

# 概要

近年，機械翻訳の分野で統計翻訳が注目されている．統計翻訳においては，対訳文などの学習データの量が多ければ多いほど翻訳精度が高くなる．しかし，対訳文の収集のコストは大きい．

この問題に対し，対訳文と対訳句を学習データに追加する手法の研究が行われている．そして，この研究において，自動評価値が向上したとの報告がある．この研究は，セルビア語英語間，スペイン語英語間 [1] そしてブルトン語フランス語間 [2] における統計翻訳において行われている．また日野らの研究においても，日本語英語間の統計翻訳における対訳文と対訳句の学習データへの追加の効果が報告されている [3] ．

しかし，これらの先行研究の実験環境においては，対訳文とは別の新たな対訳句を学習データに追加したものを学習に用いている．この場合，翻訳精度向上の理由として考えられるのは，新たな対訳句の学習により正しい対訳句の対応が強調されたことと，新たな対訳句の学習により単に対訳データの語彙量が増加したことの，大きく2つの要因が考えられる．そのため，実際に正しい対訳句の対応が強調されたことが，どの程度，翻訳精度の向上に効果があったのかはわからない．また，新たな対訳句の学習には，対訳句の収集にコストがかかるため，学習に用いるデータの収集の面でも問題が残る．

そこで，本研究では，対訳データの語彙量の増加が生じない実験環境においても，正しい対訳句の対応が強調されていれば翻訳精度が向上すると予想する．そして，対訳データの語彙量の増加が生じない実験環境において，対訳文と対訳句を学習データに追加したものをを用いた統計翻訳の有効性の調査を行う．具体的に，対訳データの語彙量の増加が生じない実験環境とは，学習に用いる対訳文から対訳句を抽出することにより実現する．この実験環境において，正しい対訳句の対応が強調されたことが，どの程度，翻訳精度の向上に効果があったのか調査する．また，この実験環境で翻訳精度が向上すれば，あらかじめ用意する対訳データとしては対訳文のみを用意すればよい．対訳データ収集のコストを増やすことなく，翻訳精度を向上させることができる．さらに，対訳句の抽出方法として，手動抽出と自動抽出の2通りの抽出方法を試し，対訳句の抽出精度やその対訳句を用いた場合の統計翻訳の翻訳精度の調査を行う．日英対訳文は重文複文 [4] を用いる．日英対訳句として日英対訳文から抽出した対訳句を用いる．

その結果，自動評価と人手評価のどちらにおいても提案手法の翻訳精度は向上した．よって，対訳データの語彙量の増加が生じない実験環境においても，正しい対訳句の対応が強調されていれば翻訳精度が向上することがわかり．対訳文と対訳句を学習データ

へ追加した対訳データを用いた統計翻訳の有効性が示された。

# 目次

第1章	はじめに	1
第2章	統計翻訳システム	3
2.1	概要	3
2.2	単語に基づく統計翻訳	3
2.3	句に基づく統計翻訳	5
2.3.1	GIZA++	6
2.3.2	言語モデル	6
2.3.3	翻訳モデル	6
2.3.4	フレーズテーブル作成法	7
2.3.5	パラメータチューニング	11
2.4	評価方法	11
2.4.1	自動評価法	11
2.4.1.1	BLEU	11
2.4.1.2	METEOR	12
2.4.1.3	RIBES	12
2.4.2	人手評価	13
第3章	提案手法	15
3.1	日英統計翻訳の流れ	16
3.2	日英対訳句抽出システム	17
3.2.1	手動抽出	17
3.2.2	自動抽出	18
3.2.2.1	日英対訳単語の作成	18
3.2.2.2	日英対訳文パターンの作成	19
3.2.2.3	日英対訳句の抽出	20
3.3	提案手法の設定	21

第4章	実験環境	22
4.1	実験データ	22
4.2	実験ツール	22
4.3	評価方法	23
第5章	日英対訳句の抽出	24
5.1	手動抽出	24
5.2	自動抽出	25
第6章	翻訳実験	27
6.1	実験手法	27
6.2	自動評価結果	28
6.3	人手評価結果	28
第7章	自動抽出の改良に伴う自動手法の有効性の調査	36
7.1	単語翻訳確率の閾値の変更	36
7.1.1	対訳句の自動抽出	36
7.1.2	自動評価結果	37
7.2	頻度情報の活用	38
7.2.1	対訳句の自動抽出	38
7.2.2	自動評価結果	39
第8章	考察	40
8.1	精度向上の原因分析	40
8.2	自動抽出と手動抽出の比較	41
8.3	自動抽出の改良に伴う自動手法の有効性の考察	41
第9章	おわりに	42

# 目 次

2.1	日英統計翻訳の流れ	5
3.1	日英統計翻訳の流れ	16
3.2	日英対訳単語作成の例	19
3.3	日英対訳文パターン作成の例	20
3.4	日英対訳句抽出の例	21

# 表 目 次

2.1	日英方向の単語対応 . . . . .	3
2.2	英日方向の単語対応 . . . . .	4
2.3	日英方向の単語対応 . . . . .	7
2.4	英日方向の単語対応 . . . . .	7
2.5	intersection の例 . . . . .	8
2.6	union の例 . . . . .	8
2.7	grow の例 . . . . .	9
2.8	grow-diag の例 . . . . .	9
2.9	grow-diag-final の例 . . . . .	10
2.10	grow-diag-final-and の例 . . . . .	10
2.11	対比較例評価出力例 . . . . .	14
2.12	対比較評価基準 . . . . .	14
2.13	手法 A VS 手法 B . . . . .	14
3.1	日英対訳文の例 . . . . .	17
3.2	日英対訳文パターンの例 . . . . .	17
3.3	日英対訳文と日英対訳文パターンの比較により抽出された日英対訳句の例	18
4.1	日英対訳文の内訳 . . . . .	22
4.2	日英対訳文の例 . . . . .	22
5.1	手動抽出した日英対訳句の内訳 . . . . .	24
5.2	手動抽出した日英対訳句の例 . . . . .	24
5.3	手動抽出した日英対訳句の評価結果と評価例 . . . . .	25
5.4	自動抽出した日英対訳句の内訳 . . . . .	26
5.5	自動抽出した日英対訳句の評価結果と評価例 . . . . .	26
6.1	自動評価結果 . . . . .	28

6.2	baseline 手法 VS 手動手法	29
6.3	baseline 手法 VS 自動手法	29
6.4	手動手法 VS 自動手法	29
6.5	baseline 手法 > 手動手法	30
6.6	baseline 手法 < 手動手法	31
6.7	baseline 手法 $\approx$ 手動手法	31
6.8	baseline 手法 > 自動手法	32
6.9	baseline 手法 < 自動手法	33
6.10	baseline 手法 $\approx$ 自動手法	33
6.11	手動手法 > 自動手法	34
6.12	手動手法 < 自動手法	35
6.13	手動手法 $\approx$ 自動手法	35
7.1	閾値 ( $\alpha = 0.10$ ) で自動抽出した日英対訳句の内訳	36
7.2	閾値 ( $\alpha = 0.10$ ) で自動抽出した日英対訳句の評価結果と評価例	37
7.3	閾値 ( $\alpha = 0.20$ ) で自動抽出した日英対訳句の内訳	37
7.4	閾値 ( $\alpha = 0.20$ ) で自動抽出した日英対訳句の評価結果と評価例	37
7.5	自動評価結果 (%)	38
7.6	頻度情報を考慮して自動抽出した日英対訳句の内訳	38
7.7	頻度情報を考慮して自動抽出した日英対訳句の評価結果と評価例	39
7.8	自動評価結果 (%)	39
8.1	フレーズテーブル	40

# 第1章 はじめに

近年，機械翻訳の分野で統計翻訳が注目されている．統計翻訳においては，対訳文などの学習データの量が多ければ多いほど翻訳精度が高くなる．しかし，対訳文の収集のコストは大きい．

この問題に対し，対訳文と対訳句を学習データに追加する手法の研究が行われている．そして，この研究において，自動評価値が向上したとの報告がある．この研究は，セルビア語英語間，スペイン語英語間 [1] そしてブルトン語フランス語間 [2] における統計翻訳において行われている．また日野らの研究においても，日本語英語間の統計翻訳における対訳文と対訳句の学習データへの追加の効果が報告されている [3] ．

