

# 概要

統計翻訳は翻訳の際に翻訳モデルと言語モデルを用いる。翻訳モデルは日本語の句と英語の句との対応をフレーズテーブルで管理している。フレーズテーブルは、学習データの日本語文と英語文を自動的に対応付けをして作成する。そのため作成されたフレーズテーブルのカバー率は高いが信頼性は低い。フレーズテーブルの信頼性を高めるためには、句対応の信頼性が高い人手で作成された対訳フレーズ辞書の利用が考えられる。

先行研究として鏡味らは、人手で作成した対訳フレーズ辞書である鳥バンク [1] を直接フレーズテーブルに追加し、翻訳精度が向上したことを報告している [2]。また、東江らは、人手で作成した対訳フレーズ辞書である英辞郎 [3] を直接フレーズテーブルに追加し、翻訳精度が向上したことを報告している [4]。

そこで、本研究では人手で作成された対訳フレーズ辞書を用いて、学習データ中の英語フレーズと日本語フレーズをそれぞれ一つのまとまりとした対訳フレーズデータを作成する。対訳フレーズデータを用いると、人手で作成されたフレーズを含むフレーズテーブルが作成できる。その結果、フレーズテーブルの信頼性が向上し、翻訳精度が向上すると考える。本研究では提案手法で作成したフレーズテーブルを用いて単文コーパスと重文複文コーパスに対してそれぞれ、日英統計翻訳と英日統計翻訳を行った。

その結果、人手評価において鳥バンクを用いた提案手法の単文と重文複文の英日翻訳と英辞郎を用いた提案手法の日英翻訳の重文複文の評価結果はベースラインより優れており、それ以外の翻訳実験はベースラインより劣っているという結果になった。

一方、自動評価において英辞郎を用いた提案手法の全ての実験においてはBLEU, METEORの値がベースラインより向上した。鳥バンクを用いた提案手法はほとんどの自動評価においてベースラインよりスコアが低くなった。

結果から、ほとんどの実験において提案手法はベースラインより劣っていた。しかし鳥バンクを用いた提案手法は英日統計翻訳において有効であることが確認できた。

# 目次

第1章	はじめに	1
第2章	統計翻訳システム	3
2.1	基本概念	3
2.2	翻訳モデル	4
2.2.1	IBMモデル	4
2.3	GIZA++	5
2.4	フレーズテーブルの作成方法	5
2.5	言語モデル	8
2.5.1	$N$ -gramモデル	8
2.6	デコーダ	9
2.7	評価方法	9
2.7.1	自動評価法	9
2.7.2	BLEU	10
2.7.3	NIST	10
2.7.4	METEOR	10
2.7.5	人手評価法	11
第3章	先行研究	12
3.1	先行研究手法	12
3.2	実験データ	15
3.2.1	実験結果	15
3.2.1.1	自動評価結果	15
3.2.1.2	人手評価結果	15
第4章	提案手法	17
4.1	提案手法の流れ	17

<b>第5章</b>	<b>実験環境</b>	<b>20</b>
5.1	実験内容 . . . . .	20
5.2	実験データ . . . . .	20
5.3	学習データの数 . . . . .	21
5.4	翻訳モデルの学習 . . . . .	21
5.5	言語モデルの学習 . . . . .	22
5.6	デコーダのパラメータ . . . . .	22
5.7	対訳フレーズ辞書 . . . . .	22
5.7.1	鳥バンク . . . . .	22
5.7.2	英辞郎 . . . . .	23
<b>第6章</b>	<b>翻訳実験</b>	<b>24</b>
6.1	対訳フレーズデータ . . . . .	24
6.2	実験結果 . . . . .	24
6.2.1	自動評価 . . . . .	24
6.2.2	人手評価 . . . . .	25
6.2.3	出力例 . . . . .	26
<b>第7章</b>	<b>考察</b>	<b>32</b>
7.1	提案手法 (鳥バンク) と提案手法 (英辞郎) . . . . .	32
7.2	人手評価と自動評価 . . . . .	32
7.3	対訳フレーズデータの効果の分析 . . . . .	33
7.3.1	文質の低下によってベースラインより訳質が低下した場合 . . . . .	33
7.3.2	主語の消失によってベースラインより訳質が低下した場合 . . . . .	34
7.3.3	動詞の消失によってベースラインより訳質が低下した場合 . . . . .	34
7.3.4	文質の向上によってベースラインより訳質が向上した場合 . . . . .	35
7.3.5	動詞の出現によってベースラインより訳質が向上した場合 . . . . .	36
7.3.6	未知語の消失によってベースラインより訳質が向上した場合 . . . . .	36
7.3.7	主語の出現によってベースラインより訳質が向上した場合 . . . . .	37
7.4	対訳フレーズデータのみを用いた翻訳実験 . . . . .	37
7.4.1	自動評価結果 . . . . .	37
<b>第8章</b>	<b>おわりに</b>	<b>39</b>

# 目 次

2.1 日英統計翻訳の流れ . . . . .	3
4.1 提案手法の流れ . . . . .	18

# 表 目 次

2.1	フレーズテーブルの例 . . . . .	4
2.2	日英方向の単語対応 . . . . .	5
2.3	英日方向の単語対応 . . . . .	5
2.4	積集合 intersection . . . . .	6
2.5	和集合 union . . . . .	6
2.6	ヒューリスティック grow-diag-final . . . . .	7
2.7	作成されたフレーズ対の例 . . . . .	7
2.8	作成されたフレーズテーブルの例 . . . . .	7
3.1	grow-diag-final で作成したフレーズテーブルの例 . . . . .	12
3.2	intersection で作成したフレーズテーブルの例 . . . . .	13
3.3	対訳フレーズ対の形式をフレーズテーブルの形式に変換した例 . . . . .	13
3.4	対訳フレーズ対に翻訳確率を付与した例 . . . . .	14
3.5	手順1で作成したフレーズテーブルに翻訳確率を付与した翻訳対を追加した例 . . . . .	14
3.6	先行研究で使用した実験データ . . . . .	15
3.7	先行研究の自動評価結果 . . . . .	15
3.8	対比較実験結果 (鳥バンク) . . . . .	16
3.9	対比較実験結果 (英辞郎) . . . . .	16
4.1	マッチングの例 . . . . .	17
4.2	対訳フレーズデータの例 . . . . .	19
4.3	“-”を取り除く前のフレーズテーブル . . . . .	19
4.4	“-”を取り除いた後のフレーズテーブル . . . . .	19
5.1	対訳文の例 . . . . .	21
5.2	鳥バンクフレーズ対の例 . . . . .	23

5.3	クリーニング後の英辞郎のフレーズ対の例	23
6.1	対訳フレーズデータの数	24
6.2	自動評価結果(日英翻訳)	25
6.3	自動評価結果(英日翻訳)	25
6.4	人手評価結果(鳥バンク 日英)	25
6.5	人手評価結果(鳥バンク 英日)	26
6.6	人手評価結果(英辞郎 日英)	26
6.7	人手評価結果(英辞郎 英日)	26
6.8	提案手法(鳥バンク)日英統計翻訳結果(単文)	27
6.9	提案手法(鳥バンク)日英統計翻訳結果(重文複文)	27
6.10	提案手法(鳥バンク)英日統計翻訳結果(単文)	28
6.11	提案手法(鳥バンク)英日統計翻訳結果(重文複文)	29
6.12	提案手法(英辞郎)日英統計翻訳結果(単文)	30
6.13	提案手法(英辞郎)日英統計翻訳結果(重文複文)	30
6.14	提案手法(英辞郎)英日統計翻訳結果(単文)	31
6.15	提案手法(英辞郎)英日統計翻訳結果(重文複文)	31
7.1	ベースラインより翻訳品質が低下した原因	33
7.2	ベースラインより翻訳品質が向上した原因	33
7.3	“文質の低下”の例	33
7.4	“主語の消失”の例	34
7.5	“動詞の消失”の例	35
7.6	“文質の向上”の例	35
7.7	“動詞の出現”の例	36
7.8	“未知語の消失”の例	36
7.9	“主語の出現”の例	37
7.10	自動評価結果(日英翻訳)	38
7.11	自動評価結果(英日翻訳)	38

# 第1章 はじめに

現在、機械翻訳において、対訳データから自動的に翻訳規則を生成し、翻訳を行う統計翻訳が注目されている。統計翻訳は獲得した翻訳規則を翻訳モデルと言語モデルで管理する。翻訳モデルは、原言語から目的言語への条件付き確率を計算するモデルである。言語モデルは、単語の列や文字の列が起こる確率を計算するモデルである。翻訳モデルは日本語と英語の句対応を管理するフレーズテーブルを用いる。通常フレーズテーブルは、学習データの日本語文と英語文を自動的に対応付けをして作成する。そのため作成されたフレーズテーブルのカバー率は高いが信頼性は低い。そこで、フレーズテーブルの信頼性を高めるために、人手で作成された辞書の利用が考えられる。人手で作成された辞書は日本語の句と英語の句との対応の信頼性が高い。そのため、統計翻訳に組み込むことによって、フレーズテーブルの信頼性と翻訳精度が向上するのではないかと考えた。

