68

翻訳品質が向上した場合以外の文(単文 40 文、重文複文 46 文)において、ベースラインと提案手法の翻訳品質はほぼ同等であり、提案手法の翻訳品質が低下した例は存在しなかった。

表 5.11 では“憂鬱”や“勝ち誇っ”が翻訳されることで翻訳品質が向上している。したがって、未知語に対する提案手法の有効性が認められる。

表 5.11: 未知語改善による翻訳品質向上の例

(a) 単文	
入力文	憂鬱な天気が続いている。
参照文	The weather has been gloomy .
ベースライン	The weather has been a 憂鬱 .
提案手法	The weather has been a dark .
入力文	万事用意がととのっていた。
参照文	All was ready .
ベースライン	The ととのつ there for everything .
提案手法	There were ready for everything .
入力文	太陽の光はプリズムで七色に分けられる。
参照文	The light of the sun can be split up into seven colours with a prism .
ベースライン	The heat of the sun in プリズム is divided into seven colors .
提案手法	The prism breaks is divided into seven colors in the heat of the sun .

(a) 重文複文	
入力文	彼らは勝ち誇って帰って来た。
参照文	They came home in triumph .
ベースライン	They 勝ち誇つ came home .
提案手法	They came back with a triumphant air .
入力文	ぼくはやっと背泳ぎができるようになった。
参照文	I finally learned how to swim on my back .
ベースライン	I can 背泳ぎ at last .
提案手法	I finally got the backstroke .
入力文	薬の効果を盲信する人が多い。
参照文	Many people have blind faith in the power of medicine .
ベースライン	Many people are 盲信 the effect of the medicine .
提案手法	Many people are faith in the effect of the medicine .

5.6.3 システム全体に対する提案手法の影響

5.6.3.1 全出力文中の未知語数

ベースラインと提案手法それぞれの出力文 10,000 文において、未知語として出力された単語数を表 5.12 に示す。

表 5.12: 未知語数

(a) 単文	
翻訳手法	未知語数
ベースライン	2,486
提案手法	974

(b) 重文複文	
翻訳手法	未知語数
ベースライン	2,400
提案手法	950

結果より、単文および重文複文それぞれの実験において、未知語として出力された単語の減少が確認できる。したがって、出力文全体においても提案手法の効果が確認できる。

5.6.3.2 自動評価

表 5.13 に翻訳実験の自動評価結果を示す。表中のベースラインとはフレーズテーブル作成に “grow-diag-final-and” を用いた結果である。

表 5.13: 自動評価結果

(a) 単文			
翻訳手法	BLEU	METEOR	RIBES
ベースライン	0.1618	0.5333	0.7275
提案手法	0.1649	0.5331	0.7328

結果より、単文実験において、提案手法の効果がわずかに認められる。しかし、重文複文実験においては、提案手法の効果は認められない。

(b) 重文複文

翻訳手法	BLEU	METEOR	RIBES
ベースライン	0.1626	0.4896	0.6940
提案手法	0.1627	0.4824	0.6931

5.6.4 人手評価結果

提案手法の出力文 10,000 文からランダムに抽出した 100 文に対して、ベースラインとの対比較評価を行う。判断基準を以下に示す。

提案手法○ 提案手法の出力結果がベースラインの出力結果よりも優れている

提案手法× 提案手法の出力結果がベースラインの出力結果よりも劣っている

差なし 文質に明確な差がない

同一出力 完全に同一の出力

5.6.4.1 人手評価結果

表 5.14 に人手による対比較評価の結果を示す。

表 5.14: 人手評価結果

(a) 単文

提案手法○	提案手法×	差なし	同一出力
7	4	45	44

(b) 重文複文

提案手法○	提案手法×	差なし	同一出力
5	2	53	40

5.6.4.2 対比較評価出力例

対比較調査におけるそれぞれの評価の文例を表 5.15～表 5.18 に示す。

表 5.15: 提案手法○ (単文)

入力文	国王は退位させられた。
参照文	The King has been deposed .
ベースライン	The king was sent 退位 .
提案手法	The king abdicated the throne .
入力文	彼は親より先に死んだ。
参照文	He preceded his parents in death .
ベースライン	He was killed ahead of her parents .
提案手法	He died ahead of her parents .
入力文	何とか金の工面が ついた。
参照文	I managed to raise the money .
ベースライン	I 工面 of money .
提案手法	I somehow could manage of money .