しかし，これらの先行研究の実験環境においては，対訳文とは別の新たな対訳句を学習データに追加したものを学習に用いている．この場合，翻訳精度向上の理由として考えられるのは，新たな対訳句の学習により正しい対訳句の対応が強調されたことと，新たな対訳句の学習により単に対訳データの語彙量が増加したことの，大きく2つの要因が考えられる．そのため，実際に正しい対訳句の対応が強調されたことが，どの程度，翻訳精度の向上に効果があったのかはわからない．また，新たな対訳句の学習には，対訳句の収集にコストがかかるため，学習に用いるデータの収集の面でも問題が残る．

そこで，本研究では，対訳データの語彙量の増加が生じない実験環境においても，正しい対訳句の対応が強調されていれば翻訳精度が向上すると予想する．そして，対訳データの語彙量の増加が生じない実験環境において，対訳文と対訳句を学習データに追加したものをを用いた統計翻訳の有効性の調査を行う．具体的に，対訳データの語彙量の増加が生じない実験環境とは，学習に用いる対訳文から対訳句を抽出することにより実現する．この実験環境において，正しい対訳句の対応が強調されたことが，どの程度，翻訳精度の向上に効果があったのか調査する．また，この実験環境で翻訳精度が向上すれば，あらかじめ用意する対訳データとしては対訳文のみを用意すればよいので，対訳データ収集のコストを増やすことなく，翻訳精度を向上させることができる．さらに，対訳句の抽出方法として，手動抽出と自動抽出の2通りの抽出方法を試し，対訳句の抽出精度やその対訳句を用いた場合の統計翻訳の翻訳精度の調査を行う．日英対訳文は重文複文



[4] を用いる．日英対訳句として日英対訳文から抽出した対訳句を用いる．

その結果，自動評価と人手評価のどちらにおいても提案手法の翻訳精度は向上した．よって，対訳データの語彙量の増加が生じない実験環境においても，正しい対訳句の対応が強調されていれば翻訳精度が向上することがわかり．対訳文と対訳句を学習データへ追加した対訳データを用いた統計翻訳の有効性が示された．

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，評価方法について述べる．3章では，本研究の提案手法について述べる．3.2章では，対訳句の抽出システムの概要を示し，各抽出方法について述べる．4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，8章では，考察を行う．最後に，9章で結論を述べ，まとめる．

## 第2章 統計翻訳システム

### 2.1 概要

統計翻訳とは、機械翻訳手法の一種である。原言語と目的言語の対訳文を大量に収集した対訳データにより、自動的に翻訳規則を獲得し翻訳を行う。

統計翻訳には“単語に基づく統計翻訳”と“句に基づく統計翻訳”があり、初期の統計翻訳では単語に基づく統計翻訳が用いられていた。しかし近年、句に基づく統計翻訳が提案され、単語に基づく統計翻訳に比べて翻訳精度が高いことがわかった。このため現在は句に基づく統計翻訳が主流となっている。

### 2.2 単語に基づく統計翻訳

単語に基づく統計翻訳は単語対応の翻訳モデルを用いている。例として、ある日本語文を英語文に翻訳する場合を考える。日本語単語を英語単語に翻訳し、日本語文の語順と同じ並びで英語単語を並べて翻訳する。単語に基づく統計翻訳は単語対応の確率を得る IBM 翻訳モデル [5] が用いられている。以下に、IBM 翻訳モデルを用いて得た英日方向における単語対応の例と、日英方向における単語対応の例を示す。また、 は単語が対応した箇所を示す。

表 2.1: 日英方向の単語対応

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

表 2.2: 英日方向の単語対応

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

表 2.1 は全ての単語に対して対応がとれている．一方で，表 2.2 は日本語単語“は”と“に”に対応する英語単語が存在しない．単語に基づく統計翻訳は対応する単語が存在しない場合，何も無い状態から単語の発生確率を計算する．このため単語翻訳確率の信頼性が問題となっている．よって現在は句単位の対応を用いた統計翻訳(句に基づく統計翻訳)が行われている．

## 2.3 句に基づく統計翻訳

句に基づく統計翻訳は句対応の翻訳モデルを用いる。原言語文を目的言語文に翻訳する場合に、隣接する複数の単語 (フレーズ) を用いて翻訳を行う方法である。本研究では日英方向の翻訳を行うため、日英統計翻訳を説明する。日英統計翻訳システムの流れを図 2.1 に示す。

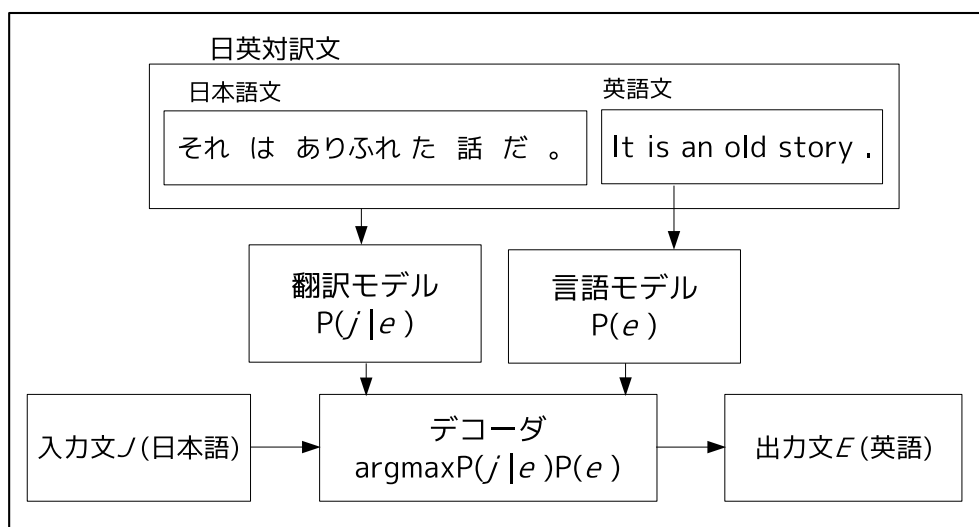


図 2.1: 日英統計翻訳の流れ

日英統計翻訳は、日本語入力文  $j$  が与えられた場合に、翻訳モデルと言語モデルの組み合わせの中から確率が最大となる英語翻訳文  $E$  を探索することで翻訳を行う。以下にその基本モデルを示す。

$$E = \operatorname{argmax}_j P(e|j) \quad (2.1)$$

$$\simeq \operatorname{argmax}_e P(j|e)P(e) \quad (2.2)$$

ここで  $P(j|e)$  は翻訳モデル、 $P(e)$  は言語モデルを示す。 $P(e)$  が単語であれば“単語に基づく統計翻訳”のモデル、 $P(e)$  が句であれば、“句に基づく統計翻訳”のモデルとなる。

また、学習データとは対訳文 (日本語文と英語文の対) を大量に用意したものである。学習データに含まれる各々のデータから、翻訳モデルと言語モデルを学習する。

### 2.3.1 GIZA++

GIZA++ [6] とは、統計翻訳で用いることを前提に作られたツールである。IBM 翻訳モデルを用いて、対訳文 (原言語文と目的言語文の対) から対訳単語と単語翻訳確率を自動的に得る。

### 2.3.2 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統計翻訳の場合、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。

### 2.3.3 翻訳モデル

翻訳モデルとは、膨大な量の対訳データを用いて日本語のフレーズが英語のフレーズへ確率的に翻訳を行うためのモデルである。この翻訳モデルはフレーズテーブルで管理されている。以下にフレーズテーブルの例を示す。

#### フレーズテーブルの例

その花		The flower		0.428571	0.0889909	0.428571	0.0907911	2.718
今晚のコンサートは		Tonight's concert is		0.5	0.000223681	0.5	0.0124601	2.718

左から日本語フレーズ、英語フレーズ、フレーズの英日方向の翻訳確率  $P(j|e)$ 、英日方向の単語の翻訳確率の積、フレーズの日英方向の翻訳確率  $P(e|j)$ 、日英方向の単語の翻訳確率の積、フレーズペナルティ (値は常に自然対数の底  $e=2.718$ ) である。

### 2.3.4 フレーズテーブル作成法

まず, GIZA++を用いて学習文から日英, 英日方向の双方向で最尤な単語アライメントを得る. 日英方向の単語対応の例を表 2.3 に示す, 英日方向の単語対応の例を表 2.4. また, は単語が対応した箇所を示す.

