

# 日英統計翻訳のための対訳フレーズ導入の検討

東江 恵介 村上 仁一 徳久 雅人  
鳥取大学大学院 工学研究科 情報エレクトロニクス専攻  
{s062041,murakami,tokuhisa}@tottori-u.ac.jp

## 1 はじめに

現在機械翻訳の分野において、対訳文対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。統計翻訳はプログラムを用いて、大量の対訳文対から対訳フレーズ対を自動的に生成する。よって、その対訳フレーズ対のカバー率は高いが、信頼性が低いという問題がある。

対訳フレーズ対の信頼性が低いという問題に対し、鏡味らは人手で作成した対訳フレーズ対を統計翻訳に導入し、翻訳性能の向上を報告した [1]。この先行研究は人手で作成した対訳フレーズ対を導入するために、自動作成した対訳フレーズ対の翻訳確率を利用する。そのために、自動作成した対訳フレーズ対と人手で作成した対訳フレーズ対のマッチングを行う。そして完全に一致した場合のみ、人手で作成した対訳フレーズ対に対して翻訳確率を付与する。よって、導入可能なフレーズ数が少ないという問題がある。

そこで本研究では、導入可能な対訳フレーズ対が少ないという問題に対処するために、完全に一致した対訳フレーズ対だけでなく、部分的に一致した対訳フレーズ対を統計翻訳に導入し、翻訳性能の向上を試みる。

## 2 日英統計翻訳

### 2.1 概要

日英統計翻訳は、与えられた日本語文  $j$  について、翻訳モデルと言語モデルの全ての組合せの中から確率値が最大となる英語文  $\hat{e}$  を探索することにより翻訳を行う。

$$\hat{e} = \operatorname{argmax}_e P(e|j) \\ \approx \operatorname{argmax}_e P(j|e)P(e)$$

$P(j|e)$  は翻訳モデル、 $P(e)$  は言語モデルと呼ばれる。また、 $\hat{e}$  を探索する翻訳システムはデコーダ [2] と呼ばれる。

### 2.2 翻訳モデル

翻訳モデルは、英語の単語の列から日本語の単語の列へ確率的に翻訳を行うためのモデルである。翻訳モデルは、フレーズテーブルと呼ばれる表により管理されている。表 1 にフレーズテーブルの例を示す。

表 1 フレーズテーブルの例

あす		Tomorrow		0.25	0.18	0.06	0.05	
いくつ	か		Some		0.05	0.01	0.01	0.04
うそ		lie		0.33	0.14	1	0.16	

左から、日本語フレーズ、英語フレーズ、フレーズの日英翻訳確率  $P(j|e)$ 、単語の日英翻訳確率の積、フレーズの英日翻訳確率  $P(e|j)$ 、単語の英日翻訳確率の積である。本研究では日本語フレーズ、英語フレーズ、各種確率の 3 つをまとめて、フレーズ対と呼ぶ。

### 2.3 言語モデル

言語モデルは単語の列が起こる確率を与えるモデルである。日英翻訳では、翻訳モデルで生成される訳文候補の中から英語として自然な文を選出する。代表的なモデルに、 $N$ -gram モデル [3] がある。

## 3 英辞郎

“英辞郎 [4]” 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、英辞郎には通常の英語辞書にない新しい語彙や複雑な言い回しも含まれる。本研究では人手で作成した対訳フレーズ対として英辞郎の対訳フレーズ対を用いる。また、英辞郎の日本語フレーズ、英語フレーズをまとめて、翻訳対と呼ぶ。本研究で使用した翻訳対の総数は 1,587,241 文対である。表 2 に翻訳対の例を示す。

表 2 翻訳対の例

awareness of the issues	: 問題意識 { もんだいいしき }
get off	{ 句動-8 } : 帰る、去るき
improve memory	: 物覚え { ものおぼえ } を良くする

## 4 提案手法

本研究では統計翻訳に翻訳対を導入するために、翻訳対に対して翻訳確率の付与を行う。翻訳確率を付与するために、2.2 節で述べたフレーズ対の翻訳確率を利用する。翻訳対がフレーズ対と完全に一致すれば、フレーズ対に付与されている翻訳確率を翻訳対に付与する。部分的に一致すれば、対訳フレーズ対に付与されている翻訳確率にペナルティをかけた値を翻訳対に付与する。