表 5.16: 提案手法× (単文)

入力文	あなたは姉さんに非常によく似ている。
参照文	You bear a strong likeness to your sister .
ベースライン	You are very much alike in the older sister .
提案手法	You are very much alike the girl .
入力文	火が消えてしまった。
参照文	The fire has gone out .
ベースライン	The fire died out .
提案手法	The fire .
入力文	こうした危険な連鎖は、断固断たねばならない。
参照文	The nation needs to cut off such a dangerous circle of crimes .
ベースライン	We must cut determined , such a dangerous chain .
提案手法	Such a dangerous , must be determined stopped chain .

表 5.17: 提案手法○ (重文複文)

入力文	この仕事はどうしても引受ける気になれぬ。
参照文	I can not bring myself to undertake the work .
ベースライン	This work is , I can not bring myself to 引受ける .
提案手法	I can not bring myself to undertake the work by all means .
入力文	ぼくたちは落ち葉をかき集めて焚き火をした。
参照文	We raked up dead leaves and built a fire .
ベースライン	We have made a fire outside かき集め the dead leaves .
提案手法	We raked up the fallen leaves in the fire outside .
入力文	英詩の研究に全身全霊をこめて打ちこんだ。
参照文	She dedicated herself wholeheartedly to the study of English poetry .
ベースライン	She 打ちこん has devoted himself to the study of English poem .
提案手法	She devoted himself to the study of English poem my soul .

表 5.18: 提案手法× (重文複文)

入力文	彼は私にそれを盗めとはっきり言った。
参照文	He told me expressly to steal it .
ベースライン	He advised me you steal the it .
提案手法	He did it to you steal the .
入力文	彼の乗った飛行機は砂の上をかなり滑走してから離陸した。
参照文	He got a good run along the sand and then took off .
ベースライン	His plane took off from the sand with considerable runway .
提案手法	His plane took off from the sand and a runway .

5.7 解析

5.7.1 intersection との比較

フレーズテーブル作成に intersection を用いた結果と提案手法の比較を行った。自動評価結果を表 5.19 に示す。

表 5.19: 自動評価結果

(a) 単文			
翻訳手法	BLEU	METEOR	RIBES
intersection	0.1552	0.5102	0.724
提案手法	0.1649	0.5331	0.7328

(b) 重文複文			
翻訳手法	BLEU	METEOR	RIBES
intersection	0.1531	0.4575	0.6863
提案手法	0.1627	0.4824	0.6931

さらに、表 5.14 の人手による対比較調査において、提案手法○と判断された文、単文 7 文例、重文複文 5 文例において、intersection の出力文との比較を行った。結果を表 5.20 に示す。また、それぞれの実験における提案手法○の例を表 5.21、表 5.22 に示す。

表 5.20: 人手評価結果

(a) 単文			
提案手法○	提案手法×	差なし	同一出力
4	0	2	1

(b) 重文複文

提案手法○	提案手法×	差なし	同一出力
5	0	0	0

表 5.21: 提案手法○ (単文)

入力文	何とか金の工面がついた。
参照文	I managed to raise the money .
ベースライン	I 工面 of money .
intersection	I somehow manage of money
提案手法	I somehow could manage of money .
入力文	捜索隊は彼の信号にただちに応答してきた。
参照文	The search party answered his signal right away .
ベースライン	The search party was immediately responded to his signals .
intersection	The search of his reply immediately signals .
提案手法	The search party responded to his signals immediately .
入力文	これはかなり洗練された車だ。
参照文	It's quite a refined car .
ベースライン	This is a considerable refined car going .
intersection	This car is a cultivated person .
提案手法	This car refined considerably .

結果より, intersection と比較においても, 提案手法の有効性が確認できる.

