

# 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

鏡味良太 村上仁一 徳久雅人 池原悟  
鳥取大学 工学部 知能情報工学科

{s042021,murakami,tokuhisa,ikehara}@ike.tottori-u.ac.jp

## 1 はじめに

現在、機械翻訳システムの分野において、対訳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が注目されている。統計翻訳では、単語列から単語列への翻訳を確率的に行うためにフレーズテーブルを用いる。しかし、フレーズテーブルのフレーズ対は、生成段階において短いフレーズ対に分割されるため、長いフレーズ対を多く含むことは困難である。また、自動作成されるフレーズテーブルは作成されるフレーズ数が多いため、カバー率は高いが、フレーズ対の信頼性は低い。一方、人手で作成する場合は翻訳対の信頼性は高いがカバー率は低い。そこで、それぞれの長所を生かすために、プログラムで自動作成したフレーズ対に人手で作成した翻訳対を追加することを考えた。

本研究では、信頼性が高く、長いフレーズ対をフレーズテーブルに追加するために、日英重文複文文型パターン辞書 [1] の対訳文対から人手で作成された翻訳対を、プログラムで自動作成したフレーズテーブルに追加し、単文と重文複文における、日英翻訳及び英日翻訳の精度評価を行う。

## 2 統計翻訳システム

### 2.1 基本概念

日英の統計翻訳は、日本語文  $j$  が与えられたとき、全ての組合せの中から確率が最大になる英語文  $\hat{e}$  を探索することによって翻訳を行う。以下にその基本モデルを示す。

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

$P(j|e)$  は翻訳モデル、 $P(e)$  は言語モデルと呼ぶ。

### 2.2 言語モデル

言語モデルは単語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 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、または、英語の単語列から日本語の単語列へ確率的に翻訳を行うためのモデルである。翻訳モデルは表 1 のようなフレーズテーブルで管理されている。

表 1 フレーズテーブルの例

桜の木の     of the cherry tree	0.1	0.01	0.33	0.03
遊び心     sense of humor	0.01	0.03	0.09	0.001
英語を学ぶ     learn english	0.02	0.01	0.5	0.4

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の積である。

## 3 実験環境

### 3.1 実験データ

実験には、日英重文複文文型パターン辞書 [1] から抽出した対訳文対、121,913 文対を用いる。尚、この辞書は日本文が重文複文であり、英文は単文や複文である。統計翻訳の前処理として、日本語文に対しては chasen[2] を用いて形態素解析を行う。英語文に対しては、句読点の前後にスペースを入れる。また、大文字の小文字化を行う。前処理を行った対訳文の例を表 2 に示す。

表 2 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 .
愛と幸せに満ちた家庭にしたいと思います。	we hope we will make our home full of love and happiness .
彼女はにっこり笑ってぼくにあいさつした。	she greeted me with a big smile .

### 3.2 人手で作成した翻訳対

人手で作成した翻訳対は、日英重文複文文型パターン辞書 [1] から抽出した対訳文対を利用する。人手で作成した翻訳対の例を表 3 に示す。尚、本稿では、人手で作成した翻訳対を翻訳対と呼ぶ。

表 3 人手で作成された翻訳対の例

コートですそ	the edge of my coat
朝晩のラッシュ時に電車に乗る	take the train during the morning and evening rush hours
国産のコートは英国製よりだいぶ落ちる	home-made coats are by far inferior to those made in britain

### 3.3 プログラムで作成するフレーズテーブル

プログラムによるフレーズテーブルの作成には, train-phrase-model.perl[3] を用いる. 尚, 本稿では, プログラムで自動作成したフレーズ対をフレーズ対と呼ぶ.

### 3.4 N-gram 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[4]" を用いる. 本研究では 5-gram モデルを用いる. また, スムージングには "ukndiscount" を用いる.

### 3.5 デコーダのパラメータ

デコーダは "moses[5]" を用いる. 本研究ではパラメータの最適化は行わない. "weight-t" はクロスエントロピーを用い, "0.5 0 0.5 0 0" とする. また, 翻訳時にフレーズの位置の変化に柔軟に対応するため, "distortion weight" は 0.2 とする.

### 3.6 評価方法

出力文の評価は自動評価法 BLEU[6] を用いる. 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.

## 4 プログラム作成したフレーズテーブルへの翻訳対の追加

### 4.1 翻訳対の追加手順

人手で作成した精度の高い翻訳対をフレーズテーブルに追加する手順を図 1 に示す.