### 4.1 部分的に一致した翻訳対の翻訳確率

本研究ではより多くの翻訳対を統計翻訳に導入するために、翻訳対とフレーズ対が部分的に一致した際の翻訳確率を計算する。より類似したフレーズ対の翻訳確率を利用するために、翻訳対とフレーズ対の DP マッチングを行う。DP マッチングにより、翻訳対と最も一致率の高いフレーズ対を抽出する。部分的に一致した翻訳対の例を表 3 に示す。

表 3 部分的に一致した翻訳対の例

翻訳対	自分で 仕事 を する     hoe my own row
フレーズ対	仕事 を する     do the work
	0.5 0.002 1 0.005
翻訳対	手術 を 受ける     have an operation
フレーズ対	手術 を 受けた     He had an operation
	1 4.4e-05 0.33 2.28e-05

### 4.2 部分的に一致した翻訳対の確率のペナルティ

翻訳対とフレーズ対が部分的に一致した際、フレーズ対に付与されている翻訳確率を翻訳対に付与することは不適切だと考える。よって、本研究ではフレーズ対のそれぞれの翻訳確率に対して、ペナルティをかける。ペナルティは日本語フレーズと英語フレーズにおいて、それぞれ一致した単語数をそれぞれの総単語数で割った値とする。ペナルティの値を翻訳確率にかけることにより、フレーズ対と部分的に一致した翻訳対に翻訳確率を与え、ペナルティの式を以下に示す。

$$\text{ペナルティ} = \frac{\text{一致した日単語数}}{\text{全日単語数}} * \frac{\text{一致した英単語数}}{\text{全英単語数}}$$

### 4.3 翻訳確率の付与手順

翻訳対に対して、翻訳確率を付与する手順の説明を以下に示し、手順の図を図1を以下に示す。

- 手順1 対訳学習文から grow-diag-final を用いてフレーズテーブルを作成
- 手順2 翻訳対の形式をフレーズテーブルの形式に変換
- 手順3 DP マッチングを用いてフレーズテーブルから翻訳確率を抽出し、ペナルティを付け、翻訳対に付与
- 手順4 翻訳確率を付与した翻訳対をフレーズテーブルに追加

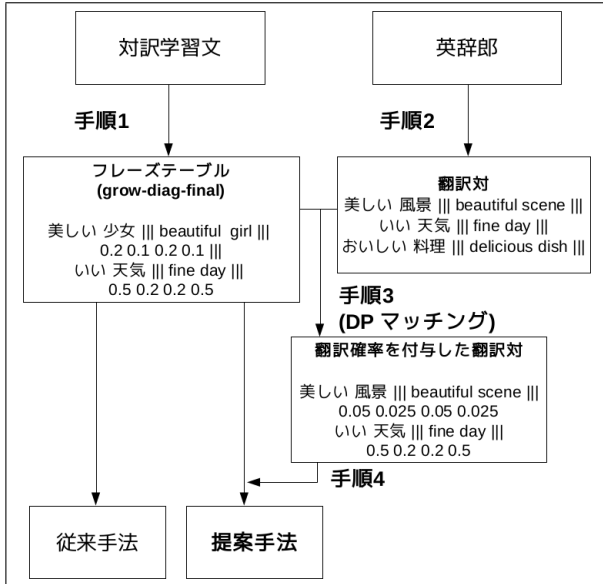


図1 提案手法の翻訳確率の付与手順

## 5 実験データ

一般に、日英統計翻訳では前処理として、日本語文を形態素解析を用いて単語に区切る。本研究では、形態素解析器として MeCab[5] を用いる。また、英語文に対して句読点の前後にスペースを入れる。この処理を対訳学習文、テスト文それぞれに行う。

### 5.1 対訳学習文

単文の実験では、対訳学習文として、辞書から抽出した単文の対訳文 100,000 文対を用いる。表4に単文の対訳学習文の例を示す。

表4 単文の対訳学習文の例

私は映画を見に行く。  
I go to see a movie.

### 5.2 テスト文

単文の日英翻訳のテスト文として、辞書から抽出した単文 1,000 文対を用いる。

### 5.3 確率を付与した翻訳対

提案手法では翻訳対の総数、1,587,241 文対のうち、単文では 69,728 文対に確率を付与した。

## 6 実験環境

### 6.1 フレーズテーブル

本研究における、フレーズテーブルの作成には、train-model.perl を用いる。なお、grow-diag-final で作成したフレーズテーブルの総数は 306,895 文対であり、intersection で作成したフレーズテーブルの総数は 12,893,466 文対である。

### 6.2 N-gram モデルの学習

言語モデルには、N-gram モデルを用いる。N-gram モデルの学習には“SRILM”[3]を用いる。本研究では、5-gram モデルを用いる。なお、スムージングには kndiscount を用いる。