表 5.22: 提案手法○ (重文複文)

入力文	この仕事はどうしても引受ける気になれぬ。
参照文	I can not bring myself to undertake the work .
ベースライン	This work is , I can not bring myself to 引受ける .
intersection	This work , but I can not bring myself to undertake it by all means .
提案手法	I can not bring myself to undertake the work by all means .
入力文	グラスが手から滑り落ちて割れた。
参照文	The glass slipped from my hand and smashed .
ベースライン	The glass fell down , and broke .
intersection	glass slid from my hand and cracked .
提案手法	The glass fell down from my hand and broke .
入力文	彼は馬のくびすを返して逃げ去った。
参照文	He turned his horse and rode away .
ベースライン	He is のくびす back the away .
intersection	He back on his heels you vacate away .
提案手法	He returned the heels vacate the horse , he ran away .

5.8 考察

実験の結果，未知語の改善における提案手法の有効性が確認できた。しかし，未知語の改善した文において，翻訳品質が向上しない文が存在した。例を表 5.23 に示す。

表 5.23 の例では“緑内障”に対して“glaucoma”，“純情”に対して“innocent”など，未知語が改善しているにも関わらず，翻訳品質は改善していない。例のように，ベースラインの文質が著しく悪い場合には，未知語が改善しても翻訳品質には影響がない。したがって，ベースラインシステムの翻訳精度がより高い場合には，未知語の改善によって翻訳品質が改善しやすいと考えられる。

表 5.23: 翻訳品質が改善しない例

(a) 単文の例

入力文	父は右目が緑内障になった。
参照文	My father had glaucoma in his right eye .
ベースライン	The 緑内障 her father .
提案手法	The her father glaucoma .
入力文	寿司のわさびを抜く。
参照文	Leave out the wasabi peppers from the sushi .
ベースライン	To remove the わさび of sushi .
提案手法	There is not enough wasabi on the sush .
入力文	桃太郎は鬼どもを退治した。
参照文	Momotaro subdued the ogres .
ベースライン	We finish off the 桃太郎 is so .
提案手法	We finish off the Momotaro is so .

(b) 重文複文の例

入力文	彼は純情な男でだまされ易い。
参照文	He is naive and easily cheated .
ベースライン	He was a man 純情 .
提案手法	He is a man was innocent .
入力文	福田さんは年は取っているが気が若い。
参照文	Fukuda is getting up there in years , but is young in spirit .
ベースライン	Mr . 福田 , but I am old .
提案手法	Mr . Fukuda , but I am old .
入力文	田中さんは首に掛けた真珠の首飾りがとても似合いました。
参照文	The pearl necklace that Mrs . Tanaka had around her neck looked very good on her .
ベースライン	Mr . Tanaka is very nice 首飾り pearl hung in the neck .
提案手法	Mr . Tanaka is very nice pearl cupidity was on the head .

5.9 まとめ

本研究では、フレーズテーブルの作成におけるヒューリスティックスの併用を行うことで、統計翻訳における未知語を減少させる手法を提案した。ヒューリスティックスの併用として、“grow-diag-final-and”のフレーズテーブルと、ベースラインシステムにおいて未知語として出力される単語に対応する“intersection”のフレーズテーブルを併用した。実験の結果、出力文全体の自動評価値に影響はなかったが、未知語の減少に大きな効果が認められた。ベースラインの出力において未知語を含む文 100 文中、単文を用いた実験では 61 文、重文複文を用いた実験では 68 文の未知語が改善した。さらに、未知語が改善した文のうち、単文実験において 61 文中 21 文、重文複文実験において 68 文中 12 文の翻訳品質が向上した。したがって、未知語問題に対して、提案手法の有効性が認められる。今後、英日方向における実験や、特許文などのデータを用いた実験における提案手法の有効性を確認する必要がある。

第6章 統語的タグを用いた統計的階層句 機械翻訳

6.1 背景と目的

従来の統計翻訳では、翻訳過程において統語的情報を用いていない。そのため、日本語と英語のような言語構造が大きく異なる言語間の翻訳において高い翻訳精度が得られない場合が多い。この問題に対して、様々な研究が行われている。