先行研究として、鏡味らは人手で作成した辞書である鳥バンクを用いた。鏡味らは、確率を付与した鳥バンクをフレーズテーブルに直接付与することにより、翻訳精度の向上を報告している [2]。また、東江らは人手で作成された辞書として英辞郎を用いた。東江らは、確率を付与した英辞郎をフレーズテーブルに直接付与することにより、翻訳精度の向上を報告している [4]。

しかし、人手で作成された辞書の利用には他にも方法が考えられる。

本研究では人手で作成された対訳フレーズ辞書を用いて、学習データ中の英語フレーズと日本語フレーズをそれぞれ一つのまとまりとした対訳フレーズデータを作成する。対訳フレーズデータを用いると、人手で作成されたフレーズを含むフレーズテーブルが作成できる。その結果、フレーズテーブルの信頼性が向上し、翻訳精度が向上すると考える。

本研究では、人手で作成された辞書として鳥バンクと英辞郎をそれぞれ用いる。また、単文コーパスと重文複文コーパスに対してそれぞれ、日英統計翻訳と英日統計翻訳を行う。つまり合計8種類の実験を行った。

その結果、人手評価において鳥バンクを用いた提案手法の単文と重文複文の英日翻訳と英辞郎を用いた提案手法の日英翻訳の重文複文の評価結果はベースラインより優れて

おり、それ以外の翻訳実験はベースラインより劣っているという結果になった。

一方、自動評価において英辞郎を用いた提案手法の全ての実験においてはBLEU, METEORの値がベースラインより向上した。鳥バンクを用いた提案手法はほとんどの自動評価においてベースラインよりスコアが低くなった。

本論文の構成は以下の通りである。まず、2章で統計翻訳システムの概要を示し、各モデルの学習について述べる。3章では、先行研究手法、実験環境と翻訳実験結果について述べる。4章では、本研究の提案手法について述べる。5章では、実験に用いるデータやツールといった実験環境について述べる。6章では、提案手法の結果を示す。そして、7章で考察を行い、最後に8章で結論を述べ、まとめる。



## 第2章 統計翻訳システム

統計翻訳システムを，原言語（翻訳の対象となる入力された言語）を日本語，目的言語（翻訳された後に出力される言語）を英語とする日英統計翻訳の場合を例として説明する．

### 2.1 基本概念

日英統計翻訳統計翻訳システムは，日本語入力文  $j$  が与えられたとき，全ての組み合わせの中から確率が最大となる英語文  $\hat{e}$  を探索することで翻訳を行う．以下に基本モデルを示す．

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

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

$P(j|e)$  は翻訳モデル， $P(e)$  は言語モデルと呼ぶ．また， $\hat{e}$  を探索する翻訳システムをデコーダと呼ぶ．図 2.1 に日英統計翻訳の流れを示す．

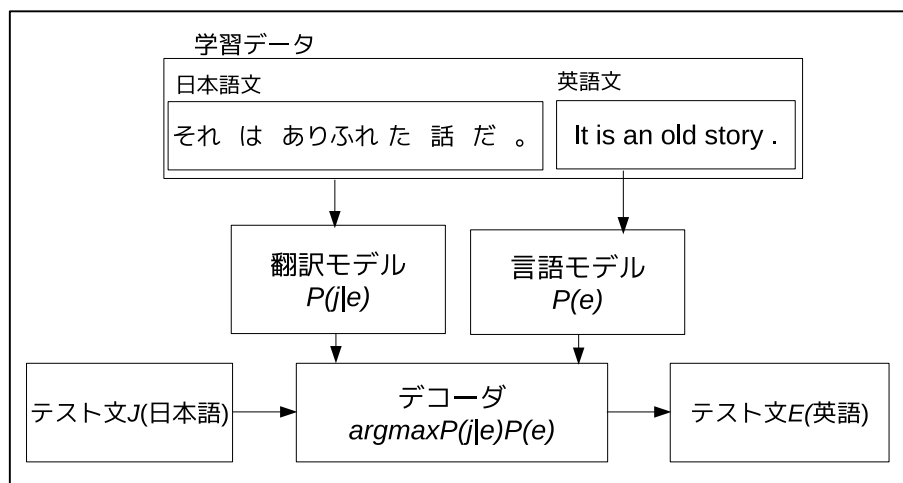


図 2.1 日英統計翻訳の流れ

## 2.2 翻訳モデル

翻訳モデルは日本語の単語列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。翻訳モデルには、大きくわけて語に基づく翻訳モデルと句に基づく翻訳モデルがある。初期の統計翻訳は、語に基づく翻訳モデルを用いていた。しかし、翻訳確率の高さから、現在では句に基づく翻訳モデルが主流となっている。句に基づく翻訳モデルは表 2.1 に示すフレーズテーブルで管理される。

表 2.1 フレーズテーブルの例

20秒		twenty seconds		0.5	0.133	1	0.013	2.718
あらし		the storm		0.0625	0.245	0.0476	0.0353	2.718
その博物館		the museum		0.125	0.037	0.142	0.201	2.718
その馬		the horse		0.25	0.0575	0.0454	0.182	2.718

左から、日本語フレーズ、英語フレーズ、フレーズの英日翻訳確率  $P(j|e)$ 、英日方向の単語の翻訳確率 (IBM モデル) の積、フレーズの日英翻訳確率  $P(e|j)$ 、日英方向の単語の翻訳確率 (IBM モデル) の積、フレーズのペナルティである。以後、フレーズペナルティは常に一定の値であるため省略する。

### 2.2.1 IBM モデル

翻訳モデルの代表例として IBM 翻訳モデルがある。IBM モデルは英語文  $e$ 、日本語文  $j$  の翻訳モデル  $P(j|e)$  を計算するためにアラインメント  $a$  と呼ばれる概念を導入し、以下のような式を考える。なお、アラインメントとはある日本語単語  $j$  と英単語  $e$  の対応関係のことを示す。

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

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

## 2.3 GIZA++

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

## 2.4 フレーズテーブルの作成方法

GIZA++よりIBMモデルを推定することで最尤な単語アラインメントを得る。これを日英、英日の両方向に対して行う。対訳文，“We ate all the apple pie”“私たちはアップルパイを全部食べてしまった”を学習データとした時の例で日英方向の単語対応の例を表2.2に、英日方向の単語対応の例を表2.3に示す。また、表中の“■”は獲得した最尤な単語アラインメントを表す。

表 2.2 日英方向の単語対応

	私	たち	は	アップルパイ	を	全て	食べ	て	しまっ	た
We	■	■								
ate							■		■	■
all						■		■		
the					■					
apple										
pie				■						

表 2.3 英日方向の単語対応

	私	たち	は	アップルパイ	を	全て	食べ	て	しまっ	た
We		■								
ate							■			
all						■				
the					■					
apple				■						
pie				■						

次に、両方向のアラインメントから、両方向に1対多の対応を認めた単語アラインメントの計算を行う。この単語アラインメントは基本的に両方向の単語対応の積集合(intersection)と和集合(union)の中間をヒューリスティックで求める。なお、積集合は

両方向ともに単語対応が存在する場合のみ対応を残し，和集合は少なくとも片方向に単語対応が存在する場合，単語対応を残す．対称な単語対応を求めるヒューリスティック (grow-diag-final など) は，まず積集合から始まり，和集合にしかない単語対応が妥当であるかを判断しながら単語対応を徐々に加える．対称化された単語アラインメントの獲得の例を表 2.4(積集合 intersection)，表 2.5(和集合 union) 表 2.6(ヒューリスティック grow-diag-final) に示す．

表 2.4 積集合 intersection

	私	たち	は	アップルパイ	を	全て	食べ	て	しまっ	た
We		■								
ate							■			
all						■				
the					■					
apple										
pie				■						

表 2.5 和集合 union

	私	たち	は	アップルパイ	を	全て	食べ	て	しまっ	た
We	■	■								
ate							■		■	■
all						■		■		
the					■					
apple				■						
pie				■						

対称化された単語アラインメントのうち，矛盾しないすべてのフレーズ対応を得る．抽出したフレーズ対の例を表 2.7 に示す．

抽出したフレーズ対応に対して確率付けを行う．日本語フレーズ  $J_{phrase}$  と英語フレーズ  $E_{phrase}$  からなるフレーズ対応の確率は以下の式で計算される．また，フレーズ対に確率を付与し，作成されたフレーズテーブルの例を表 2.8 に示す．

$$P(J_{phrase}|E_{phrase}) = \frac{\text{学習データ中で } J_{phrase} \text{ と } E_{phrase} \text{ が同時に出現した数}}{\text{学習データ中で } E_{phrase} \text{ が出現した数}} \quad (2.3)$$

表 2.6 ヒューリスティック grow-diag-final

	私	たち	は	アップルパイ	を	全て	食べ	て	しまっ	た
We	■	■								
ate							■			
all						■				
the					■					
apple				■						
pie				■						

表 2.7 作成されたフレーズ対の例

---



---

アップルパイ     apple pie
アップルパイ を 全部 食べて しまっ た     ate all the apple pie
私 たち     We
私 たち は     We
全て     all
食べて しまっ た     ate