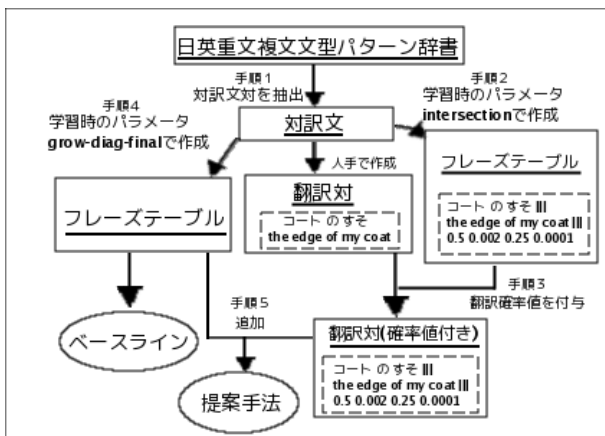


図 1 人手で作成したフレーズ対への確率値の付与方法

- 手順 1 日英重文複文型パターン辞書 [1] から対訳文対を抽出する.
- 手順 2 手順 1 で抽出した対訳文を用いてフレーズテーブルを作成する. 作成時のパラメータには intersection を用いる.
- 手順 3 手順 2 で作成したフレーズテーブルを参照して, 人手で作成した翻訳対に翻訳確率を付与する.
- 手順 4 手順 1 で抽出した対訳文を用いてフレーズテ

ブルを作成する. 作成時のパラメータには grow-diagonal を用いる.

手順 5 手順 4 で作成したフレーズテーブルに手順 3 で翻訳確率を付与した翻訳対を追加する.

尚, 本稿では, 手順 4 で作成したフレーズテーブルを用いた翻訳をベースラインと呼び, 手順 5 のフレーズテーブルを用いた翻訳を提案手法と呼ぶ.

## 5 翻訳実験

### 5.1 学習データ

学習データは日英重文複文型パターン辞書 [1] から抽出した対訳文対, 121,913 文対を用いる.

### 5.2 フレーズテーブル

ベースラインのフレーズ数, 追加したフレーズ数, 提案手法のフレーズ数を表 4 に示す.

表 4 総フレーズ数

	プログラム作成	人手作成数	追加後
日英翻訳	727,849	130,893	858,742
英日翻訳	729,410	131,514	860,924

提案手法では, 人手で作成した 261,453 個の翻訳対のうち日英翻訳には 130,893 個, 英日翻訳には 131,514 個の翻訳対に確率値が付与された. 実際に追加した翻訳対の例を表 5 に示す.

表 5 確率値を与えた翻訳対の例

コート の すそ     the edge of my coat	0.5 0.00190997 0.25 6.48554e-06
朝晩 の ラッシュ 時 に 電車 に 乗る	take the train during the morning and evening rush hours
	0.1 6.68278e-20 0.125 1.83956e-24
国産 の コート は 英国 製 より だいぶ 落ちる	home-made coats are by far inferior to those made in britain
	0.05 8.90443e-29 0.0454545 5.07768e-33

### 5.3 翻訳精度の評価

#### 5.3.1 日英翻訳の結果

日英翻訳の Open テストデータには, 単文 1,000 文及び重文複文 1,000 文を用いる. プログラムで自動作成したフレーズテーブルを用いたベースラインと, 翻訳対を追加した提案手法の翻訳精度の評価を表 6 に示す. 尚, 表中の値は BLEU の評価値である.

表 6 日英翻訳の実験結果 (テストデータ 1,000 文)

テスト文	ベースライン	提案手法
単文	0.103	0.106
重文複文	0.077	0.080

結果から, 単文, 重文複文のいずれの翻訳においても提案手法の翻訳精度が向上していることが分かる.