大西らは、文書レベルの文脈情報を用いてフレーズの並び替えを制限する手法を提案し、有意な成果を得た [3]。また、Zollmann らは統語的情報を用いた機械翻訳システムを提案し、公開している [4]。しかし、統語的情報を用いた機械翻訳の問題点として、統語ラベルの爆発的増加による、解析および翻訳時間の増加が挙げられる。

そこで、本研究では、日英統計的階層句機械翻訳において、‘浅い’統語的情報を用いる手法を提案する。浅い統語的情報として、日本語文の名詞句と動詞句にタグを付与する。タグを付与した日本語文を用いて学習することで、統語的情報を含む文法規則が生成され、翻訳精度が向上すると考えられる。

6.2 統計的階層句機械翻訳

統計的階層句機械翻訳 (以下, 階層型翻訳) とは, 階層構造を持った句を用いて翻訳を行う機械翻訳の手法である. 図 6.1 に階層型翻訳の枠組みを示す. 句に基づく統計翻訳と比較して, 大局的な語順の並びを同期文脈自由文法によって表現できるという特徴が挙げられる.

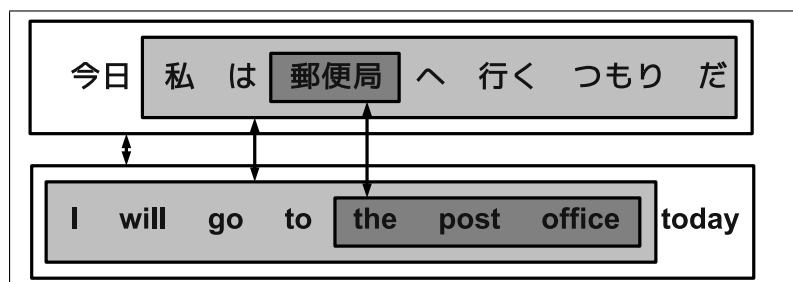


図 6.1: 階層型翻訳の枠組

6.2.1 ルールテーブル

階層型翻訳では, 翻訳モデルとして, フレーズテーブルの代わりにルールテーブルと呼ばれる表を用いる. ルールテーブルの例を表 6.1 に示す.

表 6.1: ルールテーブルの例

非終端記号	日ルール	英ルール	$P(p_j p_e)$	$P(w_j w_e)$	$P(w_e w_j)$
[X]	[X,1] と [X,2]	[X,1] and the [X,2]	1.3	2.9	1.8
[X]	[X,1] 側 [X,2]	[X,1] side [X,2]	0.6	1.8	0.5
[X]	[X,1] 側 [X,2]	[X,1] side [X,2]	0.6	2.6	1.0

左から, 非終端記号, 日本語パターン, 英語パターン, 日英フレーズ翻訳確率, 日英単語翻訳確率, 英日単語翻訳確率を示している.

6.2.2 glue-grammar

glue-grammar とは, 入力文に対して, 翻訳開始から翻訳終了までの rule-table を導くルールである. glue(接着) が示す通り, このルールがないと終端の出力まで到達しない.

表 6.2 に従来システムの glue-grammar を示す.

表 6.2: glue-grammar

非終端記号	日ルール	英ルール
[S]	[X,1]	[X,1]
[S]	[S,1] [X,2]	[S,1] [X,2]

表 6.2 は左から, 非終端記号, 日本語パターン, 英語パターンを示している.

6.3 提案手法

階層型翻訳は, 大局的な語順の並びを考慮している. しかし, 翻訳過程において, 統語的情報は用いていない. そこで提案手法では, 翻訳対象言語である日本語文の名詞句と動詞句にタグを付与することで, 統語的情報を表現する.

以下に提案手法の概略を示す.

6.3.1 提案手法の概略

手順 1 GIZA++[6] を用いて, 日本語-英語間の単語対応を獲得

手順 2 日本語文を構文解析し, 解析データを元に名詞句と動詞句にタグ (<NP>, </NP>, <VP>, </VP>) を付与

手順 3 タグが付与された部分の単語対応を修正

手順 4 タグ付与後の日本語コーパスと, 英語コーパスから文法規則を学習

手順 5 タグが付与された日本語テスト文を翻訳し, 翻訳精度を調査

6.3.2 タグ付与の手順

名詞句のタグ付与手順を以下に示す. また, タグの付与例を表 6.3 に示す.

