

日英翻訳における IBM Model 1 を用いた未知語処理

川原 宰 村上仁一

鳥取大学大学院 工学研究科 情報エレクトロニクス専攻

{s122019,murakami}@ike.tottori-u.ac.jp

1 はじめに

機械翻訳において、翻訳されない単語が未知語として出力される場合がある。これにより翻訳精度の低下につながるという問題がある。

この未知語を減少させるための手法として、対訳学習文などのパラレルコーパスを増加して行う未知語処理と、増加せずに行う未知語処理の2パターンが考えられる。前者の考えに基づいた日野らの研究 [1] では、新規の対訳句辞書を既存のコーパスに追加して未知語処理を行った。そして、未知語数の削減と翻訳精度向上が報告されている。しかし、日野らの手法では新規のコーパスが必要になるためコストがかかる。後者の考えに基づいた藤原らの研究 [2] では、句に基づく統計翻訳において、フレーズテーブル作成時のヒューリスティクスを併用して未知語処理を行った。そして、未知語数の削減と翻訳精度向上が報告されている。しかし、藤原らの手法は、句に基づく統計翻訳器にしか実装できず、他の機械翻訳器への応用が期待できない。また、川原らの研究 [3] により、句に基づく統計翻訳において、未知語を文字単位に分割し2段階翻訳を行うことで未知語数の削減と翻訳精度向上が報告されている。

そこで、本研究では、他の機械翻訳器への応用を考慮した、新規のコーパスを必要としない新たな未知語処理手法を提案する。具体的には、IBM 翻訳モデル [4] における Model 1 を用いて自動作成した対訳単語辞書による2種類の未知語処理を、一度生成した出力文に対する後処理の形で行う。提案手法は4つの手順で構成される。手順1では、任意の機械翻訳器における出力文から未知語を抽出し、未知語リスト1を作成する。手順2では、IBM Model 1 を用いて対訳学習文から対訳単語を獲得し、対訳単語辞書を作成する。手順3では、未知語リスト1に対して対訳単語辞書を適用(以下、単語翻訳)し、未知語リスト2を作成する。さらに、ここで翻訳できなかった単語の中で、漢字2文字以上の単語(以下、漢字未知語)を文字化し未知語リスト3を作成する。そして、未知語リスト3に対して対訳単語辞書を再度適用(以下、文字翻訳)し、未知語リスト4を作成する。手順3において、対訳単語辞書内に複数の訳語が存在する場合、翻訳確率が最尤のものを選択する。手順4では、未知語処理前の文に対して、未知語リスト4を適用する。以上の手法により未知語を削減し、翻訳精度向上を目指す。

2 提案手法

提案手法の特徴は以下の4点である。

- 新規のコーパスが不必要
- あらゆる機械翻訳器に実装可能
- IBM Model 1 を用いた2種類の未知語処理
- 一度生成した文に対する後処理

提案手法では、IBM 翻訳モデルにおける Model 1 を用いて自動作成した対訳単語辞書による2種類の未知語処理を、一度生成した出力文に対する後処理の形で行う。日英翻訳における提案手法の流れを図1に示す。また、具体的な手順を以下に示す。

- 手順1 任意の機械翻訳器における出力文から未知語を抽出し未知語リスト1を作成
- 手順2 IBM Model 1 を用いて対訳学習文から対訳単語を獲得し、対訳単語辞書を作成
- 手順3 未知語リスト1を単語翻訳し未知語リスト2を作成。さらに漢字未知語(漢字2文字以上の単語)を文字化し未知語リスト3を作成。そして未知語リスト3を文字翻訳し未知語リスト4を作成。(訳語が複数存在する場合は翻訳確率が最尤のものを選択)
- 手順4 未知語処理前の文に対して、未知語リスト4を適用