### 5.3.2 英日翻訳の結果

英日翻訳には、単文 100 文、重文複文 100 文のテストデータを用いる。ベースラインと提案手法の翻訳精度の評価を表 7 に示す。

表 7 英日翻訳の実験結果 (テストデータ 100 文)

テスト文	ベースライン	提案手法
単文	0.116	0.117
重文複文	0.093	0.104

結果から、単文の翻訳精度はあまり変化していないが、重文複文の翻訳精度が大きく向上していることが分かる。

### 5.4 対比較実験

表 8 の日英、英日のそれぞれの翻訳結果に対して人手による対比較実験を行う。

#### 5.4.1 判断基準

ベースラインの翻訳結果と、提案手法の翻訳結果を 100 文抽出し、どちらの翻訳が優れているかを判断する。固有名詞の未知語はローマ字変換して評価し、それ以外の未知語は存在しないものとして評価を行う。

#### 5.4.2 提案手法の翻訳結果が劣ると評価した例

提案手法の翻訳結果が、ベースラインより劣ると評価した例を表 8 に示す。

表 8 提案手法がベースラインより劣ると評価した例

入力文:彼は、トラックの運転手をしている。
正解文:he is a truck driver .
ベースライン:he is a truck driver doing .
提案手法:the truck driver , he is doing .
入力文:レポートを提出しなければ、合格点はあげません。
正解文:you cannot be given a passing grade if you do not turn in your report .
ベースライン:you must put the paper , i can not give you a point .
提案手法:the paper , you will pass the exam is not give you must submit .

#### 5.4.3 提案手法の翻訳結果が優れていると評価した例

提案手法の翻訳結果が、ベースラインより優れていると評価した例を表 9 に示す。

表 9 提案手法がベースラインより優れていると評価した例

入力文:わたしの父は、本屋をしている。
正解文:my father runs a book shop .
ベースライン:my father is that she has a 本屋 .
提案手法:my father is that she has a bookstore .
入力文:10 を 2 で割れば 5 になる。
正解文:ten divided by 2 makes 5 .
ベースライン:if you get 5 to 10 broken by two .
提案手法:ten divided by two gives ve to you .

#### 5.4.4 実験結果

ベースラインの翻訳結果の方が、文として優れていると評価した文を、“提案手法 ×” とする。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と評価した文を、“提案手法 ” とする。また、ベースラインと提案手法で翻訳結果が変化しなかった文を“変化無し” とする。対比較実験の結果を表 10 に示す。

表 10 対比較実験の結果

	テストデータ	提案手法 ×	提案手法	変化無し
日英翻訳	単文	5	18	73
	重文複文	5	15	80
英日翻訳	単文	11	14	75
	重文複文	7	15	78

結果から、全ての翻訳において、提案手法が優れていることが分かる。

## 6 翻訳対の翻訳確率の重みの最適化

提案手法の効果の更なる向上を目的として、以下の実験を行う。

### 6.1 翻訳実験

人手で作成した翻訳対の翻訳確率の重みを 2 倍、4 倍、8 倍に変化させたときの翻訳精度の評価を表 11 に示す。

表 11 翻訳確率の重みを変化させた時の翻訳実験結果

	テストデータ	ベースライン	1 倍	2 倍	4 倍	8 倍
日英翻訳	単文	0.103	0.106	0.105	0.105	0.106
	重文複文	0.077	0.080	0.081	0.085	0.085
英日翻訳	単文	0.1116	0.117	0.118	0.126	0.115
	重文複文	0.093	0.104	0.107	0.113	0.118

日英の重文複文の翻訳において、翻訳確率の重みを 4 倍にした時に翻訳精度が大きく向上した。また、英日の単文の翻訳においては、4 倍の時に大きく翻訳精度が向上したが、8 倍になると低下した。重文複文においては 2 倍、4 倍、8 倍の時に翻訳精度が向上した。

最適な翻訳確率の重みを用いたときの提案手法の翻訳精度と、ベースラインの翻訳精度の差を表 12 に示す。

表 12 ベースラインに対する提案手法の効果

	テストデータ	提案手法の効果	翻訳確率の重み
日英翻訳 1,000 文	単文	+0.003	8 倍
	重文複文	+0.008	4 倍
英日翻訳 100 文	単文	+0.010	4 倍
	重文複文	+0.025	8 倍

表 12 から、日英、英日共に重文複文の翻訳精度が大きく向上していることが分かる。

### 6.2 対比較実験

表 12 の翻訳結果 100 文に対して、5 章 2 節と同じ条件で対比較実験を行った結果を表 13 に示す。

表 10 の結果と比べて、特に英日翻訳の提案手法がベースラインよりも優れていると評価した割合が高くなっていることが分かる。

表 13 評価結果

	テストデータ	提案手法 ×	提案手法	変化無し
日英 翻訳	単文	13	28	59
	重文複文	15	35	50
英日 翻訳	単文	7	13	80
	重文複文	16	43	41