---



---

$$P(E_{phrase}|J_{phrase}) = \frac{\text{学習データ中で } J_{phrase} \text{ と } E_{phrase} \text{ が同時に出現した数}}{\text{学習データ中で } J_{phrase} \text{ が出現した数}} \quad (2.4)$$

表 2.8 作成されたフレーズテーブルの例

---



---

アップルパイ     apple pie     1 1 1 1
アップルパイ を 全部 食べて しまっ た     ate all the apple pie     1 1 1 1
私 たち     We     1 1 1 1
私 たち は     We     1 1 1 1
全て     all     1 1 1 1
食べて しまっ た     ate     1 1 1 1

---



---

## 2.5 言語モデル

言語モデルは翻訳候補の文に対して目的言語の文らしさの指標を与えるモデルである。翻訳モデルでは、訳語の選択や訳語の位置の選択に対する評価を与えることはできるが、作られた翻訳候補が目的言語の文としてふさわしいかどうかを判断する評価を与えることはできない。そのため、言語モデルでは日英翻訳の場合、より英語らしい文に対して、高い確率を与えることで、翻訳モデルで翻訳された訳文候補の中から英語として自然な文を選出する。

### 2.5.1 $N$ -gram モデル

統計翻訳では一般的に、 $N$ -gram モデルを用いる。 $N$ -gram モデルとは“単語列  $P(W_1^n) = w_1^n = w_1, w_2, w_3, \dots, w_n$  の  $i$  番目の単語  $w_i$  の生起確率  $P(w_i)$  は直前の  $(N - 1)$  の単語列  $w_{i-(N-1)}, w_{i-(N-2)}, w_{i-(N-3)}, \dots, w_{i-1}$  に依存する”という仮説に基づくモデルである。また、 $N = 1$  のモデルを uni-gram,  $N = 2$  のモデルを bi-gram,  $N = 3$  のモデルを tri-gram と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。 $N$ -gram は以下の式で表現される。ここで、 $w_i^j$  は  $i$  から  $j$  番目までの単語列を表す。

$$P(W_1^n) = P(w_1) \times P(w_2|w_1) \times P(w_3|w_1^2) \dots P(w_n|w_1^{n-1}) \quad (2.5)$$

$$\approx P(w_1) \times P(w_2|w_1) \times P(w_3|w_1^2) \dots P(w_n|w_{n-(N-1)}^{n-1}) \quad (2.6)$$

$$= \prod_{i=1}^n P(w_i|w_{i-(N-1)}^{i-1}) \quad (2.7)$$

また、 $P(w_i|w_{i-(N-1)}^{i-1})$  は以下の式で計算される。ここで  $C(w_1^i)$  は単語列  $w_1^i$  が出現する頻度を表す。

$$P(w_i|w_{i-(N-1)}^{i-1}) = \frac{C(w_{i-(N-1)}^i)}{C(w_{i-(N-1)}^{i-1})} \quad (2.8)$$

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

$$P(\text{“I have dogs .”}) \simeq P(I) \times P(\text{have}|I) \times P(\text{dogs|have}) \dots P(.|\text{dogs}) \quad (2.9)$$

tri-gram モデルであれば,  $P(\text{dogs}|I \text{ have})$ , 4-gram モデルであれば  $P(\cdot|I \text{ have dogs})$  となる.

(2.8) 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. 低頻度な語彙の場合,  $C(w_{i-(N-1)}^i), C(w_{i-(N-1)}^{i-1})$  の値が小さく, 信頼性が低い. また, 学習データ中に単語列  $w_1^i$  が存在しない場合, この単語列の出現確率は 0 と推定される. そのため, (2.8) 式から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て, 各単語列の出現数を高める必要がある. そこで, 出現頻度の少ない単語列をモデルの学習から削除 (カットオフ) する方法や, 確率が 0 となるのを防ぐために, 大きい確率を小さく, 小さい確率を大きくするスムージング手法が提案されている. スムージングの代表的な手法にバックオフ・スムージングがある. バックオフ・スムージングは学習データに出現しない  $N$ -gram の値をより低い次数の  $(N-1)$ -gram の値から推定する.

## 2.6 デコーダ

デコーダは翻訳モデルと言語モデルの全ての組み合わせの中から確率が最大となる出力文を探索して翻訳を行う. 代表的なデコーダに Moses[6] がある.

## 2.7 評価方法

本実験では, 翻訳システムによって出力した文の評価に自動評価法と人手評価法を用いる.

### 2.7.1 自動評価法

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, 翻訳システムで出力した文と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ある. 本研究では, BLEU[7], NIST[8], METEOR[9] を用いる.

## 2.7.2 BLEU

BLEU は語順 (4-gram) が正しい場合に高いスコアを出す。BLEU は以下の式で計算される。

$$BLEU_{score} = BP \times \exp \left( \sum_{n=1}^N w_n \log p_n \right) \quad (2.10)$$

$$w_n = \frac{1}{N} \quad (2.11)$$

$$p_n = \frac{\sum_i \text{出力文 } i \text{ と正解文 } i \text{ で一致した } N\text{-gram の数}}{\sum_i \text{出力文 } i \text{ の } N\text{-gram の数}} \quad (2.12)$$

ここで、 $P_n$  は出力文と正解文の  $N$ -gram の一致率を表している。BLEU はこの一致率を 1-gram から 4-gram まで計算し、その幾何平均をとる。また、出力文が正解文より短い場合、“ $\sum_i$  出力文  $i$  の  $N$ -gram の数” が小さくなり、不当にスコアが高くなる可能性がある。そこで、正解文より短い文に対するペナルティとして、 $BP$  を用いる。 $BP$  は出力文が正解文より長い場合は 1 をとなり、出力文が正解文より短い場合は 1 未満の値となる

## 2.7.3 NIST

NIST では BLEU と同様に語順の正しさを比較を行うが、5-gram を用いる。NIST は以下の式で計算される。

$$NIST_{score} = BP \times \sum_{N=1}^N \frac{\sum_i \text{出力文 } i \text{ と正解文 } i \text{ に共通する } w_1 \dots w_n}{\sum_i \text{出力文 } i \text{ と正解文 } i \text{ で一致した } N\text{-gram の数}} \text{Info}_i(w_1 \dots w_n) \quad (2.13)$$

$$\text{Info}_i(w_1 \dots w_n) = \log_2 \frac{\text{評価コーパス中 } (w_1 \dots w_{n-1}) \text{ の数}}{\text{評価コーパス中の } (w_1 \dots w_n) \text{ の数}} \quad (2.14)$$

## 2.7.4 METEOR

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。METEOR は以下の式で計算される。

$$METEOR_{score} = F_{mean} \times (1 - Pen) \quad (2.15)$$



$$F_{mean} = \frac{P \times R}{\alpha \times P + (1 - \alpha) \times R} \quad (2.16)$$

$$Pen = \gamma \times \left(\frac{c}{m}\right)^\beta \quad (2.17)$$

METEORはまず再現率  $R$  と適合率  $P$  に基づく F 値を求め、次に、単語の非連続性に対するペナルティとして関数  $Pen$  を与える。ペナルティ関数  $Pen$  において、 $m$  は出力文と正解文の単語の一致率を表す。そして、 $c$  は一致した単語を対象に、正解文と語順が同じものを1つのまとまりとして統合した場合の、まとまりの数を表す。そのため、出力文と正解文が同じ文であるとき  $c=1$  となる。また、一致率の計算において、WordNet による類義語を用いて、似た意味を持つ単語は同一であると判断される。 $\alpha$ ,  $\beta$ ,  $\gamma$  の値はパラメータである。本研究では、 $\alpha=0.9$ ,  $\beta=3.0$ ,  $\gamma=0.5$  の値を用いる。

BLEU と METEOR では 0 から 1 までの間で評価され、NIST では 0 から  $\infty$  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。

### 2.7.5 人手評価法

ベースラインの翻訳結果と提案手法の翻訳結果からそれぞれランダムに 100 文ずつ抽出し、対比較評価を行う。

提案手法の翻訳結果がベースラインの翻訳結果より優れている場合は“提案手法○”とする。一方、提案手法の翻訳結果がベースラインの翻訳結果より劣っている場合は“ベースライン○”とする。また、提案手法の翻訳結果とベースラインの翻訳結果に変化が見られたが、翻訳品質がどちらも変わらない場合は、“差なし”とし、提案手法の翻訳結果とベースラインの翻訳結果が同じ場合、“同一出力”とする。

## 第3章 先行研究

鏡味らと東江らの実験について説明する。両者の実験の違いは用いた人手で作成された対訳フレーズ辞書が鳥バンクであるか英辞郎であるかという点のみであり、手法は同じであるため、3.1節でまとめて説明する。

### 3.1 先行研究手法

先行研究では、対訳フレーズ対に翻訳確率を付与し、フレーズテーブルに追加する。通常、フレーズテーブルを作成するためのパラメータには grow-diag-final を使用する。しかし、grow-diag-final で作成されるフレーズ対は、長いフレーズ対を短いフレーズ対に分割する。長いフレーズ対が短く分割されてしまうと、長い翻訳対に対して、翻訳確率を付与することができないという問題がある。

そこで、長い翻訳対に対して翻訳確率を付与するために、パラメータ intersection で作成したフレーズテーブルのフレーズ対と対訳フレーズ対のマッチングを行う。フレーズ対と対訳フレーズ対が完全に一致した場合にのみ対訳フレーズ対にフレーズテーブルで算出された翻訳確率を付与する。そして、パラメータ grow-diag-final で作成したフレーズテーブルに、翻訳確率を付与した対訳フレーズ対を追加する。

先行研究の手法を以下に示す。

手順 1 学習データからパラメータ grow-diag-final でフレーズテーブルを作成する。grow-diag-final で作成したフレーズテーブルの例を表 3.1 に示す。

表 3.1 grow-diag-final で作成したフレーズテーブルの例

あまりに	too	0.0366	0.0452	0.2038	0.4038
あらゆる場合に	in every case	0.5	0.0006	0.1111	0.0031
あらゆる点で	in every respect	0.0294	0.0002	0.3333	0.0001

手順2 学習データからパラメータ intersection でフレーズテーブルを作成する. intersection で作成したフレーズテーブルの例を表 3.2 に示す.