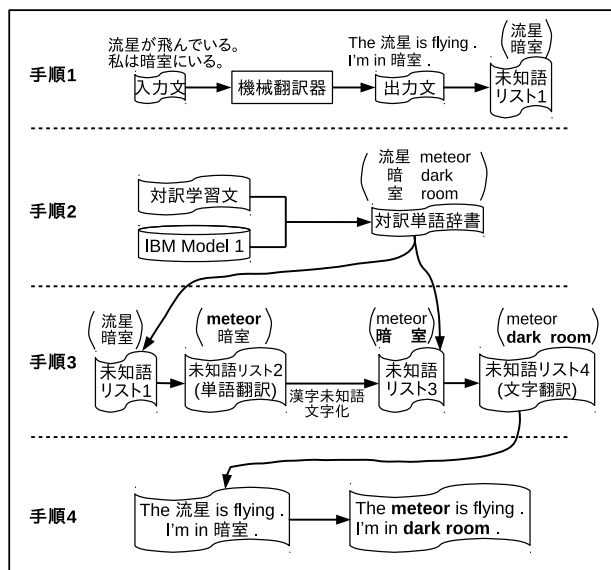


図1 日英翻訳における提案手法の流れ

3 実験環境

3.1 機械翻訳器

本実験では、図1における機械翻訳器に Pattern Based SMT [5] を用いる。

3.2 実験データ

本実験では、電子辞書などの例文より抽出した単文コーパス [6] を用いる。使用するデータの内訳を表1に示す。

表1 実験データ

対訳学習文	160,000 文
入力文	10,000 文

3.3 評価方法

本研究では、未知語処理前と処理後における未知語数と文全体の翻訳精度の比較評価を行う。文全体の翻訳精度の評価として人手評価と自動評価を行う。自動評価には BLEU, METEOR, RIBES を用いる。ここで、未知語処理前の出力文をベースラインとし、未知語処理後の出力文を提案手法とする。

4 実験結果

未知語数と未知語の翻訳精度および文全体の翻訳精度の結果を次節以降に示す。具体的には、第 4.1 節に提案手法全体の実験結果を示し、第 4.2 節に単語翻訳のみの精度調査結果を示す。そして、第 4.3 節に文字翻訳のみの精度調査結果を示す。

4.1 提案手法の実験結果

4.1.1 未知語数

ベースラインと提案手法において、未知語数の調査を行った。調査結果を表 2 に示す。

表 2 未知語数 (10,000 文中)

翻訳手法	未知語を含む文数	未知語数
ベースライン	1,644 文	1,847 単語
提案手法	501 文	521 単語

表 2 より、提案手法において 1,143 文の未知語処理に成功し、未知語数を大幅に削減することができた。この 1,143 文における未知語処理の内訳を表 3 に示す。

表 3 未知語処理の内訳 (1,143 文中)

未知語処理の種類	成功した文数	成功した単語数
単語翻訳	967 文	1,134 単語
文字翻訳	176 文	192 単語

4.1.2 未知語の翻訳精度

表 2 において未知語処理に成功した 1,143 文からランダムに抽出した 100 文において、未知語は 113 単語存在した。この未知語の翻訳精度を調査した。結果を表 4 に示す。

表 4 未知語の翻訳精度 (113 単語中)

正しい翻訳	誤った翻訳
46 単語	67 単語

また、表 4 において、正しい翻訳結果の一例を表 5 に示す。誤った翻訳結果の一例を表 6 に示す。ここで、表中の (A) は単語翻訳による翻訳例である。表中の (B) は文字翻訳による翻訳例である。

表 5 正しく翻訳できた未知語の一例 (46 単語)

(A) 表玄関	entrance
(A) 貴い	valuable
(B) 初球	first ball
(B) 原級	original grade

表 6 正しく翻訳できなかった未知語の一例 (67 単語)

(A) 現存	few
(A) 不純	custody
(B) 威圧	triumphantly compression
(B) 鉱工業	miners plumber business

4.1.3 文全体の翻訳精度 (人手評価)

第 4.1.2 項で抽出した 100 文を用いて人手評価を行った。評価の基準を以下に示す。また、評価結果を表 7 に

示す。

- 提案手法 : 入力文の意味が読み取れる
- 提案手法× : 入力文の意味が読み取れない

表 7 提案手法の評価結果 (100 文中)

提案手法	提案手法×
16 文	84 文

表 7 より、提案手法による文全体の翻訳精度向上が確認できた。

4.1.4 人手評価例

“提案手法”および“提案手法×”の場合における出力例を表 8 と 9 に示す。

表 8 では、ベースラインにおける未知語“ベニス”が正しく翻訳された場合に入力文の意味が読み取れるという状態である。提案手法により“ベニス”が“Venice”に正しく翻訳され、入力文の意味が読み取れるようになったため、“提案手法”とした。

表 8 提案手法の出力例

入力文	彼は晩年をベニスで過ごした。
参照文	He spent the afternoon of his life in Venice.
ベースライン	He spent his late years in ベニス.
提案手法	He spent his late years in Venice.