### 6.3 デコーダのパラメータ

本研究の翻訳実験では、MERT による翻訳パラメータの最適化 [9] 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 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 limit”は“-1”とする。また、“distortion weight”は“0.2”とする。

### 6.4 評価方法

本研究では自動評価と人手評価を行う。自動評価には、BLEU[6] と NIST[7] と METEOR[8] の3種類を用いる。人手評価には対比較評価を行う。

## 7 従来手法と提案手法の比較

この章では従来手法と提案手法の翻訳実験を行う。なお、本研究における従来手法と提案手法を以下に示す。

従来手法 対訳学習文から grow-diag-final で生成したフレーズテーブルを用いる実験

提案手法 4章で述べた手法で確率を付与した翻訳対を追加したフレーズテーブルを用いる実験

### 7.1 翻訳精度の自動評価

テスト文に単文を用いて翻訳実験を行う。表5に単文の翻訳実験結果を示す。

表5 単文における翻訳精度の評価

単文	BLEU	NIST	METEOR
従来手法	0.1189	4.1477	0.3730
提案手法	0.1215	4.1668	0.3735

表5の結果から提案手法は従来手法より自動評価がよいことから、提案手法は有効である。

### 7.2 翻訳精度の人手評価

人手による対比較評価では、提案手法の結果と比較対象の結果から、それぞれ 100 文ランダムに抽出し、どちらの文が入力文として適切であるかを判断する。評価基準を以下に示す。

- 提案手法 提案手法の結果が比較対象の結果より優れている場合
- 提案手法× 提案手法の結果が比較対象の結果より劣っている場合
- 差なし どちらの結果も同等に理解できるまたは同等に理解できない場合
- 同一出力 提案手法の結果と先行研究の結果が同一の文である場合

#### 7.2.1 従来手法と提案手法の人手評価

表6に従来手法と提案手法における人手評価の結果を示す。

表6 従来手法と提案手法の対比較評価結果

	提案手法	提案手法×	差なし	同一出力
単文	5	0	5	90

表6の結果から提案手法は従来手法より人手評価がよいことから、提案手法は有効である。

#### 7.2.2 提案手法が優れていると判断した例

表7に提案手法が優れていると判断した例を示す。

表7 単文において提案手法が優れていると判断した例

入力文	この仕事は細心の注意を要する。
参照文	This work needs close attention .
従来手法	This task requires a 細心 attention .
提案手法	This task requires a religious attention to detail .

表7において、提案手法は未知語が無いため、提案手法の方が優れている。

## 8 先行研究と提案手法の比較

この章では先行研究 [1] と提案手法の翻訳実験を行う。なお、実験データは5章と同様であり、実験環境は6章と同様である。先行研究では翻訳対の総数、1,587,241文対のうち、単文では18,077文対に確率を付与した。なお本研究における先行研究と提案手法を以下に示す。

先行研究 8.1節で述べる手法で確率を付与した翻訳対を追加したフレーズテーブルを用いる実験

提案手法 4章で述べた手法で確率を付与した翻訳対を追加したフレーズテーブルを用いる実験

### 8.1 先行研究の手法

先行研究では翻訳対に対して翻訳確率を付与するために、intersection で作成したフレーズテーブルのフレーズ対と英辞郎で作成した翻訳対のマッチングを行う。フレーズ対と翻訳対が完全に一致した場合にのみ翻訳対にフレーズテーブルで翻訳確率を付与する。そして、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翻訳対を追加する。先行研究の手順の説明を以下に示し、手順の図を図2に示す。