表 3.2 intersection で作成したフレーズテーブルの例

あらゆる	every	0.028	0.0215	0.2888	0.134
あらゆる	took every possible	0.5	0.02155	0.0444	4.1243
いつ	when	0.002	0.0025	0.0102	0.0098
いつも	as usual ,	0.04545	0.3188	0.00087	2.394
いつも	always	0.0425	0.3188	0.0034	0.0002
いつもの	as usual	0.0425	0.0216	0.0784	0.0002
いつもの	eleven as usual	0.1428	0.0216	0.0196	4.1846
いつものように	as usual	0.0106	1.1908	0.5	0.0001
いつものように	usual	0.0056	1.1908	0.5	0.0294

手順3 対訳フレーズ対の形式をフレーズテーブルの形式に変換する. 対訳フレーズ対の形式をフレーズテーブルの形式に変換した例を表 3.3 に示す.

表 3.3 対訳フレーズ対の形式をフレーズテーブルの形式に変換した例

あらゆる	all
あらゆる	every
いつ	when
いつから	how long?
いつも	always
いつも	at all times
いつものように	as usual
いつものように	as always

手順4 手順2で作成したフレーズテーブルを参照して対訳フレーズ対に翻訳確率を付与する. 対訳フレーズ対に翻訳確率を付与した例を表 3.4 に示す.

対訳フレーズ対の日本語フレーズと英語フレーズが intersection で作成したフレーズテーブルのフレーズと一致した場合に, その翻訳確率を対訳フレーズ対に付与する. たとえば, 表 3.3 の 2 行目 “あらゆる ||| every |||” に翻訳確率を付与する場合,

表 3.4 対訳フレーズ対に翻訳確率を付与した例

あらゆる	every	0.028	0.0215	0.2888	0.134
いつ	when	0.002	0.0025	0.0102	0.0098
いつも	always	0.0425	0.3188	0.0034	0.0002
いつものように	as usual	0.0106	1.1908	0.5	0.0001

表 3.2 の “あらゆる ||| every ||| 0.028 0.0215 0.2888 0.1340 ” 1 行目の日本語フレーズと英語フレーズと一致する。よって，対訳フレーズ対 “あらゆる ||| every |||” に翻訳確率 “0.028 0.0215 0.2888 0.134” を付与する。他の対訳フレーズ対に対しても同様にして翻訳確率を付与する。

手順 5 手順 1 で作成したフレーズテーブルに翻訳確率を付与した対訳フレーズ対を追加する。手順 1 で作成したフレーズテーブルに翻訳確率を付与した翻訳対を追加した例を表 3.5 に示す。

表 3.5 手順 1 で作成したフレーズテーブルに翻訳確率を付与した翻訳対を追加した例

あまりに	too	0.0366	0.0452	0.2038	0.4038
あらゆる場合に	in every case	0.5	0.0006	0.1111	0.0031
あらゆる点で	in every respect	0.0294	0.0002	0.3333	0.0001
あらゆる	every	0.028	0.0215	0.2888	0.1340
いつ	when	0.002	0.0025	0.0102	0.0098
いつも	always	0.0425	0.3188	0.0034	0.0002
いつものように	as usual	0.0106	1.1908	0.5	0.0001

手順 6 手順 5 で追加したフレーズテーブルを用いて統計翻訳を行う。

3.1 節の手順 1 のフレーズテーブルを用いた翻訳を先行研究のベースラインとする。また，人手で作成された対訳フレーズ辞書として鳥バンクを用いた先行研究手法を先行研究 (鳥バンク) と呼び，英辞郎を用いた先行研究手法を先行研究 (英辞郎) と呼ぶ。

## 3.2 実験データ

先行研究 (鳥バンク) と先行研究 (英辞郎) はそれぞれ単文コーパスと重文複文コーパスに対して日英統計翻訳実験を行った。それぞれの翻訳実験で用いた学習データとテストデータを表 3.6 に示す。

表 3.6 先行研究で使った実験データ

	先行研究 (鳥バンク)		先行研究 (英辞郎)	
	単文	重文複文	単文	重文複文
学習データ	100,000	121,913	100,000	100,000
テストデータ	1,000	100	10,000	10,000

### 3.2.1 実験結果

#### 3.2.1.1 自動評価結果

先行研究の翻訳結果に対して自動評価を行った結果を 3.7 に示す。なお、自動評価として先行研究 (鳥バンク) は BLEU を用い、先行研究 (英辞郎) は BLUE と METEOR を用いる。

表 3.7 先行研究の自動評価結果

評価方法	先行研究 (鳥バンク)		先行研究 (英辞郎)			
	単文	重文複文	単文)		重文複文	
	BLEU	BLEU	BLEU	METEOR	BLEU	METEOR
ベースライン	0.125	0.077	0.118	0.360	0.082	0.323
先行研究手法	0.134	0.085	0.121	0.368	0.084	0.329

#### 3.2.1.2 人手評価結果

先行研究の翻訳結果からランダムに 100 文抽出し、ベースラインとの対比較評価を行った。人手評価において、“先行研究○” は先行研究手法がベースラインより優れている場合とし、“ベースライン○” は先行研究手法がベースラインより劣っている場合とする。先行研究 (鳥バンク) の人手評価において、“変化なし” は先行研究手法とベースラインの

翻訳結果が変化しなかった場合とする。また、先行研究 (英辞郎) の人手評価において、“差なし” は先行研究手法とベースラインの翻訳結果に変化はあったが、文質が変わらなかった場合とし、“同一出力” は先行研究手法の出力とベースラインの出力が同じである場合とする。先行研究 (鳥バンク) の人手評価結果を表 3.8 に示し、先行研究 (英辞郎) の人手評価結果を 3.9 に示す。

表 3.8 対比較実験結果 (鳥バンク)

	先行研究○	ベースライン○	変化なし
単文	5	23	72
重文複文	15	35	50

表 3.9 対比較実験結果 (英辞郎)

	先行研究○	ベースライン○	差なし	同一出力
単文	5	3	4	88
重文複文	9	0	5	86

## 第4章 提案手法

本研究では，信頼性の高いフレーズテーブルを作成するために，人手で作成されたフレーズ辞書を用いて学習データ中の英語フレーズと日本語フレーズを“-”を用いてそれぞれ一つのまとまりとした対訳フレーズデータを作成する．そして，対訳フレーズデータを用いてフレーズテーブルを作成し，統計翻訳を行う．対訳フレーズデータを用いてフレーズテーブルは作成した場合，“-”でまとめられたフレーズは1単語として扱われる．そのため，不自然に途切れたフレーズは作成されず，人手で作成されたフレーズ辞書を含むフレーズテーブルが作成可能である．

### 4.1 提案手法の流れ

日英統計翻訳の場合の提案手法の流れを図4.1に示す．

手順1 人手で作成された対訳フレーズ辞書のフレーズ対と，学習データ(日本語文(1)と英語文(1))のマッチングを行う．一致した場合は手順2を行う．例を表4.1に示す．表4.1の例では，学習データの英語文に“old story”日本語文に“ありふれた話”が含まれる文がマッチングする．また，学習データの英語文に“I”日本語文に“私は”が含まれる文と，英語文に“coming of age ceremony”日本語文に“成人式”が含まれる文とマッチングする．

表 4.1 マッチングの例

対訳フレーズ対	: old story     ありふれた話 : I     私は : coming of age ceremony     成人式
学習データ 日本語文(1)	: それはありふれた話だ。
学習データ 英語文(1)	: It is an old story .
学習データ 日本語文(1)	: 私は成人式に出席した。
学習データ 英語文(1)	: I attended a coming of age ceremony .