準備 日本語文を構文解析

手順 1 係り受け関係の連続しない後置詞句において, 名詞を含み, 動詞・助動詞を含まない部分を抽出

手順2 係り受け関係の連続する後置詞句において、最後部の句が名詞を含み、動詞・助動詞を含まない部分を抽出

手順3 手順1, 手順2で抽出したものから、末尾の助詞・記号を削除

手順4 手順3で得られた名詞句の中から単語数が最大かつ単語数が2以上の句にタグを付与

表 6.3: タグ付与の例

原文	そんなやり方をすれば必ず苦情が起こる。
タグ付与後	<NP> そんなやり方 </NP> をすれば必ず <VP> 苦情が起こる </VP>。

6.3.3 提案手法の glue-grammar

従来の階層型翻訳で用いる glue-grammar については 6.2.2 節で述べた。しかし、従来の glue-grammar では、タグが付いた学習文やテスト文に対して正しいルールが選択できない。そこで、本研究では glue-grammar に新たなルールを追加し、実験を行う。表 6.4 に本研究で用いる glue-grammar を示す。

表 6.4: 提案手法における glue-grammar

非終端記号	日ルール	英ルール
[S]	[X,1]	[X,1]
[S]	[S,1] [X,2]	[S,1] [X,2]
[X]	<NP> [X,1] </NP>	[X,1]
[X]	<VP> [X,1] </VP>	[X,1]

6.4 実験環境

日本語文の構文解析，および分かち書きには Cabocha[21] を用いる．また，デコーダとして Joshua[22] を用いる．なお，Joshua のパラメータはディベロップメントデータを用いて最適化を行う．また，自動評価法として BLEU[8] を用いる．

6.4.1 実験データ

本研究では，単文・重文複文を用いた実験に加え，特許文を用いた実験を行う．実験データの内訳を表 6.5 に示す．

表 6.5: データの内訳

	単文	重文複文	特許文
対訳学習データ	100,000 文	100,000 文	300,000 文
テスト文	10,000 文	10,000 文	1,251 文
ディベロップメント文	1,000 文	1,000 文	2,000 文

また，表 6.6 にそれぞれのデータにおけるタグ付与の例を示す．

表 6.6: タグ付与例

(a) 単文

日本語文	政治は国民の生活に深い関係を持っている。
タグ付与後	政治は <NP> 国民の生活 </NP> に <VP> 深い関係を持っている </VP> 。
英語文	Politics have a deep bearing on the lives of the citizens.

(b) 重文

日本語文	校長先生が突然訪ねてきたので、母は即席で食事を用意した。
タグ付与後	<NP> 校長先生 </NP> が <VP> 突然訪ねてきた </VP> ので、母は即席で食事を用意した。
英語文	When the principal of our school paid a surprise visit to our house, my mother quickly set out some food for him.

(c) 複文

日本語文	家に帰る途中で食事を済ませた。
タグ付与後	<NP> 家に帰る途中 </NP> で <VP> 食事を済ませた </VP>。
英語文	I had a meal on my way home.

(d) 特許文

日本語文	この挿入穴 17 に上記プレート支持部材 23 の他端部が挿入嵌合されている。
タグ付与後	この挿入穴 17 に <NP> 上記プレート支持部材 23 の他端部 </NP> が <VP> 挿入嵌合されている </VP>。
英語文	Said other end part of the plate support member 23 is fitted into the insertion hole 17.