表 2.3: 日英方向の単語対応

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

表 2.4: 英日方向の単語対応

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

次に, 得られた双方向の単語アライメントを用いて, 複数単語のアライメントを得る. このアライメントは双方向の単語対応の和集合と積集合から求める. ヒューリスティックスとして双方向ともに対応する単語対応を用いる “intersection”, 双方向のどちらか一方でも対応する単語対応を全て用いる “union” がある. 表 2.4 と表 2.3 を用いた “intersection” の例を表 2.5, “union” の例を表 2.6 に示す.

表 2.5: intersection の例

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

表 2.6: union の例

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

また “intersection” と “union” の中間のヒューリスティックスとして “grow” と “grow-diag” がある . これら 2 つのヒューリスティックスでは “intersection” の単語対応と “union” の単語対応を用いる . “grow” は縦横方向 , “grow-diag” は縦横対角方向に , “intersection” の単語対応から “union” の単語対応が存在する場合にその単語対応も用いる . “grow” の例を表 2.7 に , “grow-diag” の例を表 2.8 に示す .

表 2.7: grow の例

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

表 2.8: grow-diag の例

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

“grow-diag” の最後に行う処理として “final” と “final-and” がある．“final” は少なくとも片方の言語の単語対応がない場合に，“union” の単語対応を追加する．また，“final-and” は，両側言語の単語対応がない場合に，“union” の候補対応点を追加する．“grow-diag-final” の例を表 2.9 に，“grow-diag-final-and” の例を表 2.10 に示す．



表 2.9: grow-diag-final の例

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

表 2.10: grow-diag-final-and の例

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

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

### 2.3.5 パラメータチューニング

パラメータチューニングは、デコーダで用いるパラメータを最適化することである。一般的に Minimum Error Rate Training (MERT) [16] という手法が用いられる。MERT は目的の評価関数 (一般的には BLEU) を最大にする翻訳結果が選ばれるように、パラメータ調整を行う。この際、ディベロップメントデータと呼ばれる、試し翻訳を行うデータを与える。そして、各文に対して上位 100 個程度の翻訳候補を出力し、よりよい翻訳候補が上位にくるようパラメータを調整する。

## 2.4 評価方法

本研究では、翻訳システムによって出力した文の評価に自動評価法と人手評価法を用いる。

### 2.4.1 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ある。本研究では、BLEU[8]、METEOR[9]、RIBES[10] を用いる。

#### 2.4.1.1 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。

$$BLEU_{score} = BP \times \exp \left( \sum_{n=1}^N \frac{1}{N} \log p_n \right) \quad (2.3)$$

$$p_n = \frac{\sum_i \text{出力文 } i \text{ と正解文 } i \text{ で一致した } N\text{-gram の数}}{\sum_i \text{出力文 } i \text{ の } N\text{-gram の数}} \quad (2.4)$$

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、“ $\sum_i \text{出力文 } i \text{ の } N\text{-gram の数}$ ” 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．

#### 2.4.1.2 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．

$$METEOR_{score} = F_{mean} \times (1 - Pen) \quad (2.5)$$

$$F_{mean} = \frac{P \times R}{\alpha \times P + (1 - \alpha) \times R} \quad (2.6)$$

$$Pen = \gamma \times \left(\frac{c}{m}\right)^\beta \quad (2.7)$$

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．

BLEUとMETEORとRIBESでは0から1までの間で評価され，NISTでは0から1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．なお，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．

#### 2.4.1.3 RIBES

RIBESは，正解文と出力文の間で，共通単語の出現順序を順位相関係数で評価を行う評価法である．RIBESは以下の式で計算される．

$$RIBES(S) = NSR \times P^\alpha \quad (2.8)$$

$$RIBES(K) = NKT \times P^\alpha \quad (2.9)$$

$$P = \frac{n}{h} \quad (2.10)$$

ここで， $NSR$ はスピアマンの順位相関係数であり， $NKT$ はケンドールの順位相関係数である．また $P$ はペナルティであり， $n$ は出力文と正解文との間で共通な単語の数， $h$

は出力文の単語数で計算する． $\alpha$  はペナルティに対して主として使用され， $0 \leq \alpha \leq 1$  の値である．単語の出現順を順位相関係数を用いて評価することで，文全体の語順に着目することができる．

## 2.4.2 人手評価

人手評価として，対比較評価がある．対比較評価は，二つの文を相対的に比較して，どちらがより正しい文であるかを人手で選択する評価方法である．二つの翻訳システムの出力で優劣を判断する場合に有効である．

対比較を行う例を表 2.11 に示す．表 2.11 より，“手法 A”と“手法 B”を比較すると，正しい翻訳をしている文は“手法 B”である．よって，この場合は評価者によって，表 2.12 の選択肢から“手法 B”が選択される．また，“手法 A”と“手法 B”の対比較評価を 100 文に対して行った場合の結果の例を表 2.13 に示す．表 2.13 では“手法 B”が優れていると見て取ることができる．

また，人手評価は評価者によって評価に差が出てくる可能性がある．しかし，表 2.11 のように正解文を用意しておき，参照して対比較評価を行うことで，信頼性を高める．

表 2.11: 対比較例評価出力例

例 1	
入力文	これは卒業論文です。
正解文	This is a graduation thesis.
手法 A	This is a student yearbook.
手法 B	This is a thesis of graduation.

表 2.12: 対比較評価基準

基準名	意味
手法 A	手法 A の方が良い
手法 B	手法 B の方が良い
差なし	手法 A と手法 B の出力文の単語対応や伝わる意味に差がない
一致	手法 A と手法 B の出力文が完全一致

表 2.13: 手法 A VS 手法 B

手法 A	手法 B	差なし	一致
3 文	12 文	78 文	7 文

## 第3章 提案手法

本研究では、翻訳モデル作成の際の学習データとして日英対訳文とその日英対訳文から抽出された日英対訳句を学習データへ追加した日英対訳データを用いる。具体的な日英対訳句の抽出方法としては、手動で日英対訳文から日英対訳句を抽出する手動抽出と、自動で日英対訳文から日英対訳句を抽出する自動抽出がある。そして、提案手法として2通りの実験手法を試みる。日英対訳文から日英対訳句を抽出する際に手動抽出を用いた場合を手動手法、日英対訳文から日英対訳句を抽出する際に自動抽出を用いた場合を自動手法とする。

日英統計翻訳の流れの手順を以下に示す。

手順1 英語文を学習データとして言語モデルを作成する

手順2 日英対訳文とその日英対訳文から抽出した日英対訳句を組み合わせた日英対訳データを作成する

手順3 手順2で作成した日英対訳データを学習データとして翻訳モデルを作成する

手順4 手順1と手順3で作成したモデルを用いて統計翻訳を行う

### 3.1 日英統計翻訳の流れ

図 3.1 に日英統計翻訳の流れを示す .

