

概要

機械翻訳には様々な手法がある。安場らは、新たな手法として、“ 相対的意味論に基づく変換主導型統計機械翻訳 (以下, TDSMT)[1] を提案した。TDSMT は学習文対を変換し、翻訳を行うため文法性の高い翻訳が期待できる。しかし、TDSMT は、未知語が出現した際に入力文に対する出力文の数 (以下カバー率) を向上させるため、未知語を原言語のまま出力している。このため、出力文中に未知語が日本語のまま出力され、相当する文意が読み取れないという問題がある。

そこで本研究では、TDSMT における未知語の翻訳をニューラル機械翻訳 (以下, NMT) によって行い、翻訳精度の向上を目指す。実験の結果、提案手法は未知語を含む文章の翻訳において、従来手法よりも優れた精度を示した。しかし、未知語を含まない文章の翻訳においては、NMT によって、かえって精度が低下した。

今後は、後編集など、適切な場面でのみ NMT による未知語処理を適用する方法を考える必要がある。

目次

第1章	はじめに	1
第2章	従来の研究	2
2.1	統計翻訳	2
2.1.1	概要	2
2.1.2	単語に基づく統計翻訳	2
2.1.3	IBM 翻訳モデル	2
2.1.4	単語に基づく統計翻訳の問題点	7
2.1.5	GIZA++	8
2.2	句に基づく統計翻訳	9
2.2.1	翻訳モデル	10
2.2.2	フレーズテーブル作成法	11
2.2.3	言語モデル	14
2.2.4	デコーダ	18
2.3	相対的意味論に基づく変換主導統計機械翻訳 (TDSMT) ^[1]	19
2.3.1	TDSMT の手順	20
2.3.2	学習の手順	20
2.3.3	翻訳の手順	23
2.3.4	問題点	25
2.4	Neural Machine Translation	26
2.4.1	概要	26
2.4.2	Encoder-Decoder モデル	27
2.4.3	Attention モデル	28
第3章	未知語処理における NMT の利用	30
3.1	概要	30
3.2	未知語出力用変換テーブルの作成手順	30

3.2.1	NMTによる翻訳	30
3.2.2	翻訳テーブルの作成	31
第4章	実験	33
4.1	実験目的と方法	33
4.2	実験条件	34
4.2.1	実験データ	34
4.2.2	カバー率の調査の実験条件	35
4.2.3	翻訳精度の調査の実験条件	35
4.3	実験結果	36
4.3.1	カバー率調査の結果	36
4.3.2	翻訳精度の調査の結果	37
第5章	考察	40
5.1	翻訳精度の自動評価結果	40
5.2	翻訳精度の人手評価	40
5.2.1	提案手法の方が優れていた出力	40
5.2.2	提案手法と従来手法のどちらが優れているか一概に判断の できない出力	41
5.2.3	従来手法の方が優れていた出力	42
5.3	未知語の出力	43
5.4	従来手法の出力において, 未知語を含んでいた文章における提案手法の評価	43
第6章	おわりに	45

目次

2.1	日英統計翻訳の枠組み	9
2.2	デコーダの動作例	18
2.3	TDSMT の流れ図	24
3.1	出力の流れ	32

表目次

2.1	英日方向の単語対応	8
2.2	日英方向の単語対応	8
2.3	日英方向の単語対応	11
2.4	英日方向の単語対応	11
2.5	intersection の例	12
2.6	union の例	12
2.7	grow-diag の例	13
2.8	grow-diag-final-and の例	13
2.9	対訳単語作成に用いる学習文対	20
2.10	作成される対訳単語	20
2.11	単語レベル文パターンの作成例	21
2.12	変換テーブルの作成例	21
2.13	日本語側変換テーブルの適用例	23
2.14	英語変換テーブルの適用例	23
2.15	未知語が翻訳されていない例	25
2.16	適応された文パターン	25
2.17	変換する学習文対	25
2.18	X1 に適用される変換テーブル	25
2.19	X2 に適用される変換テーブル	25
3.1	連続未知単語での分割の例	31
3.2	未知語出力用変換テーブル	31
4.1	実験データ	34
4.2	学習文対の例	34
4.3	入力文の例	34
4.4	カバー率	36