表 9 では、ベースラインにおける未知語“流言”が正しく翻訳されたとしても入力文の意味が読み取れないという状態である。提案手法により“雨露”が“rain dew”に正しく翻訳されたが、入力文の意味が読み取れないままであるため、“提案手法×”とした。

表 9 提案手法×の出力例

入力文	その小屋はどうか雨露をしのげた。
参照文	The hut kept the rain and dew out after a fashion.
ベースライン	The hut while 雨露 a fashion.
提案手法×	The hut while rain dew a fashion.

4.1.5 文全体の翻訳精度 (自動評価)

未知語処理に成功した 1,143 文を用いて自動評価を行った。表 10 に結果を示す。

表 10 自動評価結果. 精度が高い方を太字で示す

翻訳手法	BLEU	METEOR	RIBES
ベースライン	0.068	0.314	0.661
提案手法	0.072	0.335	0.678

表 10 より、全ての自動評価において、提案手法による精度向上が確認できた。

4.2 単語翻訳のみの精度調査結果

表 3 における単語翻訳の結果 (967 文) に対して精度調査を行った結果を以下に示す。

4.2.1 未知語の翻訳精度

単語翻訳の結果 (967 文) からランダムに抽出した 100 文において、未知語は 109 単語存在した。この未知語の翻訳精度の調査結果を表 11 に示す。

表 11 単語翻訳における未知語の翻訳精度 (109 単語中)

正しい翻訳	誤った翻訳
54 単語	55 単語

4.2.2 文全体の翻訳精度 (人手評価)

第 4.2.1 項で抽出した 100 文を用いて人手評価を行った。評価の基準は第 4.1.3 項と同様である。評価結果を表 12 に示す。

表 12 単語翻訳の評価結果 (100 文中)

単語翻訳	単語翻訳 ×
19 文	81 文

表 12 より、単語翻訳による文全体の翻訳精度向上が確認できた。一方で、表 4 と比べて未知語の翻訳精度が高いにも関わらず、81/100 文は精度が向上しなかった。これは、ベースライン (未知語処理前の出力文) の翻訳精度に大きく依存するためである。したがって、仮に未知語の翻訳精度が 100% であったとしても、文全体の翻訳精度の向上には限界があると考えられる。

4.2.3 文全体の翻訳精度 (自動評価結果)

単語翻訳の結果 (967 文) を用いて自動評価を行った。表 13 に結果を示す。

表 13 自動評価結果. 精度が高い方を太字で示す

翻訳手法	BLEU	METEOR	RIBES
ベースライン	0.068	0.321	0.663
単語翻訳	0.072	0.342	0.680

表 13 より、全ての自動評価において、単語翻訳による精度向上が確認できた。

4.3 文字翻訳のみの精度調査結果

表 3 における文字翻訳の結果 (176 文) に対して精度調査を行った結果を以下に示す。

4.3.1 漢字未知語の翻訳精度

文字翻訳の結果 (176 文) からランダムに抽出した 100 文において、漢字未知語は 104 単語存在した。この漢字未知語の翻訳精度の調査結果を表 14 に示す。ここで、表中の“文字単位で正しい翻訳”とは、文字単位では正しい翻訳であるが単語単位では誤った翻訳であることを意味する。

表 14 文字翻訳における漢字未知語の翻訳精度 (104 単語中)

正しい翻訳	文字単位で正しい翻訳	誤った翻訳
20 単語	16 単語	68 単語

表 14 における文字単位で正しい翻訳結果の一例を表 15 に示す。

表 15 文字単位で正しい翻訳結果の一例 (16 単語)

正 夢	positive dream
緑 青	green blue
甘 党	sweet party
仏 前	France before

4.3.2 文全体の翻訳精度 (人手評価)

第 4.3.1 項で抽出した 100 文を用いて人手評価を行った。評価の基準は第 4.1.3 項と同様である。評価結果を表 16 に示す。

表 16 文字翻訳の評価結果 (100 文中)