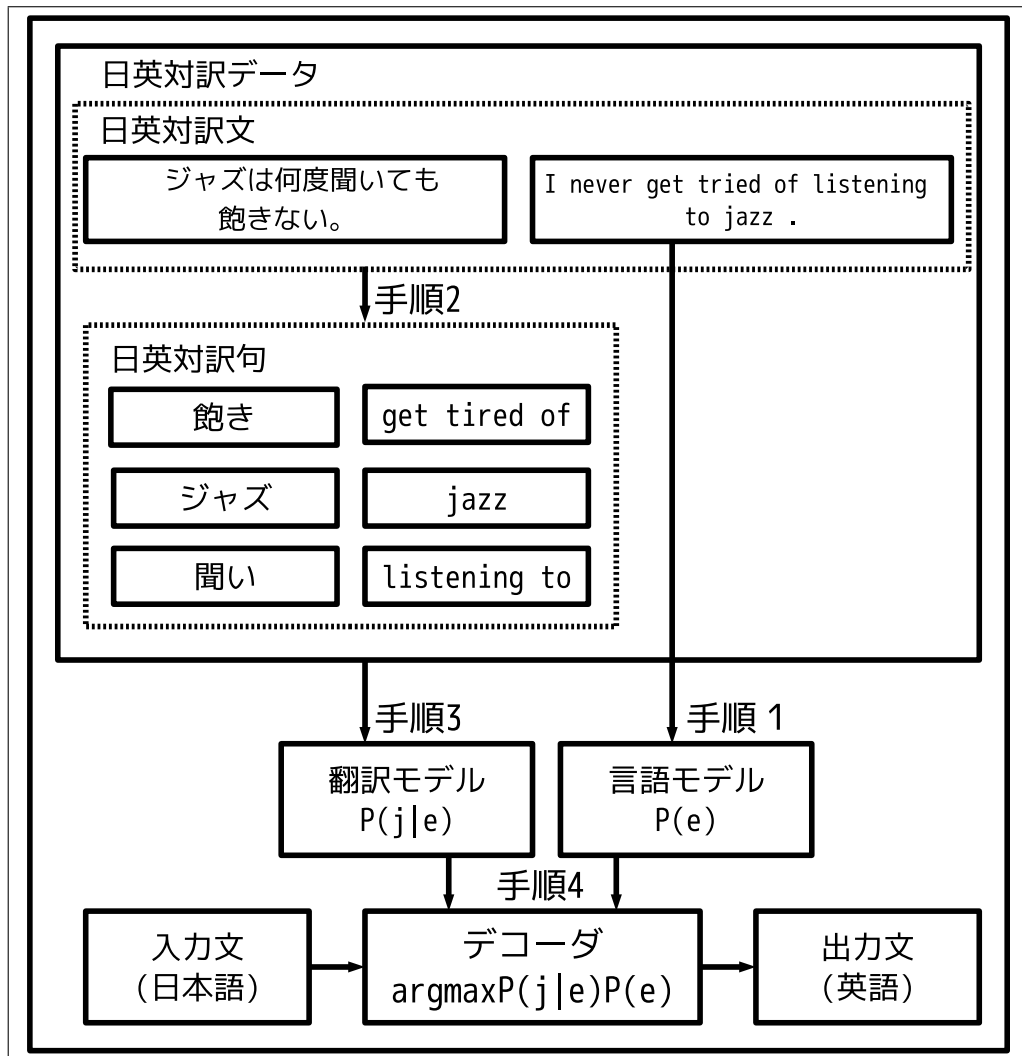


図 3.1: 日英統計翻訳の流れ

## 3.2 日英対訳句抽出システム

提案手法において，日英対訳文から日英対訳句を抽出する必要がある．日英対訳文から日英対訳句を抽出する手法には大きく分けて手動抽出と自動抽出の2通りの手法がある．本章では，この2通りの手法の説明を行う．

### 3.2.1 手動抽出

日英対訳句の手動抽出の方法として，日英対訳文とその日英対訳文から手動で作成した日英対訳文パターンを用いる方法がある．具体的には，日英対訳文とその日英対訳文から手動で作成した日英対訳文パターンを比較した変数部から品詞が *AJ*(形容詞)，*V*(動詞)，*N*(名詞)，*ADV*(副詞)，*AJV*(形容動詞)，*VP*(動詞句)，*AJVP*(形容動詞句)，*ADVP*(形容詞句)，*NP*(名詞句) の日英対訳句を抽出し作成する．また，日英対訳句は基本的に句単位で抽出しているが，単語単位や節単位の形で抽出されたものも含まれる．なお，詳しい日英対訳文パターン作成方法は池原らの論文 [12] を参照のこと．

日英対訳文の例を表 3.1，日英対訳文パターンの例を表 3.2，日英対訳文と日英対訳文パターンの比較により抽出された日英対訳句を表 3.3 に示す．

表 3.1: 日英対訳文の例

---

彼のお母さんがああ若いとは思わなかった。
I never expected his mother to be so young .

---

表 3.2: 日英対訳文パターンの例

---

$N_1 N_2$ が ああ $N_3$ とは $N_4$ なかった。
I never $N_4$ $N_1 N_2$ to be so $N_3$ .

---



表 3.3: 日英対訳文と日英対訳文パターンの比較により抽出された日英対訳句の例

日本語句	英語句
彼の	his
お母さん	mother
彼の お母さん	his mother
若い	young
思わ	expected

### 3.2.2 自動抽出

日英対訳句の自動抽出の方法として、日英対訳文とその日英対訳文から自動で作成した日英対訳文パターンを用いる方法がある。具体的には、まず、GIZA++を用いて、日英対訳単語を作成する。次に、その日英対訳単語と日英対訳文を用いて日英対訳文パターンを作成する。そして、日英対訳文と日英対訳パターンを用いて日英対訳句を抽出する。