4.5	実験結果	37
4.6	人手評価の結果	37
4.7	とした例 1	38
4.8	とした例 2	38
4.9	とした例 1	38
4.10	とした例 2	39
4.11	×とした例 1	39
4.12	×とした例 2	39
5.1	○の出力例	40
5.2	適用された文パターンと変換に使用された学習文対	41
5.3	従来手法において適応された変換テーブル	41
5.4	提案手法において適応された変換テーブル	41
5.5	の出力例	42
5.6	×の出力例	42
5.7	適用された文パターンと変換に使用された学習文対	43
5.8	従来手法において X_I に適応された変換テーブル	43
5.9	提案手法において X_I に適応された変換テーブル	43
5.10	人手評価の結果	44
5.11	○の出力例	44

第1章 はじめに

統計翻訳とは、翻訳する言語の文とその対訳学習文を大量の対訳データを使用して、統計的モデルを学習し、自動翻訳システムを構築する機械翻訳の1つである。

TDSMT は学習文対と変換テーブルを用いて翻訳を行う手法である。学習文対を変換し、翻訳を行うため文法性の高い翻訳が期待できる。しかし、変換テーブルの学習文対と入力文が完全一致しなければ変換の適用が不可能なため、入力文の数に対する出力文の数(以下カバー率)が少ない。

出力可能な入力を増やすために、安場らは未知語出力用変換テーブルを自動作成する手法を提案した。しかしながら、この手法では出力文中に未知語が日本語のまま出力される。そのため、相当する文意が読み取れないという問題がある。そこで本研究では、TDSMT における未知語の翻訳をニューラル機械翻訳(以下、NMT)によって行い、翻訳精度の向上を目指す。統計翻訳では、未知語が出現した際に、その出力が得られないが、NMT はあらゆる入力に対して、必ずなんらかの出力が存在する。本研究ではこの性質を未知語処理に利用する。

本論文の構成を以下に示す。第2章で、従来の研究について述べる。第3章で、統計機械翻訳の未知語処理におけるニューラルネットワークの利用について述べる。第4章で、本研究の実験結果を述べる。第5章で、本研究の考察を述べる。

第2章 従来の研究

2.1 統計翻訳

本節は西尾ら [2] の抜粋である。

2.1.1 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文より、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が、翻訳精度は高くなかった。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。

2.1.2 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。日本語単語を英語に翻訳し、日本語単語の語順と同じ並びで英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 翻訳モデルが用いられている。

2.1.3 IBM 翻訳モデル

IBM 翻訳モデルを以下に示す。これは、力久ら [8] の抜粋である。統計翻訳の代表的なモデルとして、IBM の Brown らによる仏英翻訳モデルがある。IBM 翻訳モデルは、単語に基づく統計翻訳を想定して作成された、単語対応の確率モデルである。この翻訳モデルは順に複雑な計算を行うモデル 1 から 5 の 5 つのモデルで構成される。

本章では、原言語であるフランス語文を F 、目的言語である英語文を E として定義する。

IBM モデルでは、フランス語文 E 、英語文 F の翻訳モデル $P(F|E)$ を計算するために、アライメント a を用いる。以下に IBM モデルの基本式を示す。

$$P(F|E) = \sum_a P(F, a|E) \quad (2.1)$$

アライメントとは仏単語と英単語の対応を意味している。IBM モデルのアライメントでは、各仏単語 f に対応する英単語 e は 1 つあり、各英単語 e に対応する仏単語は 0 から n 個ある。また仏単語 f において適切な英単語と対応しない場合、英語文の先頭に空単語 e_0 があると仮定し、その仏 NMT 参考文献単語 f と空単語 e_0 を対応づける。