手順1 対訳学習文から grow-diag-final を用いてフレーズテーブルを作成

手順2 対訳学習文から intersection を用いてフレーズテーブルを作成

手順3 翻訳対の形式をフレーズテーブルの形式に変換

手順4 フレーズテーブルのフレーズ対と完全に一致した翻訳対に翻訳確率付与

手順5 翻訳確率を付与した翻訳対をフレーズテーブルに追加

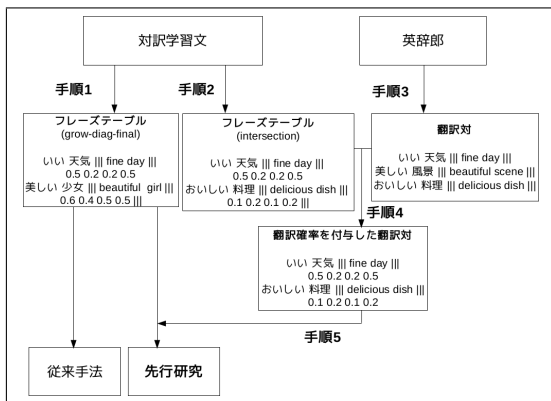


図2 先行研究の翻訳確率の付与手順

### 8.2 翻訳精度の自動評価

テスト文に単文を用いて翻訳実験を行う。表8に単文の翻訳実験結果を示す。

表8 単文における翻訳精度の評価

単文	BLEU	NIST	METEOR
先行研究	0.1267	4.3047	0.3850
提案手法	0.1215	4.1668	0.3735

表8の結果から提案手法は先行研究より自動評価が悪いことから、先行研究と比較すると有効ではない。

### 8.3 従来手法と提案手法の人手評価

表9に従来手法と提案手法における人手評価の結果を示す。なお、評価基準は7.2節と同様である。

表9 先行研究と提案手法の対比較評価結果

	提案手法	提案手法 ×	差なし	同一出力
単文	3	5	8	84

表9の結果から提案手法は先行研究より人手評価が悪いことから、先行研究と比較すると有効ではない。

## 9 先行研究と提案手法を組み合わせた翻訳実験

8章の実験結果において、提案手法は先行研究より翻訳精度が悪いという結果であった。そこで、この章では先行研究と、先行研究と提案手法を組み合わせた手法の翻訳実験を行う。なお、実験データは5章と同様であり、実験環境は6章と同様である。以後、この実験を追加実験と呼ぶ。また、本研究における、先行研究と追加実験を以下に示す。

先行研究 8.1節で述べた手法で確率を付与した翻訳対を追加したフレーズテーブルを用いる実験

追加実験 9.1節で述べる手法で確率を付与した翻訳対を追加したフレーズテーブルを用いる実験

### 9.1 追加実験の手法

追加実験は、先行研究で追加した翻訳対と提案手法で追加した翻訳対を grow-diag-final で作成したフレーズテーブルに追加した翻訳実験を行う。追加実験の手順の説明を以下に示し、手順の図を図3に示す。

手順1 対訳学習文から grow-diag-final を用いてフレーズテーブルを作成

手順2 対訳学習文から intersection を用いてフレーズテーブルを作成

手順3 翻訳対の形式をフレーズテーブルの形式に変換

手順4 DP マッチングを用いてフレーズテーブルから翻訳確率を抽出し、ペナルティをかけ、翻訳対に付与

手順5 フレーズテーブルのフレーズ対と完全に一致した翻訳対に翻訳確率付与

手順6 翻訳確率を付与した翻訳対をフレーズテーブルに追加

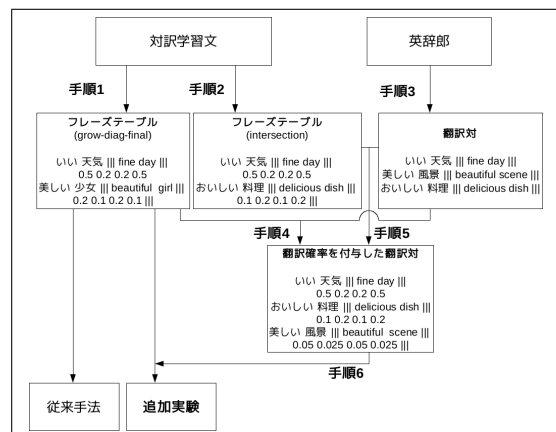


図3 追加実験の翻訳確率の付与手順

## 9.2 翻訳精度の自動評価

テスト文に単文を用いて翻訳実験を行う。表 10 に単文の翻訳実験結果を示す。

表 10 単文における翻訳精度の評価

単文	BLEU	NIST	METEOR
先行研究	0.1267	4.3047	0.3850
追加実験	0.1261	4.2965	0.3822

表 10 の結果から追加実験は先行研究より自動評価がよい。しかし、先行研究と比べると悪い。

## 9.3 先行研究と追加実験の人手評価の結果

表 11 に先行研究と追加実験における人手評価の結果を示す。

表 11 先行研究と追加実験の対比較評価結果

	追加実験	追加実験 ×	差なし	同一出力
単文	4	0	5	91

表 11 の結果から追加実験は先行研究より人手評価がよいことから、追加実験は人手評価では有効である。

### 9.3.1 追加実験が優れていると判断した例

表 12 に追加実験が優れていると判断した例を示す。

表 12 単文において追加実験が優れていると判断した例

入力文	難局を収拾する道はそれしかない。
参照文	It is the only way to save the difficult situation .
先行研究	The situation 収拾 only the way it .
追加実験	The only way that save a difficult situation .