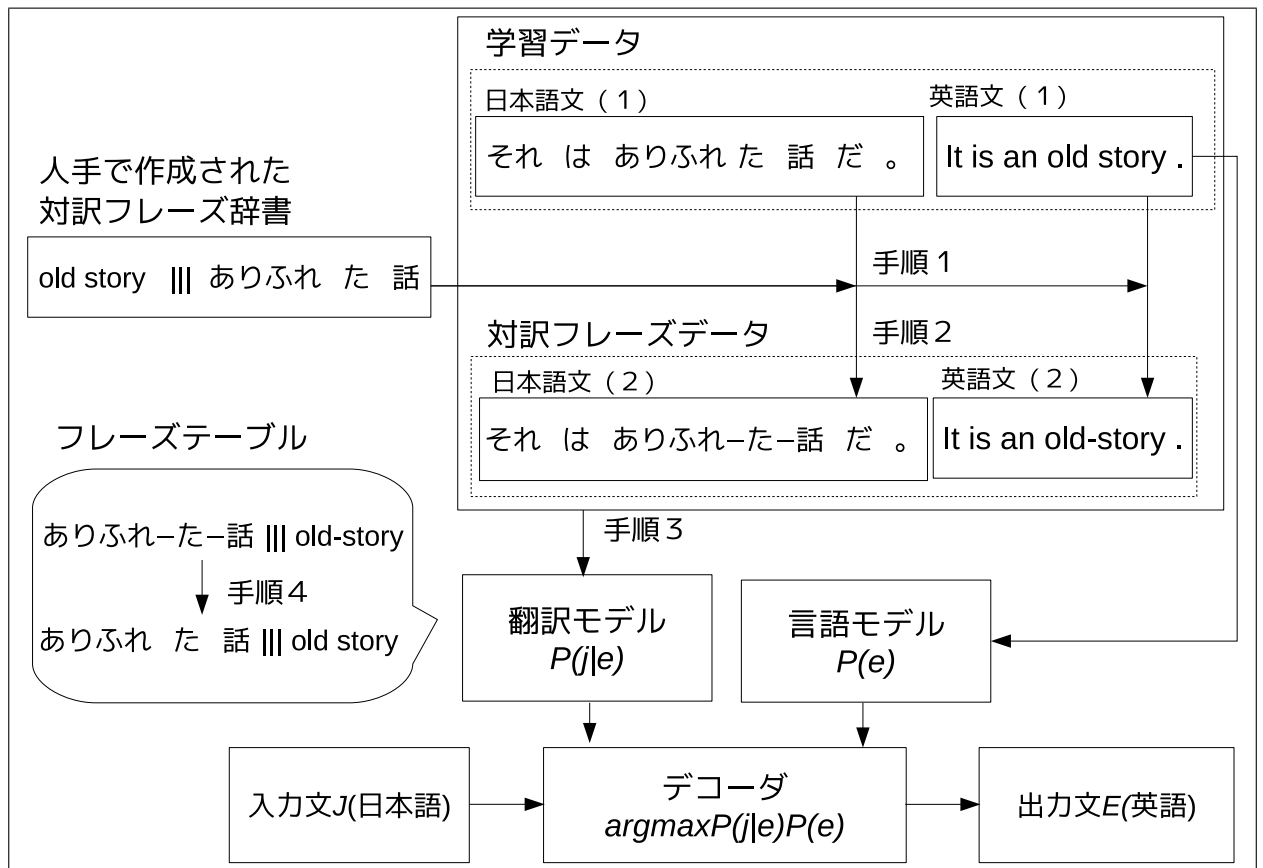


図 4.1 提案手法の流れ



手順2 学習データの一致した英語フレーズと日本語フレーズをそれぞれ一つのまとまりとするために，単語間のスペースを“-”に置き換えた対訳フレーズデータの日本語文(2)と英語文(2)を作成する．例を表4.2に示す．

表 4.2 対訳フレーズデータの例

対訳フレーズデータ 日本語文(2)	: それは ありふれ-た-話 だ 。
対訳フレーズデータ 英語文(2)	: It is an old-story .
対訳フレーズデータ 日本語文(2)	: 私-は 成人-式 に 出席 した 。
対訳フレーズデータ 英語文(2)	: I attended a coming-of-age-ceremony .

手順3 学習データと手順2で作成した対訳フレーズデータを用い，フレーズテーブルを作成する．

手順4 学習データのみから学習した言語モデルを用いるため，フレーズテーブルに含まれる“-”を取り除く．例を表4.3と表4.4に示す．

表 4.3 “-”を取り除く前のフレーズテーブル

それは     It is	
ありふれ-た-話 だ 。	an old-story .
ありふれ-た-話     old-story	
私-は     I	
成人-式     coming-of-age-ceremony	
出席 した     attended	

表 4.4 “-”を取り除いた後のフレーズテーブル

それは     It is	
ありふれた 話 だ 。	an old story .
ありふれた 話     old story	
私は     I	
成人 式     coming of age ceremony	
出席 した     attended	

手順5 手順4で作成したフレーズテーブルを用いて日英統計翻訳を行う．

## 第5章 実験環境

### 5.1 実験内容

本研究では、人手で作成された対訳フレーズ辞書として鳥バンク [1] と英辞郎 [3] の2種類を用いる。また、単文コーパスと重文複文コーパスに対してそれぞれ、日英統計翻訳と英日統計翻訳を行う。したがって、合計8種類の翻訳実験を行う。また、人手で作成された辞書から得た対訳フレーズデータを用いず、学習データのみから作成したフレーズテーブルを翻訳モデルとして行う統計翻訳をベースラインとする。

### 5.2 実験データ

単文の実験には辞書の例文より抽出した単文コーパス [10] 181,988 文からベースラインの学習データとして 100,000 文、テストデータとして 10,000 文用いる。また、重文複文の実験では重文複文コーパス [11] 121,719 文からベースラインの学習データとして 100,000 文、テストデータとして 10,000 文用いる。

単文、重文複文の両方において、提案手法の学習データはベースラインで用いた 100,000 文に 6.1 節で作成した対訳フレーズデータをそれぞれ加えたデータを学習データとする。

単文コーパスには日本語文は単文であるが英語文は重文複文である文も含まれる。また、重文複文コーパスには日本語文は重文複文であるが英語文は単文である文も含まれる。

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[12] を用いて形態素解析を行う。また、英語文に対して “tokenizer.perl[6]” を用いて分かち書きを行う。前処理を行った対訳文の例を表 5.1 に示す。

表 5.1 対訳文の例

石油の発見でその国は裕福になった。
The discovery of oil enriched the country .
梅雨が始まった。
The rainy season has set in .
彼は食料品店を営んでいる。
He runs a grocery store .

### 5.3 学習データの数

以下にベースラインと提案手法の言語モデルと翻訳モデルを作成する学習データの数を示す。

	ベースライン	提案手法
言語モデル	100,000 文対	100,000 文対
翻訳モデル	100,000 文対	100,000 文対 + 対訳フレーズデータの文数

対訳フレーズデータの数は 6.1 節で記述している。

### 5.4 翻訳モデルの学習

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学習には、多くの方法がある。本実験では、Moses の付録である “train-factored-phrase-model.perl” [6] を用いる。このプログラムは GIZA++ [5] を利用し、IBM モデルの計算を行う。本研究では、フレーズテーブルのヒューリスティックとして “grow-diag-final-and” を用いる。また、フレーズテーブルを作成する際、フレーズテーブル内の日本語と英語のフレーズ中の単語数の上限として、max-phrase-length が定義されている。たとえば、max-phrase-length の値が 20 の場合、日本語か英語のいずれかのフレーズ中の単語数が、21 以上のフレーズ対は作成されない。本研究で max-phrase-length の値として、はデフォルト値の 20 を用いる。

## 5.5 言語モデルの学習

言語モデルは,  $N$ -gram モデルを用いる.  $N$ -gram モデルの学習には “SRILM[13]” の ngram-count を用いる.  $N$ -gram モデルの次数は, 先行研究により 5-gram が有効であることがわかっている. そのため, 本研究でも, 5-gram の言語モデルを用いる. なお, スムージングに “-kndiscount” を用いる.

## 5.6 デコーダのパラメータ

デコーダは “Mises[6]” を使用する. Moses はいくつかのパラメータを設定することができる. パラメータの最適化には, Minimum Error Rate Training(MERT)[14] を用いることが一般的である. MERT は目的の評価関数 (通常は BLEU) を最大とするような翻訳結果が選ばれるようにパラメータを調節する. この際, development データと呼ばれる試し翻訳用のデータを与え, 各文について, 上位 100 文程度の翻訳候補の中で重みを変え, 各文について上位にくるようにパラメータを調節する. しかし, パラメータの最適化は development データに依存し, また, 多くの時間がかかるため, 本研究ではパラメータの最適化は行わない. クロスエントロピーを用いるため, “weight-t” は “0.5 0.0 0.5 0.0” とする. 翻訳時にフレーズの位置変化に柔軟に対応するため, “distortion weight” は 0.2 とする.

## 5.7 対訳フレーズ辞書

本実験では人手で作成された対訳フレーズ辞書として鳥バンクと英辞郎を用いる.

### 5.7.1 鳥バンク

鳥バンク [1] は自然言語処理のための言語知識ベースを収録したデータバンクであり, 日本語の重文と複文を対象とする「意味類型パターン辞書」が収録されている. 本実験では鳥バンクから抽出した 698,472 フレーズ対 [2] を用いる. 本実験では鳥バンクを用いて統計翻訳を行う実験を提案手法 (鳥バンク) とする. フレーズ対の例を表 5.2 に示す.