・モデル 1

(2.1) 式は以下の式に分解することができる。 m はフランス語文の長さ、 a_1^{j-1} はフランス語文における、1 番目から $j-1$ 番目までのアライメント、 f_1^{j-1} はフランス語文における、1 番目から $j-1$ 番目まで単語を表している。

$$P(F, a|E) = P(m|E) \prod_{j=1}^m P(a_j|a_1^{j-1}, f_1^{j-1}, m, E) P(f_j|a_1^j, f_1^{j-1}, m, E) \quad (2.2)$$

(2.2) 式ではとても複雑であるので計算が困難である。そこで、モデル 1 では以下の仮定により、パラメータの簡略化を行う。

- フランス語文の長さの確率 ϵ は m, E に依存しない

$$P(m|E) = \epsilon$$

- アライメントの確率は英語文の長さ l に依存する

$$P(a_j|a_1^{j-1}, f_1^{j-1}, m, E) = (l+1)^{-1}$$

- フランス語の翻訳確率 $t(f_j|e_{a_j})$ は、仏単語 f_j に対応する英単語 e_{a_j} に依存する

$$P(f_j|a_1^j, f_1^{j-1}, m, e) = t(f_j|e_{a_j})$$

パラメータの簡略化を行うことで、 $P(F, a|E)$ と $P(F, E)$ は以下の式で表される。

$$P(F, a|E) = \frac{\epsilon}{(l+1)^m} \prod_{j=1}^m t(f_j|e_{a_j}) \quad (2.3)$$

$$P(F|E) = \frac{\epsilon}{(l+1)^m} \sum_{a_1=0}^l \cdots \sum_{a_m=0}^l \prod_{j=1}^m t(f_j|e_{a_j}) \quad (2.4)$$

$$= \frac{\epsilon}{(l+1)^m} \prod_{j=1}^m \sum_{i=0}^l t(f_j|e_{a_j}) \quad (2.5)$$

モデル 1 では翻訳確率 $t(f|e)$ の初期値が 0 以外の場合, Expectation-Maximization(EM) アルゴリズムを繰り返し行うことで得られる期待値を用いて最適解を推定する. EM アルゴリズムの手順を以下に示す.

手順 1 翻訳確率 $t(f|e)$ の初期値を設定する.

手順 2 仏英対訳対 $(F^{(s)}, E^{(s)})$ (但し, $1 \leq s \leq S$) において, 仏単語 f と英単語 e が対応する回数の期待値を以下の式により計算する.

$$c(f|e; F, E) = \frac{t(f|e)}{t(f|e_0) + \cdots + t(f|e_l)} \sum_{j=1}^m \delta(f, f_j) \sum_{i=0}^l \delta(e, e_i) \quad (2.6)$$

$\delta(f, f_j)$ はフランス語文 F 中で仏単語 f が出現する回数, $\delta(e, e_i)$ は英語文 E 中で英単語 e が出現する回数を表している.

手順 3 英語文 $E^{(s)}$ の中で 1 回以上出現する英単語 e に対して, 翻訳確率 $t(f|e)$ を計算する.

1. 定数 λ_e を以下の式により計算する.

$$\lambda_e = \sum_f \sum_{s=1}^S c(f|e; F^{(s)}, E^{(s)}) \quad (2.7)$$

2. (2.7) 式より求めた λ_e を用いて, 翻訳確率 $t(f|e)$ を再計算する.

$$\begin{aligned} t(f|e) &= \lambda_e^{-1} \sum_{s=1}^S c(f|e; F^{(s)}, E^{(s)}) \\ &= \frac{\sum_{s=1}^S c(f|e; F^{(s)}, E^{(s)})}{\sum_f \sum_{s=1}^S c(f|e; F^{(s)}, E^{(s)})} \end{aligned} \quad (2.8)$$

手順 4 翻訳確率 $t(f|e)$ が収束するまで手順 2 と手順 3 を繰り返す.