## 7 考察

### 7.1 提案手法の効果の分析

表 13 の対比較実験の結果において、翻訳結果が変化した 170 文中、119 文が提案手法が優れていると評価した。提案手法が優れていると評価した理由は、文質の向上と未知語の減少に分けることができる。

#### 7.1.1 文質の向上

ベースラインと比較すると、提案手法の翻訳結果が文として好ましくなり、文質が向上した例を表 14 に示す。

表 14 文として好ましくなった例

入力文:この学校では、毎年五百名の学生が卒業していく。
正解文: ve hundred students graduate from this school every year .
ベースライン:this is the name of 500 students graduate school year , and go .
提案手法:this school is 500 students graduate of the year , and go .
入力文:友達のお父さんに、駅まで車で送っていただきました。
正解文:the father of my friend drove me to the station .
ベースライン:i would like to the station by a friend of mine , to your father .
提案手法:i would like to the station , so send it to my friend 's father .

ベースラインと比較して、提案手法の翻訳結果が文として好ましくなっている。

#### 7.1.2 未知語の減少

ベースラインと比較して、未知語が減少したことにより翻訳精度が向上した例を表 15 に示す。

表 15 未知語が減少した例

入力文:日本と大陸はかつてつながっていた。
正解文:the continent and the island of japan were linked in the old days .
ベースライン:he was once a つながっ japan and the continent .
提案手法:japan is linked up to the continent was once .
入力文:地球は温暖化しつつある。
正解文:global warming is developing .
ベースライン:the earth is 温暖 they are .
提案手法:they are warming the earth .

未知語が翻訳されたことで翻訳精度が向上している。

#### 7.1.3 文質が向上した文と未知語が減少した文の比較

提案手法が優れていると評価した 119 文において、未知語が減少した文数と、文質が向上したと思われる文数を表 16 に示す。

表 16 文質が向上した文数と未知語が減少した文数の比較

	テストデータ	未知語減少	文質が向上
日英 翻訳	単文	4	31
	重文複文	11	17
英日 翻訳	単文	2	11
	重文複文	5	38

未知語数と関係なく翻訳精度が向上した文が、119 文中 97 文と約 8 割を占めていることが分かる。結果から、人手で作成した翻訳対をプログラムで自動作成したフレーズテーブルに追加した提案手法は、主に文質の向上に効果があったと言える。

### 7.2 最適化したパラメータを用いた翻訳実験

翻訳精度全体の向上を目的として、reordering モデルのパラメータも含めたパラメータの最適化を行った。尚、フレーズテーブルの作成には mooses に付属している train-factored-phrase-model.perl を用いた。

単文の翻訳には development データに単文 100 文を使用して最適化を行ったパラメータを用いる。重文複文の翻訳には development データに重文複文 1,000 文を使用して最適化を行ったパラメータを用いる。翻訳実験の結果を表 17 に示す。

表 17 最適化したパラメータを用いた実験結果

テストデータ	ベースライン		提案手法	
	最適化前	最適化後	最適化前	最適化後
1,000 文				
単文	0.106	0.110	0.110	0.117
重文複文	0.076	0.084	0.083	0.086

結果から、最適化したパラメータを用いることで、単文、重文複文共に翻訳精度が向上していることが分かる。また、いずれの翻訳においても提案手法の翻訳精度が高いことが分かる。

## 8 おわりに

本研究では、人手で作成した信頼性の高い翻訳対をプログラムで自動作成したフレーズテーブルに追加し、単文と重文複文における日英、英日翻訳の精度評価を行った。約 13 万の翻訳対を追加し、追加した翻訳対の翻訳確率の重みを最適化した結果、BLEU 値が日英翻訳の単文は 0.3 %、重文複文では 0.8% 向上し、英日翻訳では単文は 1%、重文複文は 2.5% 向上した。結果から、提案手法が有効であることが示された。

### 参考文献

- [1] 鳥バンク <http://unicorn.ike.tottori-u.ac.jp/toribank/>. 2007.
- [2] chasen, <http://chasen-legacy.sourceforge.jp/>
- [3] training-release-1.3.tgz, <http://statmt.org/wmt06/shared-task/baseline>
- [4] SRILM, The SRI Language Modeling Toolkit, <http://www.speech.sri.com/projects/srilm>
- [5] Moses, mooses.2007-05-29.tgz, <http://www.statmt.org/moses/>
- [6] NIST Open MT Scoring, <http://www.nist.gov/speech/tests/mt/>