表 5.2 鳥バンクフレーズ対の例

コート の すそ	the edge of my coat
偉大 な 学者	become a great scholar
カメラ を 買う	buy a camera
朝晩 の ラッシュ 時 に 電車 に 乗る	take the train during the morning and evening rush hours

### 5.7.2 英辞郎

英辞郎 [3] は, EDP(Electronic Dictionary Project) がアップデートし続けている英和・和英辞書である. そのため, 英辞郎には通常英語辞書にない新しい単語や複雑な言い回しも含まれる. 英辞郎のデータには対訳フレーズ対の他に翻訳例や注釈であったり, 本来の文に出てこない “ ~ ” 等の記号が含まれる. 本実験では英辞郎のクリーニングを行い, 必要な英語と日本語のフレーズ対のみの形にした 1,366,575 フレーズ対を用いる. 英辞郎を用いて統計翻訳を行う実験を提案手法 (英辞郎) とよぶ. 表 5.3 にクリーニング後の英辞郎のフレーズ対の例を示す.

表 5.3 クリーニング後の英辞郎のフレーズ対の例

come out from	から 出 て くる
come out from	の 結果 として 生じる
obtain information on	に 関する 情報 を 得る
in view of the good quality and attractive designs	良い 品質 で 魅力的 な デザイン である こと を 考慮 すると

## 第6章 翻訳実験

### 6.1 対訳フレーズデータ

学習データの英語文と日本語文，人手で作成された辞書の英語フレーズと日本語フレーズのそれぞれをマッチングを行った．マッチした場合は，学習データ中のフレーズの単語間のスペースを“-”に置き換えた対訳フレーズデータを作成する．ベースラインで用いる学習データ 100,000 文対に対し，提案手法 (鳥バンク) で作成された対訳フレーズデータの数と提案手法 (英辞郎) で作成された対訳フレーズデータの数を表 6.1 に示す．

表 6.1 対訳フレーズデータの数

	単文	重文複文
提案手法 (鳥バンク)	97,649	99,902
提案手法 (英辞郎)	83,017	84,782

鳥バンクを用いて作成した対訳フレーズデータは，英辞郎を用いて作成した対訳フレーズデータより多くなった．これは鳥バンクの対訳フレーズ辞書が単文コーパス及び重文複文コーパスと分野が同じであるため，各コーパスにより多くマッチングしたためと考えている．

### 6.2 実験結果

#### 6.2.1 自動評価

日英統計翻訳の出力文の評価は自動評価法の BLEU, NIST, METEOR を用いる．また，英日統計翻訳の出力文の評価は自動評価法の BLEU, NIST を用いる．日英統計翻訳の結果を表 6.2 に示す．英日統計翻訳の結果を表 6.3 に示す．

表 6.2 自動評価結果 (日英翻訳)

	単文			重文複文		
	BLEU	NIST	METEOR	BLEU	NIST	METEOR
ベースライン	0.1091	3.9823	0.4597	0.0916	3.6270	0.2645
提案手法 (鳥バンク)	0.1068	3.8390	0.4566	0.0894	3.3828	0.2526
提案手法 (英辞郎)	0.1114	3.9891	0.4625	0.0944	3.6269	0.2667

表 6.3 自動評価結果 (英日翻訳)

	単文		重文複文	
	BLEU	NIST	BLEU	NIST
ベースライン	0.1519	4.1148	0.1132	3.4877
提案手法 (鳥バンク)	0.1522	4.0884	0.1102	3.3543
提案手法 (英辞郎)	0.1527	4.1090	0.1157	3.5117

結果より，提案手法 (英辞郎) はベースラインと比べて BLEU, METEOR のいずれの自動評価においてもスコアが向上し，提案手法の有効性を確認することができた．しかし，提案手法 (鳥バンク) は日英翻訳の単文の BLEU のみベースラインの評価値より優れているという結果になった．

## 6.2.2 人手評価

人手による対比較評価を 6.1 節の実験結果に対して行う．提案手法 (鳥バンク) の日英統計翻訳の結果を表 6.4 に，英日統計翻訳の結果を表 6.5 に示す．提案手法 (英辞郎) の日英統計翻訳結果を表 6.6 に，英日統計翻訳の結果を表 6.7 に示す．

表 6.4 人手評価結果 (鳥バンク 日英)

	ベースライン○	提案手法○	差無し	同一出力
単文	5	0	70	25
重文複文	4	2	90	4

表 6.5 人手評価結果 (鳥バンク 英日)

	ベースライン○	提案手法○	差無し	同一出力
単文	3	8	66	23
重文複文	2	4	88	6

表 6.6 人手評価結果 (英辞郎 日英)

	ベースライン○	提案手法○	差無し	同一出力
単文	8	1	52	39
重文複文	1	5	83	11

表 6.7 人手評価結果 (英辞郎 英日)

	ベースライン○	提案手法○	差無し	同一出力
単文	3	3	58	36
重文複文	4	2	73	21

人手翻訳結果において提案手法 (英辞郎) の日英統計翻訳はベースラインより優れているという結果になった。また、提案手法 (鳥バンク) の英日統計翻訳はベースラインより優れているという結果になった。

### 6.2.3 出力例

6.2.2 節の翻訳実験に対して出力例を示す。提案手法 (鳥バンク) 日英統計翻訳の単文の出力例を表 6.8 に、重文複文の出力例を表 6.9 に示す。提案手法 (英辞郎) 日英統計翻訳の単文の出力例を表 6.10 に、重文複文の出力例を表 6.11 に示す。提案手法 (鳥バンク) 英日統計翻訳の単文の出力例を表 6.12 に、重文複文の出力例を表 6.13 に示す。提案手法 (英辞郎) 英日統計翻訳の単文の出力例を表 6.14 に、重文複文の出力例を表 6.15 に示す。



表 6.8 提案手法 (鳥バンク) 日英統計翻訳結果 (単文)

ベースライン○の翻訳例	
入力文	金 1 0 万円 まさに 受け取り ました。
正解文	Received the sum of a hundred thousand yen .
ベースライン	I have just received the money one hundred thousand yen .
提案手法 (鳥バンク)	I received the money one hundred thousand yen .
差なしの翻訳例	
入力文	クラブ は 役員 を 選出 した。
正解文	The club elected their officers .
ベースライン	The club an elect .
提案手法 (鳥バンク)	The club an choose .

表 6.9 提案手法 (鳥バンク) 日英統計翻訳結果 (重文複文)

ベースライン○の翻訳例	
入力文	都合 が つき しい 払い いただきます。
正解文	You shall pay me at your convenience
ベースライン	Please pay me at your earliest convenience .
提案手法 (鳥バンク)	me at your earliest convenience .
提案手法○の翻訳例	
入力文	水泳 大会 で 優勝 して みせると 決心 していた。
正解文	He was determined to be top dog at the swimming meet .
ベースライン	Swimming championship to show .
提案手法 (鳥バンク)	Swimming championship would make up his mind .
差なしの翻訳例	
入力文	僕 は 酒 を 口 に した ことが 無い。
正解文	I never tasted strong drink in my life .
ベースライン	I drink in his mouth .
提案手法 (鳥バンク)	Wine has to say .

表 6.10 提案手法 (鳥バンク) 英日統計翻訳結果 (単文)

ベースライン○の翻訳例	
入力文	This town was not prepared for an earthquake .
正解文	この町は地震に対して無防備だった。
ベースライン	この町は地震の用意しなかった。
提案手法 (鳥バンク)	この町は地震でいた。
提案手法○の翻訳例	
入力文	I shall never forget your kindness .
正解文	ご親切は決して忘れません。
ベースライン	私は後々まで忘れません。
提案手法 (鳥バンク)	ご恩は決して忘れません。
差なしの翻訳例	
入力文	The cork popped .
正解文	コルク栓がポンと音をたてた。
ベースライン	このコルクた。
提案手法 (鳥バンク)	コルク栓た。

表 6.11 提案手法 (鳥バンク) 英日統計翻訳結果 (重文複文)

ベースライン○の翻訳例	
入力文	Ensure that the disk is not exposed to dust , dirt , fingerprints , etc .
正解文	ディスクにほこりや指紋、汚れが付かないように注意してください。
ベースライン	それは、ディスクは、指紋が汚れないようにしてください。
提案手法 (鳥バンク)	そのことはありませんが、など、土に指紋がない。
提案手法○の翻訳例	
入力文	When the window is opened , the mountain could be seen clearly .
正解文	窓を開けると山がよく見えた。
ベースライン	窓を開けるとがはっきり見えた。
提案手法 (鳥バンク)	窓を開けて、山がはっきり見えた。
差なしの翻訳例	
入力文	The school was established by the union of five villages .
正解文	この学校は五か村連合してできた。
ベースライン	学校で定められたのは組合がある。
提案手法 (鳥バンク)	学校で5人の村をした。

表 6.12 提案手法 (英辞郎) 日英統計翻訳結果 (単文)

ベースライン○の翻訳例	
入力文	この おかげ で 私 は いっそう 思い どり 選択 できる 。
正解文	This will afford me wider freedom of choice .
ベースライン	This helped me with the choice .
提案手法 (英辞郎)	This to me with his choice .
提案手法○の翻訳例	
入力文	急い で 朝食 を 飲み こんだ 。
正解文	I gulped down a quick breakfast .
ベースライン	He hurried 飲み こん breakfast .
提案手法 (英辞郎)	I swallowed up the breakfast .
差なしの翻訳例	
入力文	物事 は そう 理論 どり には いか ない 。
正解文	We can hardly settle things by theory alone .
ベースライン	Things is according to theory .
提案手法 (英辞郎)	Everything is according to theory .