文字翻訳	文字翻訳 ×
7 文	93 文

表 16 より、文字翻訳による文全体の翻訳精度向上が確認できた。ただし、表 12 と比べて精度向上した文数は少ない。これは、文字翻訳における漢字未知語の翻訳精度が低いためであると考えられる。

4.3.3 人手評価例

“文字翻訳”および“文字翻訳 ×”の場合における出力例を表 17 と 18 に示す。

表 17 では、ベースラインにおける未知語“大功”が正しく翻訳された場合に入力文の意味が読み取れるという状態である。文字翻訳により“大功”が“great success”に正しく翻訳され、入力文の意味が読み取れるようになったため、“文字翻訳”とした。

表 17 文字翻訳 の出力例

入力文	国家に大功があった。
参照文	There was a great act in the state.
ベースライン	There was a 大功 to nation .
文字翻訳	There was a great success to nation .

表 18 では、ベースラインにおける未知語“日傘”が正しく翻訳されたとしても入力文の意味が読み取れないという状態である。文字翻訳により“日傘”が“day umbrella”となったが、文字単位で正しい翻訳であるため、入力文の意味が読み取れない。したがって“文字翻訳 ×”とした。

表 18 文字翻訳 × の出力例

入力文	彼女は日傘をさした。
参照文	She put up a parasol .
ベースライン	She was 日傘 pointed .
文字翻訳 ×	She was day umbrella pointed .

4.3.4 文全体の翻訳精度 (自動評価結果)

文字翻訳の結果 (167 文) を用いて自動評価を行った。表 19 に結果を示す。

表 19 自動評価結果. 精度が高い方を太字で示す

翻訳手法	BLEU	METEOR	RIBES
ベースライン	0.067	0.284	0.654
文字翻訳	0.069	0.299	0.667

表 19 より、全ての自動評価において、文字翻訳による精度向上が確認できた。

5 考察

5.1 IBM 翻訳モデル

IBM 翻訳モデルには Model 1 ~ Model 5 までの 5 つのモデルが存在する。本研究では、最も基本的な Model 1 を用いて対訳単語辞書を自動作成した。しかし、他のモデルを用いた実験は行なっておらず、対訳単語辞書の精度が変化する可能性が残っている。よって、今後は他のモデルを用いた場合の実験結果との比較が必要であると考えられる。

5.2 未知語の傾向

第 4.1 節で獲得した未知語 (1,847 単語) の傾向を調べるため品詞の分布を調査した。品詞の解析には“MeCab”を用いた。上位 3 位の結果を表 20 に示す。

表 20 未知語の品詞分布

順位	品詞	数
1	名詞	1,311
2	動詞	364
3	副詞	80

表 20 における名詞の中でも特に一般名詞が多く存在し、名詞全体の約 7 割を占めていた。残り 3 割の内、サ変接続名詞が約 2 割を占めており、固有名詞が約 1 割を占めていた。また、表 21 に動詞と副詞の未知語例を示す。

表 21 動詞と副詞の未知語例

品詞	例 1	例 2	例 3
動詞	償う	偲ん	吹き替える
副詞	徐徐に	なにやら	あらためて

5.3 単語翻訳と文字翻訳の効果

単語翻訳と文字翻訳を併用した提案手法全体の結果(表 7)と単語翻訳のみの結果(表 12)を比較すると、単語翻訳のみの方が翻訳精度が高いことが分かる。この原因は以下の 2 点である。1 つ目は、表 7 における 16/100 文が文字翻訳で処理されたが、この 16 文の翻訳精度が向上しなかった点である。2 つ目は、表 11 より、単語翻訳の段階で未知語の翻訳精度が約 50% しかないため、単語数が増加する文字翻訳ではその精度が更に低下し約 20% となった点である。以上より、提案手法全体の試験では、単語翻訳の方が文字翻訳よりも翻訳精度が高くなったと考えられる。したがって、本実験環境においては、単語翻訳と文字翻訳を併用せず、単語翻訳のみで未知語処理を行う方が良いと考える。ただし、本実験よりも多量の対訳学習文を用いた環境で実験した場合には、単語翻訳時における未知語の翻訳精度が向上すると考えられる。したがって、対訳学習文の量が多量であるほど文字翻訳の効果が現れると考える。