#### 3.2.2.1 日英対訳単語の作成

GIZA++を用いて、日英対訳単語を作成する。手順を以下に示す。

手順 1 GIZA++を用いて日英対訳文から日英方向と英日方向の単語対応を得る。

手順 2 単語対応より日英対訳単語を得る。

手順 3 日英方向と英日方向の単語の翻訳確率を掛け合わせ、日英対訳単語の翻訳確率(以下、日英対訳単語翻訳確率と表記)を得る。

手順 4 日英対訳単語翻訳確率が一定の閾値 ( $\alpha$ ) 以上である日英対訳単語を抽出する。

日英対訳単語の作成の例を図 3.2 に示す。

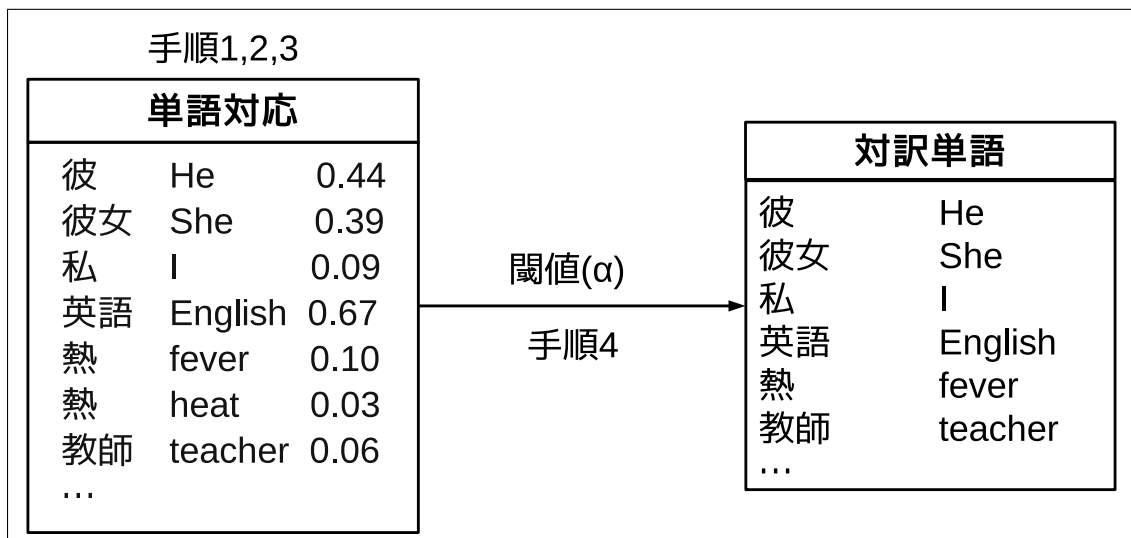


図 3.2: 日英対訳単語作成の例

### 3.2.2.2 日英対訳文パターンの作成

日英対訳単語と日英対訳文を用いて日英対訳文パターンを作成する。手順を以下に示す。

手順 1 3.2.2.1 節で抽出した日英対訳単語が日英対訳文中で適合した場合、変数化を行い、日英対訳文パターンを得る。

手順 2 日英対訳文パターンの英文パターンにおいて、変数の直前に冠詞がある場合、冠詞を除去する。

なお、変数が連続しない日英対訳文パターンのみを本研究で用いる日英対訳文パターンとする。日英対訳文パターンの作成の例を図 3.3 に示す。

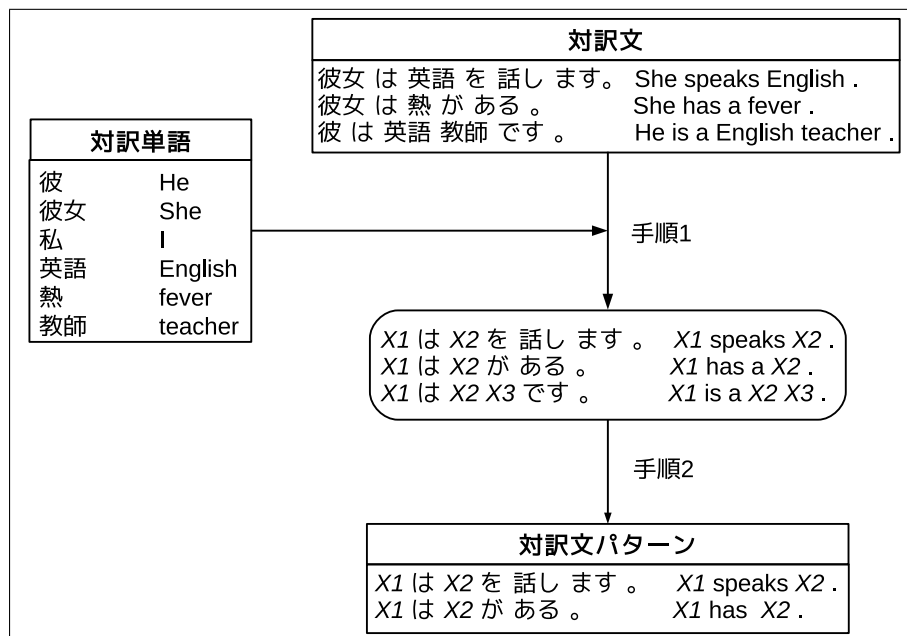


図 3.3: 日英対訳文パターン作成の例

### 3.2.2.3 日英対訳句の抽出

日英対訳文と日英対訳文パターンを用いて日英対訳句を抽出する．手順を以下に示す．

手順 1 日英対訳文と日英対訳文パターンを照合する．

手順 2 日英対訳文が日英対訳文パターンに適合した場合，日英対訳文パターンの変数部に対応する対を日英対訳句として抽出する．

日英対訳文パターンの作成の例を図 3.4 に示す．

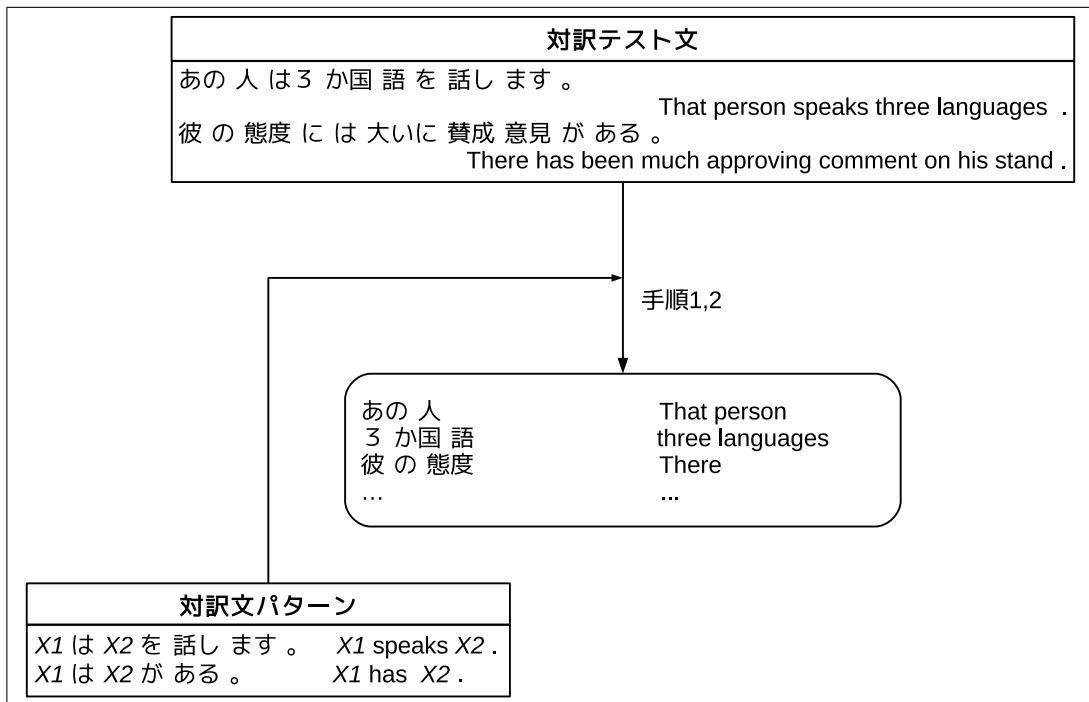


図 3.4: 日英対訳句抽出の例

### 3.3 提案手法の設定

本研究の、提案手法として、2通りの実験手法を設定する。2通りの実験手法を以下に示す。

**手動手法** 日英統計翻訳の流れの手順2において、日英対訳文から日英対訳句を抽出する際に手動抽出を用いた場合の手法

**自動手法** 日英統計翻訳の流れの手順2において、日英対訳文から日英対訳句を抽出する際に自動抽出を用いた場合の手法

## 第4章 実験環境

### 4.1 実験データ

本研究では、実験データとして日英対訳文の重文複文 [4] を用いる。本研究で用いる日英対訳文の内訳を表 4.1 に示す。日英対訳文の例を表 4.2 に示す。

表 4.1: 日英対訳文の内訳

	日本語	英語
学習データ (文)	100,000	
単語数 (語)	1,404,366	1,137,163
異なり単語数 (語)	38,530	37,035
平均文長 (文)	14.0	11.4
テストデータ (文)	10,000	
単語数 (語)	140,671	113,586
異なり単語数 (語)	13,765	12,176
平均文長 (文)	14.1	11.4
デベロップメントデータ (文)	1,000	
単語数 (語)	14,237	1,11,591
異なり単語数 (語)	3,438	3,047
平均文長 (文)	14.3	11.6

表 4.2: 日英対訳文の例

あの男はあいさつもせずに帰ってしまった。 That man left without so much as saying goodbye .
---------------------------------------------------------------------------

### 4.2 実験ツール

統計翻訳の前処理として、日本語句と日本語文に対して、MeCab[13] を用いて形態素解析を行う。また、英語句と英語文に対して “tokenizer.sed [14]” を用いて分かち書きを

行う。

翻訳モデルの学習には，“train-model.perl[16]”を用いる。言語モデルの学習には，“SRILM[15]”の“ngram-count”を用いる。本研究では， $N$ -gram モデルは 5-gram とする。本実験では統計翻訳のデコーダとして，“moses[16]”を用いる。moses のパラメータは，“mert-moses.pl[16]”を用いてチューニングを行う。また，統計翻訳においては，moses の設定ファイル“moses.ini”の“distortion-limit”の値を“-1”に変更する。distortion-limit はフレーズの並び替えを制限する値である。-1 にすることで，無制限とする。

### 4.3 評価方法

本研究では，出力文の評価として自動評価と人手評価を行う。自動評価は自動評価法“BLEU[8]”，“METEOR[9]”，“RIBES[10]”を用いる。人手評価は出力文からランダムに 100 文抽出し，対比較評価を行う。

## 第5章 日英対訳句の抽出

本研究では，日英対訳句として前章で示した学習データとして用いた日英対訳文の重文複文から抽出したものをを用いる [11]．本章では，翻訳実験で用いる日英対訳句の抽出を行う．具体的な日英対訳句の抽出方法としては，手動で日英対訳文から日英対訳句を抽出する手動抽出と，自動で日英対訳文から日英対訳句を抽出する自動抽出がある．

### 5.1 手動抽出

手動抽出した日英対訳句として，鳥バンク [11] の手動で作成したパターンと原文を比較した変数部から品詞が *AJ*(形容詞)，*V*(動詞)，*N*(名詞)，*ADV*(副詞)，*AJV*(形容動詞)，*VP*(動詞句)，*AJVP*(形容動詞句)，*ADVVP*(形容詞句) の日英対訳句を抽出したものをを用いる．本研究で用いる，手動抽出した日英対訳句の内訳を表 5.1 に示す．日英対訳句の例を表 5.2 に示す．

表 5.1: 手動抽出した日英対訳句の内訳

	日本語	英語
学習データ (句)	549,694	
単語数 (語)	1,058,949	1,031,265
異なり単語数 (語)	35,613	34,823
平均句長 (語)	1.93	1.88

表 5.2: 手動抽出した日英対訳句の例

日本語句	英語句
帰っ	left
あの	That
あの男	That man
男	man
あいさつ	saying goodbye

手動抽出した日英対訳句には誤りが存在する．そこで，ランダムに 100 句抽出して人手評価を行った．評価基準を以下に示す．評価結果と評価例を表 5.3 に示す．

適切な対応をとる日英対訳句

部分的に適切な対応をとる日英対訳句

× 不適切な対応をとる日英対訳句

表 5.3: 手動抽出した日英対訳句の評価結果と評価例

評価	句数	日英対訳句例	
	83 句	この分野	this field
		紋章	a coat of arms
		外国為替市場	the foreign exchange market
	12 句	私	My
		両社	companies
		少しも	a stroke
×	5 句	びたっと	all
		日本	it
		ナポレオン	his