6.5 実験結果

6.5.1 自動評価

各実験データにおける自動評価結果を表 6.7 に示す。表中におけるベースラインは、タグ付与を行わないコーパスを用いた翻訳結果である。

表 6.7: 自動評価結果

(a) 単文実験

翻訳手法	BLEU
ベースライン	10.19
提案手法	9.46

結果より、全ての実験において提案手法の有効性は認められなかった。

(b) 重文複文実験

翻訳手法	BLEU
ベースライン	7.41
提案手法	5.88

(c) 特許文実験

翻訳手法	BLEU
ベースライン	26.05
提案手法	24.75

6.5.2 対比較調査

単文と重文複文の実験において、ベースラインと提案手法の出力結果に対して人手による対比較調査を行った。評価は、出力結果からランダムに抽出した20文を用いる。結果を表6.8に示す。

表 6.8: 人手評価結果

(a) 単文

提案手法○	提案手法×	差なし	同一出力
0	0	14	6

結果より、ベースラインと提案手法には、差が認められなかった。表6.5.2に単文における出力文例を示す。また、表6.5.2に重文複文における出力文例を示す。

(b) 重文複文

提案手法○	提案手法×	差なし	同一出力
0	0	19	1

表 6.9: 単文における出力文例

ベースライン入力文	こちらへおいでの節はぜひお立ち寄りください。
提案手法入力文	こちらへ <NP> おいでの節 </NP> は <VP> ぜひお立ち寄りください </VP>。
参照文	Please drop in on us if you happen to come this way .
ベースライン出力文	When to the , I drop .
提案手法出力文	Please of of the to the , I drop .

表 6.10: 重文複文における出力文例

ベースライン入力文	電車にかさを忘れたのを思い出し、駅に取って返した。
提案手法入力文	電車に <VP> かさを忘れたのを思い出し </VP>、駅に取って返した。
参照文	I remembered I left my umbrella on the train , so I hurried back to the station .
ベースライン出力文	I remembered my umbrella in the train , I to the station .
提案手法出力文	I forgot the umbrella in the train , to the station .

6.6 考察

実験結果より，提案手法において有効性は認められなかった．本節では以下の5つの視点から原因の考察を行う．

6.6.1 生成されたルールテーブルの考察

表 6.6.1 に，生成されたルールテーブルの例を示す．なお，表中の [X,N] は非終端記号を表している．

表 6.11: 生成されたルールの例

日本語ルール	英語ルール
<NP> 日本 の [X,1] < /NP>	The [X,1] in Japan
<VP> 新聞 に 載っ た < /VP> 。	was printed in the newspaper .
< /NP> の <VP>	finding fault with

表 6.6.1 において，1 行目と 2 行目の例は正しく統語的情報が付与されているといえる．しかし，3 行目と 4 行目の例は明らかな誤りである．したがって，明らかに誤りのあるルールをフィルタリングすることによって翻訳精度の向上に繋がる可能性がある．

6.6.2 デコーダが考慮するフレーズ長の問題

Joshua のパラメータの 1 つである MaxPhraseSpan は，各終端記号における最大終端記号数を表している．しかし，タグを付与することにより，単語数が増加し，MaxPhraseSpan を上回る句が増加した．そのため，デコーダが考慮するフレーズが減少し，翻訳精度が減少したと考えられる．表 6.6.2 に MaxPhraseSpan を 10 から 14 まで増加させた場合におけるベースラインと提案手法の自動評価結果を示す．

表 6.12: MaxPhraseSpan の増加による翻訳精度の変化

翻訳手法	MaxPhraseSpan				
	10	11	12	13	14
ベースライン	26.05	26.33	26.60	26.63	26.71
提案手法	24.75	25.52	25.76	26.00	26.20

表 6.6.2 より, MaxPhraseSpan の増加に伴い, ベースラインと提案手法における差の減少が認められる.

6.6.3 動詞句タグの有効性

動詞句タグの有効性を調査するため, 名詞句のみにタグを付与した場合と提案手法との比較を行った. 自動評価における比較結果を表 6.6.3 に示す. なお, 実験には単文を用いている.

表 6.13: 動詞句タグの効果

手法	BLEU
名詞句タグのみ	9.21
提案手法 (名詞句 + 動詞句)	9.46

表 6.6.3 より, 動詞句へのタグ付与は, 名詞句のみへのタグ付与よりも有効であることが示された.