5.4 文字翻訳における 2 つの問題

5.4.1 複合名詞の問題

文字翻訳において、3 文字以上の漢字未知語における複合名詞の問題が存在する。例えば、本実験で得られた漢字未知語において“基本給”が存在するが、文字翻訳の結果は“based book salary”となり失敗している。このような場合、本来ならば“基本”と“給”に分割した状態で翻訳する必要があるため、文字翻訳では正しく翻訳することが困難であることが分かる。したがって、単語翻訳の段階で複合名詞を分割する処理が今後必要になると考えられる。実際に、対訳単語辞書において、“基本”と“basic”の対訳が最尤の翻訳確率で存在し、“給”と“salary”の対訳が最尤の翻訳確率で存在しているため、仮に実装できた場合には精度向上が期待できる。

5.4.2 固有名詞の問題

文字翻訳において、固有名詞の問題が存在する。具体例を表 22 に示す。ここでは、固有名詞“苗場”が未知語として出力されている。文字翻訳の結果が“seedlings spot”となっており、文字単位では正しい翻訳であるが単語単位では誤った翻訳である。つまり、未知語が固有名詞である場合、表 22 のような単語単位で誤った翻訳にしかならないため、精度向上に繋がらない。ただし、漢字未知語における固有名詞のみローマ字に変換する処理を追加することで解消できると考える。

表 22 固有名詞の文字翻訳例

入力文	私は苗場へスキーに行った。
参照文	I went skiing at Naeba.
ベースライン	I went to 苗場 to skiing.
文字翻訳	I went to seedlings spot to skiing.

5.5 句に基づく統計翻訳 (Moses) での実験

図 1 における機械翻訳器に Moses を用いた場合の実験を行った。その結果、未知語を含む 941 文の内 361 文の未知語処理に成功した。しかし、人手評価の結果、文全体の翻訳精度はほとんど向上しなかった。ここで、PBSMT では文全体の翻訳精度が向上したが、Moses では文全体の翻訳精度が向上しなかった原因として、未知語処理前の文構造の差が考えられる。本研究は、一度生成された出力文に対して後処理の形で未知語処理を行っている。そのため、未知語処理前の文の翻訳精度に大きく左右される。そして、PBSMT は未知語処理前の文構造が Moses に比べて明白であるため、正しい未知語翻訳ができた場合に文全体の翻訳精度が向上しやすくなる。したがって、PBSMT と Moses の間に差が生じたと考えている。

6 おわりに

本研究では、他の機械翻訳器への応用を考慮した、新規のコーパスを必要としない新たな未知語処理手法を提案した。具体的には、抽出した未知語に対して単語翻訳を行い、その後に漢字未知語を抽出し文字翻訳を行った。実験の結果、PBSMT において 1,143/1,644 文、Moses において 361/941 文の未知語処理に成功した。さらに、人手評価と自動評価の結果より、PBSMT において未知語処理による文全体の翻訳精度の向上が確認できた。また、本実験環境において、単語翻訳は文字翻訳に比べて効果があることが分かった。一方で、未知語処理前の翻訳精度が低い場合には、未知語が正しく翻訳できたとしても文全体の翻訳精度は向上しないことが分かった。今後は、未知語処理前の翻訳精度向上や、文字翻訳における複合名詞と固有名詞の問題解消に取り組みたい。

参考文献

- [1] 日野聡子, 村上仁一, 徳久雅人, 村田真樹: “日英統計翻訳における対訳句コーパスの効果”, 言語処理学会第 18 回年次大会発表論文集, pp491-494, 2012.
- [2] 藤原勇, 村上仁一, 徳久雅人: “句に基づく統計翻訳における未知語処理の 1 手法”, 言語処理学会第 19 回年次大会発表論文集, pp580-583, 2013.
- [3] 川原宰, 村上仁一: “文字ベース翻訳による未知語処理”, 言語処理学会第 22 回年次大会発表論文集, pp207-210, 2016.
- [4] Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, Robert L. Mercer: “The mathematics of statistical machine translation: Parameter Estimation”, Computational Linguistics, 1993.
- [5] 村上仁一: “パターンに基づく統計機械翻訳の概要と問題点について”, 電子情報通信学会技術研究報告, 言語理解とコミュニケーション, NLC2017-3, pp.13-18, 2017.
- [6] 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.