## 5.2 自動抽出

自動抽出した日英対訳句として，日英対訳文とその日英対訳文から自動で作成した日英対訳文パターンを用いて抽出したものをを用いる．具体的には，まず，GIZA++を用いて，日英対訳単語を作成する．次に，その日英対訳単語と日英対訳文を用いて日英対訳文パターンを作成する．そして，日英対訳文と日英対訳パターンを用いて日英対訳句を抽出する．また，3.2.2.1 節で説明した，日英対訳単語翻訳確率の値に対しての閾値として  $\alpha = 0.15$  を用いる．

自動抽出によって得られた日英対訳句の内訳を表 5.4 に示す．

自動抽出した日英対訳句からランダムに 100 句抽出して人手評価を行った．評価結果と評価例を表 5.5 に示す．



表 5.4: 自動抽出した日英対訳句の内訳

	日本語	英語
学習データ (句)	18,192	
単語数 (語)	56,885	43,638
異なり単語数 (語)	9,003	7,770
平均句長 (語)	3.13	2.4

表 5.5: 自動抽出した日英対訳句の評価結果と評価例

評価	句数	日英対訳句例
	29 句	5 年間 守ってくれる印 その問題を解くことができ
		five years symbol against evil solved the problem
	25 句	遠く いかにも精力的 この時刻に
		long aggressive at this hour
×	46 句	ある わたしの主義 職場を
		people My lack of perseverance

## 第6章 翻訳実験

本章の実験では、翻訳モデル作成の際の学習データとして、日英対訳文とその日英対訳文から抽出した日英対訳句を組み合わせた日英対訳データを用いる。また、日英対訳句の抽出方法として、手動抽出と自動抽出を用いる。手動抽出した日英対訳句を用いた手法を手動手法、自動抽出した日英対訳句を用いた手法を自動手法とする。翻訳モデル作成の際の学習データとして、日英対訳文のみを用いる場合を `baseline` 手法として、手動手法と自動手法との翻訳精度の比較を行う。

### 6.1 実験手法

以下に翻訳実験における、実験手法を示す。

`baseline` 手法 日英対訳文のみを用いて学習を行った場合の手法

手動手法 日英統計翻訳の流れの手順2において、日英対訳文から日英対訳句を抽出する際に手動抽出を用いた場合の手法

自動手法 日英統計翻訳の流れの手順2において、日英対訳文から日英対訳句を抽出する際に自動抽出を用いた場合の手法

## 6.2 自動評価結果

実験の翻訳結果に対する自動評価結果を表 6.1 に示す。

表 6.1: 自動評価結果

自動評価手法	baseline 手法	手動手法	自動手法
BLEU(%)	11.98	<b>14.18</b>	12.54
METEOR(%)	39.62	<b>44.71</b>	42.55
RIBES(%)	66.54	<b>69.30</b>	68.02

自動評価結果のまとめを以下に示す。

- 自動手法と手動手法共に baseline 手法と比較すると自動評価値が向上している。
- 自動手法と手動手法を比較すると手動手法の方が自動評価値が向上している。

## 6.3 人手評価結果

対比較評価結果の表記の説明を以下に示す。

- baseline 手法  
対比較評価において、比較対象の翻訳結果の翻訳精度よりも baseline 手法がより優れていると判断した場合
- 手動手法  
対比較評価において、比較対象の翻訳結果の翻訳精度よりも手動手法がより優れていると判断した場合
- 差なし  
対比較評価において、比較する両実験の翻訳結果の翻訳精度に明確な差がないと判断した場合
- 一致  
対比較評価において、比較する両実験の翻訳結果が完全に同一出力の場合

人手評価手法として出力結果からランダムに 100 文抽出して対比較評価を行う。baseline 手法と手動手法の対比較評価結果を表 6.2，baseline 手法と自動手法の対比較評価結果を表 6.3，手動手法と自動手法の対比較評価結果を表 6.4 に示す。

表 6.2: baseline 手法 VS 手動手法

baseline 手法	手動手法	差なし	一致
3 文	16 文	77 文	4 文

表 6.3: baseline 手法 VS 自動手法

baseline 手法	自動手法	差なし	一致
7 文	17 文	64 文	12 文

表 6.4: 手動手法 VS 自動手法

手動手法	自動手法	差なし	一致
10 文	7 文	76 文	7 文

人手評価結果のまとめを以下に示す。

- 自動手法と手動手法共に baseline 手法と比較すると人手評価が高い。
- 自動手法と手動手法を比較すると手動手法の方が人手評価が高い。

表 6.5 baseline 手法と手動手法の翻訳結果の対比較評価において，baseline 手法が良いと判断した例

表 6.6 baseline 手法と手動手法の翻訳結果の対比較評価において，手動手法が良いと判断した例

表 6.7 baseline 手法と手動手法の翻訳結果の対比較評価において，差なしと判断した例

表 6.5: baseline 手法 > 手動手法

入力文	あんなやつを負かすのは何でもない。
正解文	I can beat him hands down .
baseline 手法	on my mettle , I can beat him hands down .
手動手法	I can beat him .
入力文	パンを買うお金をいくらか下さい。
正解文	Give me something to buy bread with .
baseline 手法	Please put some money to buy bread .
手動手法	Please some money to buy bread .
入力文	この問題が解決されるまでには何年も待たねばなるまい。
正解文	We would have to wait a good number of years before this problem is solbed .
baseline 手法	I will have to wait years before it can be solved the problem .
手動手法	This problem can be solved by will have to wait for years .

表 6.6: baseline 手法 < 手動手法

入力文	警官は私に止まれと合図した。
正解文	The policeman gave me a sign to stop .
baseline 手法	The policeman to me to stop .
手動手法	The policeman signaled to me to stop .
入力文	世間の口はうるさいものだ。
正解文	People will talk .
baseline 手法	I find it of the world .
手動手法	It is mouth of the world .
入力文	彼は愉快な人だ。
正解文	He is a lot of fun .
baseline 手法	He is a person .
手動手法	He is fun .

表 6.7: baseline 手法 ≈ 手動手法

入力文	いつ死んでもよい。
正解文	I am ready to die .
baseline 手法	I would rather die than ever .
手動手法	It is all gone forever .
入力文	まさに降ってわいた幸運だった。
正解文	It was quite a windfall .
baseline 手法	It was nobody knew with what .
手動手法	He sprang up if it was lucky .
入力文	雨のため外出せず家に止まった。
正解文	The rain kept me indoors .
baseline 手法	The rain stopped at home to go out .
手動手法	He left out of the rain stopped at home .

表 6.8 baseline 手法と自動手法の翻訳結果の対比較評価において，baseline 手法が良いと判断した例

表 6.9 baseline 手法と自動手法の翻訳結果の対比較評価において，自動手法が良いと判断した例

表 6.10 baseline 手法と自動手法の翻訳結果の対比較評価において，差なしと判断した例

表 6.8: baseline 手法 > 自動手法	
入力文	パンを買うお金をいくらか下さい。
正解文	Give me something to buy bread with .
baseline	Please put some money to buy bread .
自動手法	Please some money to buy bread .
入力文	生き物をいじめるのは悪い。
正解文	It is wrong to torment any creature .
baseline 手法	Cruelty to living in the wrong .
自動手法	It is wrong to living .
入力文	暑さにやられて食欲がなくなってしまった。
正解文	The sun has scorched up my appetite for food .
baseline 手法	I had no appetite and by heat .
自動手法	Iwas destroyed the heat of appetite .

表 6.9: baseline 手法 < 自動手法

入力文	どうもそう思えてしょうがない。
正解文	I am strongly inclined to think so .
baseline	I can not think so .
自動手法	I am strongly inclined to think so .
入力文	ぜいたくをしていれば結局貧乏になる。
正解文	Extravagance ends in want .
baseline 手法	If you have turned out to be a luxury .
自動手法	If you extravagance turned out to be poor .
入力文	彼はよい選び方をしたと思う。
正解文	I think he made a good choice .
baseline 手法	He is a good one of my choice .
自動手法	I think he is a nice choice .

