

日英統計翻訳における英辞郎の効果

東江恵介 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

{s062041,murakami,tokuhisa,ikehara}@ike.tottoru-u.ac.jp

1 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。統計翻訳では、単語の列から単語の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。フレーズテーブルはプログラムにより自動作成される。そのため、フレーズ対のカバー率は高いが、信頼性は低い。

一方、人手で作成されるフレーズ対は信頼性は高いが、フレーズ対の総数が少ないため、カバー率は低い。先行研究 [1] で、プログラムで自動作成されたフレーズテーブルに、人手で作成したフレーズ対を追加し、翻訳精度の向上が確認された。しかし、先行研究では、人手で作成したフレーズ対の総数が少ないため、フレーズテーブルの増加量が少ない。

そこで、本研究では、翻訳対の総数が多い“英辞郎 [2]”を利用し、フレーズテーブルの増加量を多くすることにより、翻訳精度の向上を試みる。なお、本稿では、“英辞郎”のフレーズ対を翻訳対と呼ぶ。

2 日英統計翻訳システム

2.1 概要

日英統計翻訳システムは、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、 $P(e)$ は言語モデルである。

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

あの日		that day		0.05	0.0301124	1	0.0472132
いつか		some day		0.4	0.0081328	0.4	0.0266272
けがで		injury		0.1	0.0181661	1	0.0455217

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。

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 モデルがある。

3 提案手法

本研究では、“英辞郎”で作成した翻訳対を用いて、統計翻訳を行うために、翻訳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“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のみ翻訳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翻訳対を追加する。

3.1 翻訳対の追加手順

プログラムで作成したフレーズテーブルに翻訳対を追加する手順を図 1 に示す。

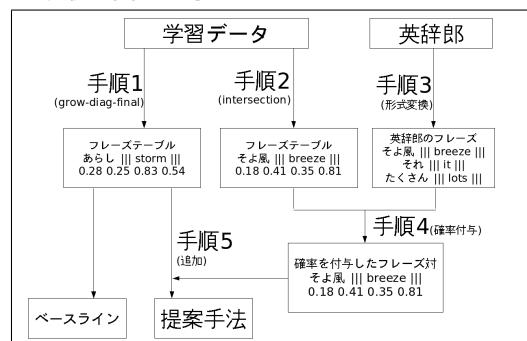


図 1 翻訳対の追加手順

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

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

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

手順 2 で作成したフレーズテーブルを参照して翻訳対に翻訳確率を付与

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翻訳対を追加

なお、本稿では手順 1 のフレーズテーブルを用いて翻訳した結果をベースライン、手順 5 のフレーズテーブルを用いて翻訳した結果を提案手法と呼ぶ。

4 実験環境

4.1 実験データ

4.1.1 英辞郎

“英辞郎”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。翻訳対の例を表 2 に示す。

表 2 翻訳対の例

in every respect : あらゆる点で、すべての点で、 どう見ても、どこまでも、万事に
in a sense : ある意味では、ある程度まで、幾分、 ある点で、一面では / 【用例】 In a sense, what you say is true. : ある意味ではあなたの言っていることは正しい。
on the right track : 正しい方向 {ほうこう} に進んで、 針路を誤っていない、やり方が正しい、妥当で、いい線いってる、 分かりかけている、的はずれではない、良い方向に向かっている

4.1.2 学習データ

単文の実験では、学習データとして、辞書から抽出した単文の対訳文 100,000 文対を用いる。また、重文複文の実験では、学習データとして、辞書から抽出した重文複文の対訳文 100,000 文対を用いる。なお、単文、重文複文それぞれの対訳文には、統計翻訳の前処理として、日本語文には chasen[3] を用いて形態素解析を行い、英語文に対しては、句読点の前後にスペースを入れる。前処理を行った学習データの例を表 3 に示す。

表 3 学習データの例

私は映画を見に行く。 I go to see a movie .
心は経験によって育つ。 The mind expands with experience .
あの人の家はすぐ見つかった。 I soon found that person's house .

4.1.3 テストデータ

単文の日英翻訳のテストデータには、単文 10,000 文対を用いる。重文複文の日英翻訳のテストデータには、重文複文 10,000 文対を用いる。テストデータには統計翻訳の前処理として、日本語文には chasen を用いて形

態素解析を行い、英語文に対しては、句読点の前後にスペースを入れる。前処理を行ったテストデータの例を表 4 に示す。

表 4 テストデータの例

信仰は山をも動かす。 Faith can movemountains .
もっと右へ寄ってください。 Please move over more to the right .
彼の考え方は極端すぎる。 His way of thinking goes too far .