・モデル2

モデル1では, 全ての単語の対応に対して, 英語文の長さ l にのみ依存し, 単語対応の確率を一定としている. そこで, モデル2では, j 番目の仏単語 f_j と対応する英単語の位置 a_j は英語文の長さ l に加えて, j と, フランス語文の長さ m に依存し, 以下のような関係とする.

$$a(a_j|j, m, l) \equiv P(a_j|a_1^{j-1}, f_1^{j-1}, m, l) \quad (2.9)$$

この関係からモデル1における (2.4) 式は, 以下の式に変換できる.

$$P(F|E) = \epsilon \sum_{a_1=0}^l \cdots \sum_{a_m=0}^l \prod_{j=1}^m t(f_j|e_{a_j}) a(a_j|j, m, l) \quad (2.10)$$

$$= \epsilon \prod_{j=1}^m \sum_{i=0}^l t(f_j|e_{a_j}) a(a_j|j, m, l) \quad (2.11)$$

モデル2では, 期待値は $c(f|e; F, E)$ と $c(i|j, m, l; F, E)$ の2つが存在する. 以下の式から求められる.

$$c(f|e; F, E) = \frac{t(f|e)}{t(f|e_0) + \cdots + t(f|e_l)} \sum_{j=1}^m \delta(f, f_j) \sum_{i=1}^l \delta(e, e_i) \quad (2.12)$$

$$= \sum_{j=1}^m \sum_{i=0}^l \frac{t(f|e) a(i|j, m, l) \delta(f, f_j) \delta(e, e_i)}{t(f|e_0) a(0|j, m, l) + \cdots + t(f|e_l) a(l|j, m, l)} \quad (2.13)$$

$$c(i|j, m, l; F, E) = \sum_a P(a|E, F) \delta(i, a_j) \quad (2.14)$$

$$= \frac{t(f_j|e_i) a(i|j, m, l)}{t(f_j|e_0) a(0|j, m, l) + \cdots + t(f_j|e_l) a(l|j, m, l)} \quad (2.15)$$

$c(f|e; F, E)$ は対訳文中の英単語 e と仏単語 f が対応付けされる回数の期待値, $c(i|j, m, l; F, E)$ は英単語の位置 i が仏単語の位置 j に対応付けされる回数の期待値を表している.

モデル2では, EM アルゴリズムで計算すると複数の極大値が算出され, 最適解が得られない可能性がある. モデル1では $a(i|j, m, l) = (l+1)^{-1}$ となるモデル2の特殊な場合であると考えられる. したがって, モデル1を用いることで最適解を得ることができる.

・モデル3

モデル3は, モデル1とモデル2とは異なり, 1つの単語が複数対応する単語の繁殖数や単語の翻訳位置の歪みについて考慮する. またモデル3では単語の位置を絶対位置と

して考える．モデル3では以下のパラメータを用いる．

- 翻訳確率 $P(f|e)$
英単語 e が仏単語 f に翻訳される確率
- 繁殖確率 $n(\phi|e)$
英単語 e が ϕ 個の仏単語と対応する確率
- 歪み確率 $d(j|i, m, l)$
英語文の長さ l , フランス語文の長さ m のとき, i 番目の英単語 e_i が j 番目の仏単語 f_j に翻訳される確率

さらに, 英単語が仏単語に翻訳されない個数を ϕ_0 とし, その確率 p_0 を以下の式で求める．このとき, 歪み確率は $\frac{1}{\phi_0!}$ で, $p_0 + p_1 = 1$ で p_0, p_1 は0より大きいとする．

$$P(\phi_0|\phi_1^l, E) = \binom{\phi_1 + \dots + \phi_l}{\phi_0} p_0^{\phi_1 + \dots + \phi_l - \phi_0} p_1^{\phi_0} \quad (2.16)$$