6.6.4 階層型翻訳におけるタグ付与の効果

先行研究と提案手法では, タグの付与方法が異なる. よって, 提案手法のタグの付与方法の有効性を調査するため, 句に基づく統計翻訳においてタグ付与実験を行った. 句に基づく統計翻訳におけるデコーダは Moses デコーダ [15] を用いる. また, タグの付与は名詞句のみに行った. 表 6.14 に実験結果を示す. なお, 表中のベースラインは通常のコーパスを用いた, 句に基づく統計翻訳である.

表 6.14: 句に基づく統計翻訳におけるタグ付与の有効性

手法	BLEU
ベースライン	11.01
タグ付与	12.21

表 6.14 より, 句に基づく統計翻訳において, タグ付与手法が有効であることが示された. よって今回のタグ付与手法では, 階層型翻訳においては有効ではないと考えられる.

6.6.5 統語的情報の不足

本研究では、統語的情報として、日本語文の名詞句と動詞句にタグを付与した。しかし、日本語文において、それぞれ最長の名詞句・動詞句のみに付与を行った。よって、複数の名詞句・動詞にタグ付与を行うことで統語的情報が増加することが考えられる。また、階層型翻訳において元言語である日本語文にタグを付与するだけでは、統語的情報としては不十分であったと考えられる。対象言語である英語にも何らかの統語的情報を付与することで、より構文的な情報を伴った文法規則が得られる可能性がある。

6.7 まとめ

本研究では、日英統計的階層句機械翻訳において、統語的情報として日本語文の名詞句・動詞句にタグを付与する手法を提案した。しかし、提案手法において有意な効果は認められなかった。原因として、日本語文へのタグ付与のみでは、統語的情報として不十分であることなどが挙げられる。今後の展開として、生成されたルールのフィルタリングを行う手法や、目的言語である英語側においても、統語的情報を付与する手法が考えられる。

第7章 結論

本論文では、統計翻訳に関する3つの研究を扱った。以下にそれぞれの研究についてまとめる。

- **パターン翻訳を用いた学習データ増加手法の検討**

本研究では、単言語データに対してパターン翻訳を行い、既存の対訳学習データに付与することで、対訳学習データを増加させる手法を提案した。実験の結果、単文データセットにおいてデコーダのチューニングを行わない実験で、自動評価および人手による対比較評価から有効性が確認できた。しかし、その他のデータセットにおいては提案手法の有効性は認められなかった。原因として、パターン翻訳の学習データに既存の対訳学習データを用いていることが考えられる。つまり、パターン翻訳を用いて日本語学習データを翻訳しても、パターン翻訳の学習に用いるデータがベースラインと同様であるため有用な情報が獲得できないのだと思われる。しかし、少量の単文データを学習データとして用いる実験など、ベースラインの翻訳精度が著しく低い場合に提案手法が有効に働く可能性がある。今後、使用するデータによる翻訳結果の差異の、より詳細な調査が必要である。

- **句に基づく統計翻訳における未知語処理の1手法**

本研究では、統計翻訳における未知語を減少させる手法を提案した。提案手法の特徴として、フレーズテーブルの作成におけるヒューリスティックスの併用を行うことで、対訳辞書データなどの外部リソースを必要としない点が挙げられる。ヒューリスティックスの併用として、“grow-diag-final-and”のフレーズテーブルと未知語として出力される単語に対応する“intersection”のフレーズテーブルを併用する。実験の結果、出力文全体の自動評価値に影響はなかったが、未知語の減少に大きな効果が認められた。ベースラインの出力において未知語を含む文100文中、単文を用いた実験では61文、重文複文を用いた実験では68文の未知語が改善した。さらに、未知語が改善した文のうち、単文実験において61文中21文、重文複文実験において68文中12文の翻訳品質が向上した。したがって、未知語問題に対して、提案手

法の有効性が認められる。今後の展開として、特許文を用いた実験など、未知語が大量に生成される場合における提案手法の有効性を確認していきたい。

- **統語的タグを用いた統計的階層句機械翻訳**

本研究では、日英統計的階層句機械翻訳において、統語的情報として日本語文の名詞句・動詞句にタグを付与する手法を提案した。統語的情報を含んだ学習文を用いて翻訳システムの学習を行い、統語的情報を含む文法規則を生成することで、システムの翻訳精度向上を目指した。しかし、単文・重文複文・特許文全ての実験において、提案手法の効果は認められなかった。原因として、日本語文へのタグ付与のみでは、統語的情報として不十分であることなどが挙げられる。今後の展開として、生成されたルールのフィルタリングを行う手法や、目的言語である英語側においても、統語的情報を付与する手法が考えられる。