4.2 フレーズテーブルの形式に変換した翻訳対

本研究では、intersection で作成したフレーズテーブルと翻訳対のマッチングを行うため、翻訳対の形式をフレーズテーブルの形式に変換する必要がある。“英辞郎”で作成し、フレーズテーブルの形式に変換した翻訳対の例を表 5 に示す。

表 5 フレーズテーブルの形式に変換した翻訳対の例

あらゆる点で in every respect
ある意味では in a sense
いい線いってる on the right track

今回使用した“英辞郎”には、1,085,344 件の翻訳対がある。なお、統計翻訳の前処理として、翻訳対の日本語フレーズに対して、chasen を用いて形態素解析を行った。

4.3 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、trainphrase-model.perl[4] を用いる。

4.4 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[5]”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。

4.5 デコーダのパラメータ

本研究では、デコーダとして“moses[6]”を用いる。また、パラメータの最適化 [7] 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 weight”は“0.2”とする。

4.6 評価方法

本研究では、出力文の評価として自動評価法“BLEU[8]”と“METEOR[9]”を用いる。なお、入力文 1 文に対して正解文 1 文を用いて評価する。また、人手による対比較実験を行う。対比較実験に関しては 6 章で述べる。

5 翻訳実験

5.1 フレーズテーブル

ベースラインの単文の実験には、単文の学習データ 100,000 文対を用い、grow-diag-final でフレーズテーブルを作成する。また、重文複文の実験には、重文複文の学習データ 100,000 文対を用い、grow-diag-final でフレーズテーブルを作成する。

一方、提案手法の実験には、単文、重文複文それぞれのベースラインのフレーズテーブルに“英辞郎”で作成した翻訳対を追加したフレーズテーブルを用いる。単文でのベースラインのフレーズ数、追加した翻訳対の数、提案手法のフレーズ数を表 6 に示す。また、重文複文でのベースラインのフレーズ数、追加した翻訳対の数、提案手法のフレーズ数を表 7 に示す。

表 6 単文でのフレーズテーブルの総数

ベースライン	翻訳対追加数	提案手法
303,480	18,237	32,1717

表 7 重文複文でのフレーズテーブルの総数

ベースライン	翻訳対追加数	提案手法
235,684	17,581	253,265

“英辞郎”で作成した翻訳対 1,085,344 件のうち、単文では 18,237 件をフレーズテーブルに追加した。また、重文複文では 17,581 件をフレーズテーブルに追加した。翻訳確率を付与した翻訳対の例を表 8 に示す。

表 8 確率を付与した翻訳対の例

あらゆる点で in every respect
0.029 0.002 0.333 0.001
ある意味では in a sense
0.167 0.001 1 0.007
いい線 いったる on the right track
0.1 0.003 0.111 0.001

5.2 日英翻訳精度の評価結果

表 9 と表 10 に日英翻訳精度の評価結果を示す。なお、表 9 の値は BLEU 値であり、表 10 の値は METEOR 値である。

表 9 翻訳精度の評価 (BLEU 値)

テスト文	ベースライン	提案手法
単文	0.118	0.121
重文複文	0.082	0.084

表 10 翻訳精度の評価 (METEOR 値)

テスト文	ベースライン	提案手法
単文	0.360	0.368
重文複文	0.323	0.329

表 9 と表 10 の結果から単文と重文複文ともに、提案手法はベースラインよりも、BLEU 値と METEOR 値が向上していることがわかる。

6 対比較実験

人手による対比較実験を 5.2 節の日英翻訳結果に対して行う。

6.1 評価方法

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 文づつ抽出し、どちら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、それ以外の未知語は存在しないものとして評価を行う。

6.2 判断基準

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、“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、“提案手法×”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は見られなかったが、文質がどちらも変わらない場合は、“文質変化なし”とし、提案手法の出力とベースラインの出力が同じ場合、“同一出力”とする。

6.3 実験結果

表 11 に実験結果を示す。

表 11 対比較実験結果

	提案手法	提案手法×	文質変化なし	同一出力
単文	5	3	4	88
重文複文	9	0	5	86

表 11 から、提案手法は、翻訳精度の向上に効果があることが示された。また、ベースラインの翻訳結果と提案手法の翻訳結果を比較して、26 文に変化が表れた。

6.4 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の翻訳結果より優れていると判断した例を表 12 に示す。

表 12 提案手法がベースラインより優れていると判断した例

日本語文 : 死傷者は数千に達した。
正解文 : Casualties numbered in the thousands .
ベースライン : 死傷 reached the thousands .
提案手法 : casualties have reached the thousands .

6.5 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の翻訳結果より劣ると判断した例を表 13 に示す。

表 13 提案手法がベースラインより劣ると判断した例

日本語文 : その講座は 1 月に終わる。
正解文 : The course finishes in January .
ベースライン : The 講座 in January ends .
提案手法 : The chair is in January .