表 6.10: baseline 手法 ≈ 自動手法

入力文	あいつは何をやらせても抜けた男だ。
正解文	Whatever you give him to do , he manages to get out of it .
baseline 手法	He is a man of whatever .
自動手法	He is a man of him to do anything .
入力文	今日ではなく、昨日行けば良かったのだ。
正解文	We should have gone yesterday , not today .
baseline 手法	I am glad you yesterday, but today .
自動手法	If you go , not today happened yesterday .
入力文	道路を横断するときは車に注意しなさい。
正解文	Watch out for the traffic when you cross the street .
baseline 手法	Be careful for cars when you cross the street .
自動手法	The meeting started than thirty minutes behind schedule .



表 6.11 手動手法と自動手法の翻訳結果の対比較評価において，手動手法が良いと判断した例

表 6.12 手動手法と自動手法の翻訳結果の対比較評価において，自動手法が良いと判断した例

表 6.13 手動手法と自動手法の翻訳結果の対比較評価において，差なしと判断した例

表 6.11: 手動手法 > 自動手法

入力文	方々探したが見つからなかった。
正解文	I looked everywhere , but I could not find it .
手動手法	I looked everywhere , but I could not find it .
自動手法	I looked for people , but could not find it .
入力文	あなたはジョンの奥さんだと思い込んでいた。
正解文	I was under the apprehension that you were John 's wife .
手動手法	He has the illusion that you are John 's wife .
自動手法	John 's wife was the illusion that he is a man of you .
入力文	大金をもうける方法を考え出した。
正解文	They worked out a plan for making a lot of money .
手動手法	He thought out a way to make a lot of money .
自動手法	thought out a way to make a large sum of money .

表 6.12: 手動手法 < 自動手法

入力文	そんなことをしては弁解の余地がない。
正解文	Such action admits of no excuse .
手動手法	If you were to do such a thing .
自動手法	There is no excuse to do such a thing .
入力文	彼が言ったことについてよく考えてみた。
正解文	I reflected upon what he had said .
手動手法	I often think about what he said .
自動手法	of about what he said .
入力文	本報告書は、ディスクメモリー産業を詳細に調べたものである。
正解文	This report takes an in-depth look at the disk memory industry .
手動手法	The report to see the disk memory industry .
自動手法	The report assesses the details of the disk memory industry .

表 6.13: 手動手法 ≈ 自動手法

入力文	子供を保護するのは親の務めだ。
正解文	It is the parent 's job to protect their children .
手動手法	It is the duty of parents to protect the children .
自動手法	It is the duty of parents to protect the child .
入力文	品物を順にまわして見てください。
正解文	Please hand on the article .
手動手法	The goods in turn , please and look .
自動手法	Please , and the goods at least .
入力文	この章を読み終えるまで待ってください。
正解文	Wait till I have finished this chapter .
手動手法	Please wait until I had finished reading this section .
自動手法	Please wait until I finish this chapter to read .

# 第7章 自動抽出の改良に伴う自動手法の有効性の調査

自動抽出においては，パラメータを調節することなどの工夫をしていくことにより，より多くの日英対訳句の抽出や，より高い精度の日英対訳句の抽出が可能であると考えられる．そこで，本節では，パラメータの調節による自動抽出の改良に伴う，自動手法の有効性の調査を行う．

## 7.1 単語翻訳確率の閾値の変更

本節では，3.2.2.1 節で説明した，対訳単語翻訳確率の値に対しての閾値 ( $\alpha$ ) を変更して実験を行う．なお，翻訳実験の自動手法の際の自動抽出で用いた閾値は ( $\alpha = 0.15$ ) となっている．今回は，閾値 ( $\alpha = 0.10$ )，閾値 ( $\alpha = 0.20$ ) で自動抽出を行い，抽出性能の評価を行う．さらに，抽出された対訳句を用いた自動手法による翻訳実験を行う．

### 7.1.1 対訳句の自動抽出

閾値 ( $\alpha = 0.10$ ) に設定した場合の自動抽出した日英対訳句の内訳を表 7.1，評価結果と評価例を表 7.2 に示す．閾値 ( $\alpha = 0.20$ ) に設定した場合の自動抽出した日英対訳句の内訳を表 7.3，評価結果と評価例を表 7.4 に示す．

表 7.1: 閾値 ( $\alpha = 0.10$ ) で自動抽出した日英対訳句の内訳

	日本語	英語
学習データ (句)	58,515	
単語数 (語)	167,143	131,103
異なり単語数 (語)	18,398	15,709
平均句長 (語)	2.86	2.24

表 7.2: 閾値 ( $\alpha = 0.10$ ) で自動抽出した日英対訳句の評価結果と評価例

評価	句数	日英対訳句例
	29 句	わが子 own child に注意 look out
	15 句	丈夫 healthy 入力 simply enter
×	56 句	新聞 Lend 大家 This piece

表 7.3: 閾値 ( $\alpha = 0.20$ ) で自動抽出した日英対訳句の内訳

	日本語	英語
学習データ (句)	6,423	
単語数 (語)	16,715	12,911
異なり単語数 (語)	3,700	2,961
平均句長 (語)	2.6	2.0

表 7.4: 閾値 ( $\alpha = 0.20$ ) で自動抽出した日英対訳句の評価結果と評価例

評価	句数	日英対訳句例
	36 句	この言葉 This word ほとんど不可能に近い be almost impossible
	29 句	我々がその問題を解く solve the problem 病んだ。 hurt her
×	35 句	恥 honourable する your years

表 7.1, 表 7.2, 表 7.3, 表 7.4 より, 閾値を低い値に設定すると抽出できる対訳句数は増加するが, 対訳句の精度は下がっている。また, その逆に閾値を高く設定すると抽出できる対訳句数は減少するが, 対訳句の精度は上がっている。つまり, 対訳句数と対訳句精度はトレードオフの関係になっていることがわかる。

### 7.1.2 自動評価結果

閾値 ( $\alpha = 0.10$ ) に設定した場合の自動手法を“自動手法 (0.1)”, 閾値 ( $\alpha = 0.15$ ) に設定した場合の自動手法を“自動手法 (0.15)”, 閾値 ( $\alpha = 0.20$ ) に設定した場合の自動手法を“自動手法 (0.2)”と表記する。

自動評価結果を表 7.5 に示す。

表 7.5: 自動評価結果 (%)

自動評価手法	自動手法 (0.1)	自動手法 (0.15)	自動手法 (0.2)
BLEU	12.07	<b>12.54</b>	12.51
METEOR	42.42	<b>42.55</b>	40.92
RIBES	67.58	<b>68.02</b>	67.25

表 7.5 より，自動評価においては，閾値の変更に伴う自動評価結果にほとんど差は見られなかった．これは，閾値の変更では，対訳句数と対訳句精度のトレードオフが生じるためであると考えられる．よって，閾値の変更では，抽出できる対訳句数の増加と同時に対訳句精度の向上を図ることは難しいと考える．

## 7.2 頻度情報の活用

翻訳実験の自動手法の際の自動抽出において抽出される対訳句では頻度情報は活用されていない．つまり，同じ対訳句は2度以上は抽出されない．しかし，頻繁に出現する対訳句ほど信頼性が高い対訳句である可能性が高いため，同じ対訳句が2度以上抽出されるようにすれば頻度情報を有効的に使い自動抽出が行えるのではないかと考える．よって，本節では，3.2.2.3 節の対訳句の抽出において，頻度情報を考慮した対訳句の抽出を行う．つまり，同じ対訳句が2度以上抽出されるようにする．さらに，抽出された対訳句を用いた自動手法による翻訳実験を行う．

### 7.2.1 対訳句の自動抽出

頻度情報を考慮して自動抽出した日英対訳句の内訳を表 7.6，評価結果と評価例を表 7.7 に示す．

表 7.6: 頻度情報を考慮して自動抽出した日英対訳句の内訳

	日本語	英語
学習データ (句)	397,908	
単語数 (語)	1,115,809	852,329
異なり単語数 (語)	9,003	7,770
平均句長 (語)	2.8	2.14

表 7.7: 頻度情報を考慮して自動抽出した日英対訳句の評価結果と評価例

評価	句数	日英対訳句例
	36 句	勝利の歓声 彼の理論が正しい
	19 句	今やいっそう険悪 私の病気
x	45 句	本 この本は彼女

表 7.6, 表 7.7 より, 頻度情報を考慮した自動抽出では考慮しない場合と比べて, 約 22 倍の量の対訳句を抽出できた。ただ, 頻度情報を考慮しているだけなので, 語彙的な情報量の増加はない。さらに, 特徴としては, 頻度情報を考慮しない場合と比べて, 平均句長はやや短かくなり, 対訳句の精度はほぼ同等か向上している。