謝辞

最後に、2年間に渡り、本研究のご指導を頂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研究室の村田真樹教授，村上仁一准教授，徳久雅人講師に深く感謝す
ると共に，厚く御礼申し上げます。また，インターンシップ中のご指導や，その後のご
助言を頂きました情報通信研究機構の渡辺太郎氏に感謝いたします。そして，日常の議
論を通じて多くの知識や示唆を頂いた同研究室の皆様方，参考にさせていただいた著書
の著者の方々に，感謝の気持ちと御礼を申し上げたく，謝辞にかえさせていただきます。

参考文献

- [1] 猪澤雅史, 村上仁一, 徳久雅人, 池原悟: “統計翻訳における単文・重文複文の翻訳精度の評価”, 言語処理学会第14回年次大会発表論文集, pp.869–872, 2008.
- [2] 日野聡子, 村上仁一, 徳久雅人, 村田真樹: “鳥バンクと英辞郎を日英対訳文に追加した統計翻訳の調査”, 言語処理学会第18回年次大会発表論文集, pp.491–494, 2012.
- [3] Takashi Onishi, Masao Utiyama, Eiichiro Sumita: “Reordering Constraint Based on Document-Level Context”, The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL HLT 2011), pp.434–438, 2011.
- [4] Andreas Zollmann, Ashish Venugopal: “Syntax Augmented Machine Translation via Chart Parsing”, Proceedings of the Workshop on Statistical Machine Translation, pp.138–141, 2006.
- [5] Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, Robert L. Mercer: “The mathematics of statistical machine translation: Parameter Estimation”, Computational Linguistics, 1993.
- [6] Franz Josef Och, Hermann Ney: “A Systematic Comparison of Various Statistical Alignment Models”, Computational Linguistics, pp.19–51, 2003.
- [7] 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.
- [8] 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.

- [9] 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.
- [10] 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.
- [11] 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 回年次大会発表論文集, pp.505–508, 2012.
- [12] Xiaoguang Hu, Haifeng Wang, Hua Wu: “Using RBMT Systems to Produce Bilingual Corpus for SMT”, Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp.287–195, 2007.
- [13] Holger Schwenk: “Investigations on Large-Scale Lightly Supervised Training for Statistical Machine Translation”, Proceedings of the IWSLT, pp.182–189, 2008.
- [14] 西村拓哉, 村上仁一, 徳久雅人, 池原悟: “日英対訳文対を用いたパターン翻訳器の自動作成法の検討”, 言語処理学会第 18 回年次大会発表論文集, pp.159–162, 2012.
- [15] 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.
- [16] Andreas Stolcke: “SRILM - an Extensible Language Modeling Toolkit”, 7th International Conference on Spoken Language Processing, pp.901–904, 2002.
- [17] 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.
- [18] 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.

- [19] 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回年次大会発表論文集, pp.263–266, 2012.
- [20] Shinya Amano, Hideki Hirakawa, Yoshinao Tsutsumi: “TAURAS: The Toshiba machine translation system”, Manuscr Program MT Summit, pp.15–23, 1987.
- [21] Taku Kudo, Yuji Matsumoto: “Japanese Dependency Analysis using Cascaded Chunking”, CoNLL 2002: Proceedings of the 6th Conference on Natural Language Learning 2002 (COLING 2002 Post-Conference Workshops), pp.63-69, 2002.
- [22] Zhifei Li, Chris Callison-Burch, Chris Dyer, Juri Ganitkevitch, Sanjeev Khudanpur, Lane Schwartz, Wren Thornton, Jonathan Weese and Omar Zaidan: “Joshua: An Open Source Toolkit for Parsing-based Machine Translation”, In Proceedings of the Workshop on Statistical Machine Translation (WMT09), pp.135–139, 2009.