7 考察

7.1 提案手法の分析

表 11 の対比較実験において、翻訳結果に変化が表れた 26 文中 14 文に対して、提案手法の翻訳結果が優れて

いと判断した。提案手法の翻訳結果が優れていると判断した 14 文を未知語の減少による翻訳品質の向上と文質の向上による翻訳品質の向上に分類した。

7.1.1 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 14 に示す。

表 14 未知語が減少した例

日本語文：あなたが褒美をもらったのは当然なことだ。
英語正解文：You have well earned your reward .
ベースライン：It is quite natural that you have got to 褒美 .
提案手法：It is quite natural that you have got a reward .
日本語文：その音楽家は寂しい晩年を送った。
正解文：The composer spent his last years in loneliness .
ベースライン：The musician sent a 寂しい his life .
提案手法：The musician a lonely sent his life .

7.1.2 文質の向上

ベースラインと比較して、文質が向上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 文質が向上した例

日本語文：金槌と同様なものがある。
英語正解文：There is something similar to a hammer .
ベースライン：As with a 金槌 is something .
提案手法：There is something As with a hammer .
日本語文：車をオーバーホールに出した。
正解文：I had my car overhauled .
ベースライン：オーバーホール stuck in the car .
提案手法：I put the car to overhaul .

7.1.3 未知語が減少した文と文質が向上した文の比較

提案手法が優れていると評価した 14 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 未知語が減少した文と文質が向上した文の比較

テストデータ	未知語減少	文質向上
単文	5	0
重文複文	7	2

未知語が減少した文が 14 文中、12 文と大多数であることがわかる。よって、プログラムで自動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提案手法は、主に未知語の減少に効果があったと言える。

7.2 先行研究との比較

先行研究では単文、重文複文の学習データとして、121,913 文対を用いた。テストデータには、1,000 文対を用いた。先行研究は、単文、重文複文共に、人手で作成したフレーズ対 261,453 件中、130,893 件をフレーズテーブルに追加した。一方、本研究は、単文、重文複文の学習データとして、100,000 文対を用いた。テストデータには、10,000 文対を用いた。本研究は、翻訳対 1,085,344

件中、単文で 18,237 件、重文複文で 17,581 件をフレーズテーブルに追加した。その他の実験環境は同一の環境で実験を行った。表 17 に先行研究の翻訳精度と本研究の翻訳精度を示す。表中の値は BLEU 値である。

表 17 先行研究との翻訳精度比較

テストデータ	先行研究		本研究	
	ベースライン	提案手法	ベースライン	提案手法
単文	0.103	0.106	0.118	0.121
重文複文	0.077	0.080	0.082	0.084

先行研究は単文で 0.3%、重文複文で 0.3% 向上している。一方、本研究では単文で 0.3%、重文複文で 0.2% 向上している。よって、ベースラインと提案手法の差はほぼ同じであることがわかる。

本研究の目的は、翻訳対の総数が多い“英辞郎”を使用し、カバー率を向上させることによる、翻訳精度の向上であった。しかし、先行研究では翻訳対を 130,893 件をフレーズテーブルに追加しているのに対し、本研究では単文では 18,237 件、重文複文では 17,581 件であり、先行研究のおよそ 1/7 しか追加していない。実験結果では、“英辞郎”で作成した翻訳対を追加した時のベースラインと提案手法の差と、先行研究において、人手で作成した翻訳対を追加した時のベースラインと提案手法の差がほぼ同じであった。よって、フレーズテーブルに翻訳対を追加して翻訳するという手法の限界と考えている。

8 おわりに

本研究では、“英辞郎”から作成したフレーズ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の精度評価を行った。単文での翻訳には翻訳対 18,237 件を追加し、重文複文での翻訳には翻訳対 17,581 件を追加した。この結果、BLEU 値が単文では 0.3%、重文複文では 0.2% 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。

今後の課題としては、翻訳確率を付与できなかった翻訳対に対して、パーシャルマッチング [10] を行い、翻訳確率を付与することで、翻訳精度の向上を目指す。

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 (2008)
- [2] 英辞郎 <http://www.alc.co.jp/>
- [3] chasen <http://www.asahi-net.or.jp/~YW3T-TRNS/namazuchi/chasen/index.htm>
- [4] training-rerease-1.3.tgz
- [5] SRILM The SRI Language Model Toolkit <http://www.speech.sri.com/projects/srilm>
- [6] Moses moses-2009-04-13.tgz <http://www.statmt.org/moses/>
- [7] Franz Josef Och “Minimum Error Rate Training in Statistical Machine Translation” Association for Computational Linguistics pp.160-167,(2003)
- [8] NIST Open MT scoring <http://www.nist.gov/speech/tests/mt/>
- [9] The METEOR Automatic Machine Translation Evaluation System <http://www-2.cs.cmu.edu/~alavie/METEOR/>
- [10] Zhongjun He Qun Liu Shouxun Lin “Partial matching strategy for phrase-based statistical machine translation” Association for Computational Linguistics pp.161-164 (2008)