表 6.13 提案手法 (英辞郎) 日英統計翻訳結果 (重文複文)

ベースライン○の翻訳例	
入力文	彼 は 医者 の 止める の も 聴か ず に 酒 を 止め ぬ 。
正解文	He keeps on drinking in defiance of his doctor's warning .
ベースライン	He keeps on drinking in defiance of his doctor's warning .
提案手法 (英辞郎)	He keeps on drinking in defiance of the .
提案手法○の翻訳例	
入力文	結果 は 期待 した より は よか った 。
正解文	The results were better than had been looked for .
ベースライン	The than I had expected .
提案手法 (英辞郎)	The result is better than I had expected .
差なしの翻訳例	
入力文	欲望 は 起こ ったり 消え たり する もの です 。
正解文	Desire comes and goes .
ベースライン	The and went to happen .
提案手法 (英辞郎)	is what and went down .

表 6.14 提案手法 (英辞郎) 英日統計翻訳結果 (単文)

ベースライン○の翻訳例	
入力文	The union will join in the strike .
正解文	組合はストライキに参加する。
ベースライン	組合はストライキ参加する。
提案手法 (英辞郎)	組合は後日参加している。
提案手法○の翻訳例	
入力文	There was a moan of protest from the students .
正解文	生徒たちから抗議のうなり声があがった。
ベースライン	から抗議のうなり声があった。
提案手法 (英辞郎)	生徒から抗議のうなり声があった。
差なしの翻訳例	
入力文	He has thrown in his lot with us .
正解文	彼は我々と運命を共にした。
ベースライン	彼はになった。
提案手法 (英辞郎)	彼は多くのである。

表 6.15 提案手法 (英辞郎) 英日統計翻訳結果 (重文複文)

ベースライン○の翻訳例	
入力文	His vocabulary showed a fairly wide range .
正解文	彼の語彙はかなり広い範囲に及んでいた。
ベースライン	彼の語彙をかなり広い範囲にわたっている。
提案手法 (英辞郎)	彼は、語彙を示した。
提案手法○の翻訳例	
入力文	You should master Chinese while you are in college .
正解文	大学に行っている間にぜひ中国語を身につけなさい。
ベースライン	中国語をマスターしている間にしている。
提案手法 (英辞郎)	大学に行っている間に中国語をマスターしなさい。
差なしの翻訳例	
入力文	Do not use this tool so roughly !
正解文	この道具をそんなに乱暴に扱っては困る。
ベースライン	この工具を使用してはいけなから困る。
提案手法 (英辞郎)	この道具は、使用しないでください。

## 第7章 考察

### 7.1 提案手法(鳥バンク)と提案手法(英辞郎)

人手評価において提案手法(鳥バンク)の英日翻訳結果はベースラインより優れていた。一方、自動評価においてスコアが低かった。

人手評価において提案手法(鳥バンク)が優れていた原因として、鳥バンクの対訳フレーズ対は単文コーパス及び重文複文コーパスと分野が同じであるため対訳フレーズデータを多く作成できた。その結果、信頼性の高いフレーズテーブルを自動的に作成することができたためであると考えている。

また、人手評価において提案手法(英辞郎)はベースラインの翻訳品質とさほど変わらなかった。しかし、自動評価においてスコアが向上した。この原因として、人手で作られた対訳フレーズ辞書を用いて作成した信頼性の高いフレーズテーブルは、ベースラインで作成したフレーズテーブルに既に存在するケースが多く、そのため殆どの人手評価で提案手法の翻訳品質が向上しなかったと考えている。

### 7.2 人手評価と自動評価

提案手法(英辞郎)は人手評価においてベースラインの翻訳品質と比べてさほど変わらなかったが、自動評価において翻訳精度の向上が確認できた。一方、提案手法(鳥バンク)は人手評価において日英統計翻訳はベースラインより優れていたが、自動評価においてベースラインよりスコアが低かった。自動評価で優れていると判断した実験結果は人手評価において優れている場合もあった。

今後自動評価の問題点もふまえて提案手法(英辞郎)と提案手法(鳥バンク)の更なる調査を考えている。

## 7.3 対訳フレーズデータの効果の分析

6.2.2 節で人手評価を行った結果について分析を行った。表 6.4 から表 6.7 の結果において、ベースラインが優れていた文は 30 文，提案手法が優れていた文は 25 文であった。それぞれの結果に対して，提案手法の訳質が低下した原因と向上した原因ごとに分類した。ベースラインが優れていた文に対しては，ベースラインより訳質が低下した原因として，“文質の低下 (フレーズテーブルの選択ミス)”，“主語の消失”，“動詞の消失” に分類した。提案手法が優れていた文として，ベースラインより翻訳品質が向上した原因として，“文質の向上”，“動詞の出現”，“未知語の消失”，“主語の出現” に分類した。分類結果を表 7.1 と表 7.2 に示す。

表 7.1 ベースラインより翻訳品質が低下した原因

文質の低下 (フレーズテーブルの選択ミス)	23/30
主語の消失	5/30
動詞の消失	2/30

表 7.2 ベースラインより翻訳品質が向上した原因

文質の向上 (フレーズテーブルの良い選択)	19/25
動詞の出現	3/25
未知語の消失	2/25
主語の出現	1/25

### 7.3.1 文質の低下によってベースラインより訳質が低下した場合

表 7.1 においての，“文質の低下” の例を表 7.3 に示す。

表 7.3 “文質の低下” の例

入力文	台風で彼はやむをえず週末の予定を変更した。
正解文	The typhoon forced him to change his weekend plan .
ベースライン	He had no choice but to change the weekend by the typhoon .
提案手法	He changed the weekend to undertake the typhoon .

表 7.3 の例では，“やむをえず”に対するフレーズテーブルとして，ベースラインは“やむをえず ||| had no choice but ”が用いられたのに対し，提案手法は“やむをえず ||| undertake”が用いられた。

ベースラインのフレーズテーブルに，提案手法で作成された“やむをえず ||| undertake”のフレーズテーブルは存在しなかった。また，提案手法のフレーズテーブルに，ベースラインで作成された“やむをえず ||| had no choice but ”のフレーズテーブルは存在しなかった。ベースラインにある句対応の正しいフレーズテーブルが作成されず，ベースラインにない句対応が不適切なフレーズテーブルを用いて翻訳されたため，翻訳品質が低下した。

### 7.3.2 主語の消失によってベースラインより訳質が低下した場合

表 7.1 においての，“主語の消失”の例を表 7.4 に示す。

表 7.4 “主語の消失”の例

入力文	He was walking down the street .
正解文	彼は通りを歩いていた。
ベースライン	彼は通りを歩いていた。
提案手法	通りを歩いていた。

表 7.4 の例では，“歩いていた”に対するフレーズテーブルとして，ベースラインは“was walking ||| 歩いていた ”が用いられたのに対し，提案手法は“He was walking ||| 歩いていた”が用いられた。提案手法では，ベースラインのように“He”に対する日本語フレーズとして“彼は”が翻訳されず，翻訳品質が低下した。

ベースラインのフレーズテーブルに，提案手法で作成された“He was walking ||| 歩いていた”のフレーズテーブルは存在しなかった。また，提案手法のフレーズテーブルに，ベースラインで作成された“was walking ||| 歩いていた ”のフレーズテーブルは存在しなかった。

### 7.3.3 動詞の消失によってベースラインより訳質が低下した場合

表 7.1 においての，“動詞の消失”の例を表 7.5 に示す。



表 7.5 “動詞の消失” の例

入力文	それは 大体 同じ 意味 に 用い られる 。
正解文	It is used in much the same sense .
ベースライン	It is used for almost the same meaning .
提案手法	That's all of the same meaning .