## 7.2.2 自動評価結果

頻度情報を考慮した自動抽出を用いた自動手法を“自動手法 (頻度)”と表記する。自動評価結果を表 7.8 に示す。

表 7.8: 自動評価結果 (%)

自動評価手法	自動手法	自動手法 (頻度)
BLEU	12.54	11.69
METEOR	42.55	40.46
RIBES	68.02	67.07

表 7.8 より, 頻度情報を考慮した対訳句の自動抽出を用いた自動手法の翻訳精度は, 頻度情報を考慮しない場合と比べて, 自動評価結果においてはやや劣っているという結果となった。この結果より, 原因としては, 頻度が多い対訳句が必ずしも信頼性が高い対訳句ではないことや, 頻度が少ない対訳句が重要な対訳句であることなどが考えられるが, 詳しい原因はわからない。よって, 頻度情報として, 単に出現回数を考慮するだけでは, 不十分であることが分かった。

## 第8章 考察

### 8.1 精度向上の原因分析

本研究では、対訳文と、その対訳文から手動抽出した対訳句を組み合わせた対訳データを学習データとして用いることにより、翻訳精度が向上することがわかった。さらに、自動抽出した対訳句を用いた場合でも同様に翻訳精度が向上することがわかった。

この原因は、対訳句を組み合わせることにより、フレーズテーブルの翻訳確率の変化に伴い、フレーズテーブルの信頼性が向上していることが考えられる。例えば、表 6.5 の手動手法 の例を見ると、手動手法で得られた出力文では、“合図し”が“signaled”に翻訳されているが、baseline 手法では、“合図し”がうまく翻訳されていない。ここで、手動手法と baseline 手法のフレーズテーブルの確認をしてみると、表 8.1 に示されるように、翻訳確率が baseline 手法で 0.125、手動手法で 0.311 となっており、翻訳確率の変化に伴い、フレーズテーブルの信頼性が向上していることがわかる。

この結果より、対訳句の学習により正しい対訳句の対応が強調されたことが翻訳精度の向上に繋がっていることがわかった。よって、対訳データの語彙量の増加が生じない実験環境において、対訳文と対訳句を組み合わせた対訳データを学習データとして用いた統計翻訳には有効性があると考えられる。

表 8.1: フレーズテーブル

実験手法	日本語句	英語句	日英翻訳確率
baseline 手法	合図し	signaled	0.125
手動手法	合図し	signaled	0.311

## 8.2 自動抽出と手動抽出の比較

自動抽出と手動抽出の対訳句の比較を行うと、抽出精度の面では手動抽出の方が優れている。しかし、手動抽出には抽出コストが大きくかかる。そのため、場合に依っては抽出コストがほとんどかからない自動抽出を行う手法は有効であると考えられる。さらに、自動抽出はプログラムのパラメータを調節することなどの工夫をしていくことにより、より多くの日英対訳句の抽出や、より高い精度の日英対訳句の抽出が可能であると考えられる。

## 8.3 自動抽出の改良に伴う自動手法の有効性の考察

7章の調査において、自動抽出の改良を試みた。その結果、翻訳精度の向上が確認できなかった。この原因として、閾値の変更に関しては、閾値の値を変化させても抽出精度と抽出できる対訳句数とのトレードオフが発生することにより、翻訳精度が向上しなかったと考えられる。よって、閾値の変更は対訳句の学習により翻訳精度の向上を目指す場合には、適当な方法ではないと考えられる。また、頻度情報の活用においては、原因として、頻度が多い対訳句が必ずしも信頼性が高い対訳句ではないことや、頻度が少ない対訳句が重要な対訳句であることなどが考えられるが、詳しい原因はわからない。しかし、単に出現回数を考慮するだけでは、不十分であることが分かった。今後は、有効的な頻度情報の活用方法を模索する必要があると考える。

この結果より、今回は自動抽出の改良は得られなかったが、さらに、プログラムの改良を含め、改良の余地はあると考える。



## 第9章 おわりに

本研究では、対訳文に対して新たな対訳句を追加する場合である、単に対訳データの語彙が増加するような実験環境において、翻訳精度の向上に対して、対訳句の学習により正しい対訳句の対応が強調されたことがどの程度影響しているかわからないという問題点があった。さらに、新たな対訳句の学習には対訳句の収集にコストがかかるため、学習に用いるデータの収集の面でも問題が残る。

そこで、対訳データの語彙量の増加が生じない実験環境において、対訳文とその対訳文から抽出した対訳句を学習データへ追加した統計翻訳の有効性の調査を行った。具体的には、学習に用いる対訳文から対訳句を抽出する。そして、対訳文と対訳句を学習データへ追加した対訳データを用いて、日英統計翻訳を行い、その翻訳精度を調査する。また、対訳句の抽出方法として、手動抽出と自動抽出の2通りの抽出方法を試し、対訳句の抽出精度やその対訳句を用いた場合の統計翻訳の翻訳精度の調査を行った。

その結果、自動評価と人手評価のどちらにおいても提案手法の翻訳精度は向上した。具体的には、手動抽出した対訳句を用いた場合の手法では、対訳文のみを用いる baseline の手法と比べて、BLEU 値で約 2%、METEOR 値で約 5%、RIBES 値で約 3%の向上が確認できた。また、自動抽出した対訳句を用いた場合の手法でも、対訳文のみを用いる baseline の手法と比べて、BLEU 値で約 0.5%、METEOR 値で約 3%、RIBES 値で約 1.5%の向上が確認できた。よって、提案手法の有効性が確認することができた。つまり、対訳データの語彙量の増加が生じない実験環境において、対訳文と対訳句を学習データへ追加した対訳データを用いた統計翻訳の有効性が示された。

この結果、対訳データの語彙量の増加が生じない実験環境において、対訳句の学習により正しい対訳句の対応が強調されることがわかった。また、あらかじめ用意する対訳データとしては対訳文のみを用意すればよいため、対訳データ収集のコストを増やすことなく、翻訳精度を向上させることができる。

さらに、自動抽出した対訳句を用いた場合の手法の有効性の確認ができたことにより、対訳句の抽出においてもコストをかけることなく翻訳精度が向上することが確認できた。これは、手動抽出の場合はコストが大きいという問題点に対する、解決策になると考え

られる。

よって、本研究結果より、対訳文と、その対訳文から抽出した対訳句を組み合わせた対訳データを学習データとして用いることにより、翻訳精度が向上することがわかった。

# 謝辞

最後に、3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] Popović Maja, Ney Hermann “Statistical Machine Translation with a small amount of bilingual training data”, 5th LREC SALT MIL Workshop on Minority Languages, 2006.
- [2] Francis M Tyers “Rule-based augmentation of training data in Breton-French statistical machine translation”, 13th Annual Conference of the European Association for Machine Translation, pp.213-217, 2009.
- [3] 日野聡子 “日英統計翻訳における対訳句コーパスの効果”, 鳥取大学平成 24 年度修士論文, 2013.
- [4] 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.
- [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, pp.137-148, 1993.
- [6] GIZA++ : <http://www.fjoch.com/GIZA++>
- [7] Franz Josef Och “Minimum Error Rate Training in Statistical Machine Translation”, 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”, 43th Annual Meeting of the Association of Computational Linguistics, 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] 鳥バンク : <http://unicorn.ike.tottori-u.ac.jp/toribank/>
- [12] 池原悟, 阿部さつき, 徳久雅人, 村上仁一 “非線形な表現構造に着目した重文と複文の日英文型パターン化 ”, 言語処理学会論文誌, Vol.11, No.3, pp.69-95 (2004-7)
- [13] MeCab: 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.
- [14] tokenizer.sed : <http://www.cis.upenn.edu/treebank/tokenizer.sed>
- [15] Andreas Stolcke, “SRILM - an Extensible Language Modeling Toolkit”, 7th International Conference on Spoken Language Processing, pp.901-904, 2002.
- [16] Philipp Koehn, Marcello Federico, Brooke Cowan, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, Evan Herbst, “Moses: Open Source Toolkit for Statistical Machine Translation”, Proceedings of the ACL 2007 Demo and Poster Sessions, pp.177-180, 2007.