表 12 において、追加実験は未知語が無くなっているため、追加実験の方が優れている。

表 13 に追加実験が優れていると判断した例を示す。

表 13 単文において追加実験が優れていると判断した例

入力文	この国では情報は自由に流れる。
参照文	In this country data flows freely .
先行研究	Information flows in this country is at .
追加実験	Information flows in this country free .

表 13 において、追加実験は“自由”に対応する単語が翻訳されているため、追加実験の方が優れている。

## 10 考察

### 10.1 翻訳候補文に使用された翻訳対の違い

提案手法では、69,728 文対の翻訳対に対して翻訳確率の付与を行い、先行研究では、18,077 文対の翻訳対に対して翻訳確率の付与を行った。提案手法と従来手法を比較すると、提案手法の翻訳精度は自動評価、人手評価共によい。しかし、先行研究と比較すると自動評価、人手評価共に悪いという結果になった。その原因は、統計翻訳において、翻訳候補として使用される翻訳対の数の違いだと考える。翻訳実験において、使用される翻訳対の数を調べたところ、先行研究では追加した翻訳対 18,077 文対中、翻訳候補文に使用された数は 2,481 文対であった。一方、提案手法は追加した翻訳対 69,728 文対中、翻訳候補文に使用された数は 1,067 文対であり、先行研究の半数以下の翻訳対しか使用されていない。

この原因として、翻訳確率にかけたペナルティーが適切ではないため、本来使用されるべき翻訳対の確率が低下したことが考えられる。今後は単語によるペナルティーだけでなく、フレーズや頻度も考慮に入れた適切なペナルティーを考えていきたい。

## 10.2 未知語の影響

表 14 にそれぞれの手法における未知語の総数と BLEU スコアを示す。

表 14 未知語の総数

	未知語	BLEU
従来手法	631	0.1189
先行研究	523	0.1267
提案手法	620	0.1215
追加実験	519	0.1261

表 14 から、未知語が減少するほど、BLEU スコアは向上している。先行研究と提案手法の未知語を比較すると、約 100 語の違いがある。よって、先行研究と提案手法の BLEU スコアの差は未知語が大きな原因だと考える。一方、先行研究と追加実験の未知語を比較すると、追加実験は僅かに未知語が少ない。しかし、BLEU スコアは僅かに先行研究の方がよい。その原因は、不適切な翻訳確率が付与された翻訳対がデコーディングの際、本来必要なパスの枝刈りを行い、悪影響を及ぼしたことだと考えている。

## 11 おわりに

本研究では、部分的に一致したフレーズを統計翻訳に導入することにより翻訳精度の向上を試みた。その結果、BLEU スコアにおいて、提案手法は従来手法と比較して単文で 0.1189 から 0.1215 へ 0.26% の向上を確認した。しかし、提案手法は BLEU スコア、人手評価共に先行研究より翻訳精度が悪い。その原因の 1 つは、翻訳候補に使用されるフレーズが少ないことであった。また、BLEU スコアにおける、提案手法と先行研究の差の原因は未知語の差が原因だと考えている。

追加実験において、先行研究より人手評価が優れているということから、提案手法は先行研究と組み合わせることで、僅かな有効性を確認した。今後は翻訳確率にかける、適切なペナルティを考えていきたい。

## 参考文献

- [1] 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 15 回年次大会 pp.224-227 (2009)
- [2] Philipp Koehn et al., “Moses: Open Source Toolkit for Statistical Machine Translation”, Association for Computational Linguistic, pp.177-180,(2007).
- [3] The SRI Language Model Toolkit <http://www.speech.sri.com/projects/srlm>
- [4] <http://www.alc.co.jp/>
- [5] <http://mecab.sourceforge.net/>
- [6] Kishore Papineni, Salim Rukos, Todd Ward, and Wei-Jing Zhu, “BLEU: a Method for Automatic Evaluation of Machine Translation”, Association for Computational Linguistics, pp.311-318, (2002).
- [7] George Doddington, “Automatic evaluation of machine translation quality using n-gram co-occurrence statistics”, Proceedings of the second international conference on Human Language Technology Research, (2002)
- [8] Satanjeev Banerjee and Alon Lavi, “METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments”, Association for Computational Linguistics, pp.65-72, (2005).
- [9] Franz Josef Och “Minimum Error Rate Training in Statistical Machine Translation”, Association for Computational Linguistics, pp.160-167, (2003).