表 7.5 の例では，“用い られる 。” に対するフレーズテーブルとして，ベースラインは“用い られる 。

ベースラインのフレーズテーブルに，提案手法で作成された“用い られる 。

### 7.3.4 文質の向上によってベースラインより訳質が向上した場合

表 7.2 においての，“文質の向上” の例を表 7.6 に示す。

表 7.6 “文質の向上” の例

入力文	I moved to the suburbs last year .
正解文	去年、 郊外 に 引っ越 しまし た 。
ベースライン	私 は 郊外 へ 移 転 し た 。
提案手法	私 は 去 年 郊 外 へ 移 動 し た 。

表 7.6 の例では，ベースラインの出力は，“to the suburbs ||| 郊外 へ” と “ last year . ||| 。” のフレーズテーブルを用いて翻訳しているのに対し，提案手法は“the suburbs last year ||| は 去年 郊外 ” のフレーズテーブルが用いられた。

ベースラインのフレーズテーブルに，提案手法で作成された“the suburbs last year ||| は 去年 郊外 ” のフレーズテーブルは存在しなかった。ベースラインにない句対応の正しいフレーズテーブルが作成され，それを用いて翻訳したため，翻訳品質が向上した。

### 7.3.5 動詞の出現によってベースラインより訳質が向上した場合

表 7.2 においての，“動詞の出現” の例を表 7.7 に示す。

表 7.7 “動詞の出現” の例

入力文	It will depend on the results of all these changes .
正解文	その出来にもかかわってくる。
ベースライン	それには、これらの結果だ。
提案手法	それは、これらの結果によるだろう。

表 7.7 の例では，“will depend on” に対するフレーズテーブルとして，ベースラインは“will depend on ||| に ” が用いられたのに対し，提案手法は“will depend ||| よる だろう ” が用いられた。

提案手法のフレーズテーブルに，ベースラインで作成された“will depend on ||| に ” のフレーズテーブルは存在しなかった。ベースラインにあった句対応の不適切なフレーズテーブルが作成されず，句対応の正しいフレーズテーブルを用いて翻訳したため，翻訳品質が向上した。

### 7.3.6 未知語の消失によってベースラインより訳質が向上した場合

表 7.2 においての，“未知語の消失” の例を表 7.8 に示す。

表 7.8 “未知語の消失” の例

入力文	急いで朝食を飲みこんだ。
正解文	I gulped down a quick breakfast .
ベースライン	He hurried 飲みこん breakfast .
提案手法	I swallowed up the breakfast .

表 7.8 の例では，ベースラインは“飲みこん” に対するフレーズテーブルが存在しなかったのに対し，提案手法は“飲みこんだ ||| swallowed up” が用いられた。

ベースラインにない句対応の正しいフレーズテーブルが作成され，それを用いて翻訳したため，翻訳品質が向上した。

### 7.3.7 主語の出現によってベースラインより訳質が向上した場合

表 7.2 においての，“主語の出現”の例を表 7.9 に示す。

表 7.9 “主語の出現”の例

入力文	There was a moan of protest from the students .
正解文	生徒たちから抗議のうなり声があがった。
ベースライン	から抗議のうなり声があった。
提案手法	生徒から抗議のうなり声があった。

表 7.9 の例では，“the students”に対するフレーズテーブルとして，ベースラインは“the students |||。”が用いられたのに対し，提案手法は“from the students ||| 生徒から”が用いられた。

提案手法のフレーズテーブルに，ベースラインで作成された“the students |||。”のフレーズテーブルは存在しなかった。ベースラインにあった句対応の不適切なフレーズテーブルが作成されず，句対応の正しいフレーズテーブルを用いて翻訳したため，翻訳品質が向上した。

## 7.4 対訳フレーズデータのみを用いた翻訳実験

5.4 節に示しているように，本研究において，提案手法はベースラインの学習データに対訳フレーズデータを加えたデータを学習データとしている。ほぼ提案手法(英辞郎)，提案手法(鳥バンク)の対訳フレーズデータはほぼ 100,000 文近く作成できた。そこで，学習データとして対訳フレーズデータのみを用い，統計翻訳を行った場合の翻訳精度の調査をする。ここで，人手で作成された辞書として鳥バンクを用いて作成した対訳フレーズデータのみを学習データとする翻訳実験を“対訳フレーズデータ実験(鳥バンク)”と呼ぶ。また，人手で作成された辞書として英辞郎を用いて作成した対訳フレーズデータのみを学習データとする翻訳実験を“対訳フレーズデータ実験(英辞郎)”と呼ぶ。

### 7.4.1 自動評価結果

対訳フレーズデータ実験に対して日英統計翻訳，英日統計翻訳を行った翻訳結果に対して自動評価を行った。日英統計翻訳の結果を表 7.10 に，英日統計翻訳の結果を表 7.11 に示す。

表 7.10 自動評価結果 (日英翻訳)

	単文			重文複文		
	BLEU	NIST	METEOR	BLEU	NIST	METEOR
ベースライン	0.1091	3.9823	0.4597	0.0916	3.6270	0.2645
提案手法 (鳥バンク)	0.1068	3.8390	0.4566	0.0894	3.3828	0.2526
対訳フレーズデータ実験 (鳥バンク)	0.1015	3.7774	0.4453	0.0761	3.2517	0.2298
提案手法 (英辞郎)	0.1114	3.9891	0.4625	0.0944	3.6269	0.2667
対訳フレーズデータ実験 (英辞郎)	0.1059	3.9431	0.4558	0.0851	3.5189	0.2557

表 7.11 自動評価結果 (英日翻訳)

	単文		重文複文	
	BLEU	NIST	BLEU	NIST
ベースライン	0.1519	4.1148	0.1132	3.4877
提案手法 (鳥バンク)	0.1522	4.0884	0.1102	3.3543
対訳フレーズデータ実験 (鳥バンク)	0.1441	3.9412	0.0944	3.0728
提案手法 (英辞郎)	0.1527	4.1090	0.1157	3.5117
対訳フレーズデータ実験 (英辞郎)	0.1469	4.0345	0.1059	3.3720

結果から、対訳フレーズデータ実験の自動評価結果はベースラインより劣っているという結果になった。この原因は、対訳フレーズデータのみでは学習データ不足であるため、評価値がベースラインより下がったためだと考えられる。

## 第8章 おわりに

本研究では、人手で作成された辞書を用い、学習データ中のフレーズを一つのまとまりとした対訳フレーズデータを作り、人手で作成されたフレーズを含むフレーズテーブルを作成した。人手で作成された辞書として英辞郎を用いた提案手法(英辞郎)と鳥バンクを用いた提案手法(鳥バンク)の翻訳実験を行った。また、単文コーパスと重文複文コーパスを用い、それぞれ日英統計翻訳と英日統計翻訳を行った。したがって、合計8種類の翻訳実験を行った。

その結果、人手評価において提案手法(英辞郎)の日英翻訳の重文複文と提案手法(鳥バンク)の単文と重文複文の英日翻訳の評価結果はベースラインより優れており、それ以外の翻訳実験はベースラインより劣っているという結果になった。

一方、自動評価において提案手法(英辞郎)の全ての実験においてはBLEU, METEORの値がベースラインより向上した。提案手法(鳥バンク)はほとんどの自動評価においてベースラインよりスコアが低くなった。

これは鳥バンクの対訳フレーズ辞書がフレーズ対が単文コーパス及び重文複文コーパスと分野が同じであるため、効果的な対訳フレーズデータを作成し、良い翻訳結果が得ることができたからと考えている。

今後、提案手法(英辞郎)と提案手法(鳥バンク)の更なる調査を行っていきたい。

# 謝辞

最後に、1年間に渡りご指導いただきました鳥取大学工学部知能情報工学科計算機工学講座C研究室の村田真樹教授，村上仁一准教授，徳久雅人講師そして計算機工学講座C研究室の方々に厚く御礼申し上げます。また，参考にさせていただいた論文の著者の方々に対して深く感謝申し上げます。

## 参考文献

- [1] 鳥バンク : <http://unicorn.iike.tottori-u.ac.jp/toribank/>
- [2] 鏡味良太, 村上仁一, 徳久雅人, 池原悟, “統計翻訳における人手で作成された大規模フレーズテーブルの効果”, 言語処理学会第 14 回年次大会, pp.224-227, 2008.
- [3] 英辞郎 : <http://www.alc.co.jp/>
- [4] 東江恵介, 村上仁一, 徳久雅人, 池原悟, “日英統計翻訳における英辞郎の効果” , 言語処理学会第 16 回年次大会発表論文集, pp.641-644, 2010.
- [5] GIZA++ : <http://www.fjoch.com/GIZA++>
- [6] Philipp Koehn, Marcello Federico, Brooke Cowan, Richard Zens, Chris Dyer, Ondřej Bojar, Alexandra Constantin, Evan Herbst, “Open Source Toolkit for Statistical Machine Translation”, Proceedings of the ACL 2007 Demo and Poster Sessions, pp.177-180, June 2007.
- [7] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu, “Bleu: a Method for Automatic Evaluation of Machine Translation”, Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, pp.311-318, July 2002.
- [8] George Doddington, “Automatic Evaluation of Machine Translation Quality Using n-gram Co-Occurrence Statistics” , In Proc. ARPA Workshop on Human Language Technology, 2002.  
<http://www.itl.nist.gov/>
- [9] Lavie, Alon and Denkowski, Michael. “ The METEOR Metric for Automatic Evaluation of Machine Translation” , 2009.

- [10] 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.
- [11] 村上仁一, 池原悟, 徳久雅人, “日本語英語の文対応の対訳データベースの作成”, 「言語, 認識, 表現」第 7 回年次研究会, 2002.
- [12] MeCab : <http://mecab.sourceforge.net/>
- [13] SRILM : The SRI Language Modeling Toolkit,  
<http://www.speech.sri.com/projects/srilm>
- [14] Franz Josef Och, “Minimum Error Rate Training in Statistical Machine Translation”, Association for Computational Linguistics 2003, pp.160-167, 2003.