したがって, モデル3は以下の式で求められる．

$$\begin{aligned} P(F|E) &= \sum_{a_1=0}^l \dots \sum_{a_m=0}^l P(F, a|E) \\ &= \sum_{a_1=0}^l \dots \sum_{a_m=0}^l \binom{m - \phi_0}{\phi_0} p_0^{m-2\phi_0} p_1^{\phi_0} \prod_{i=1}^l \phi_i! n(\phi_i|e_i) \\ &\quad \times \prod_{j=1}^m t(f_j|e_{a_j}) d(j|a_j, m, l) \end{aligned} \quad (2.18)$$

モデル3では, 全てのアライメントを計算するため, 計算量が膨大となるので期待値を近似により求める．

・モデル4

モデル4では, モデル3と異なり, 単語の位置を絶対位置ではなく, 相対位置で考える．またモデル3では考慮されていない各単語の位置, 例えば形容詞と名詞の関係を考慮する．モデル4では歪み確率 $d(j|i, m, l)$ を2つの場合で考える．

- 繁殖数が1以上である英単語に対応する仏単語の中で, 最も文頭に近い場合

$$P(\Pi_{[i]1} = j | \pi_1^{[i]-1}, \tau_0^l, \phi_0^l, E) = d_1(j - \odot_{i-1} | \mathcal{A}(e_{[i-1]}), \mathcal{B}(f_j)) \quad (2.19)$$

\odot_{i-1} は $i-1$ 番目の英単語に対応する仏単語の位置を表している．

- それ以外の場合

$$P(\Pi_{[i]k} = j | \pi_{[i]1}^{k-1}, \pi_1^{[i]-1}, \tau_0^l, \phi_0^l, E) = d_{>1}(j - \pi_{[i]k-1} | \mathcal{B}(f_j)) \quad (2.20)$$

$\pi_{[i]k-1}$ は同じ英単語に対応している直前の仏単語を表している .

- モデル 5

モデル 4 では, 単語の位置に関して直前の単語以外は考慮されていない . したがって, 複数の単語が同じ位置に生じたり, 単語の存在しない位置が生成される . モデル 5 では, この問題を避けるために, 単語を空白部分に配置するよう改善が施されている .

- 繁殖数が 1 以上である英単語に対応する仏単語の中で, 最も文頭に近い場合

$$\begin{aligned} P(\Pi_{[i]1} = j | \pi_1^{[i]-1}, \tau_0^l, \phi_0^l, E) \\ = d_1(v_j | \mathcal{B}(f_j), v_{\odot_{i-1}}, v_m - \phi_{[i]} + 1)(1 - \delta(v_j, v_{j-1})) \end{aligned}$$

v_j は j 番目までの空白数, \mathcal{A} は英語の単語クラス \mathcal{B} はフランス語の単語クラスを表している .

- それ以外の場合

$$\begin{aligned} P(\Pi_{[i]k} = j | \pi_{[i]1}^{k-1}, \pi_1^{[i]-1}, \tau_0^l, \phi_0^l, E) \\ = d_{>1}(v_j - v_{\pi_{[i]k-1}} | \mathcal{B}(f_j), v_m - v_{\pi_{[i]k-1}} - \phi_{[i]} + k)(1 - \delta(v_j, v_{j-1})) \end{aligned}$$

2.1.4 単語に基づく統計翻訳の問題点

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, \square は単語が対応した箇所を示す .

表 2.1: 英日方向の単語対応

	He	went	to	kyoto	on	business
彼						
は						
仕事						
で						
京都						
に						
行っ						
た						

表 2.2: 日英方向の単語対応

	He	went	to	kyoto	on	business
彼						
は						
仕事						
で						
京都						
に						
行っ						
た						

表 2.1 は日本語単語 “は” と “に” と “た” に対応する英単語が存在しない．一方で, 表 2.2 は全ての単語に対して対応がとれている．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．よって現在句に基づく統計翻訳が行われている．

2.1.5 GIZA++

GIZA++ 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．

2.2 句に基づく統計翻訳

本節は中村 [9] らの抜粋である。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枠組みを図 2.1 に示す。

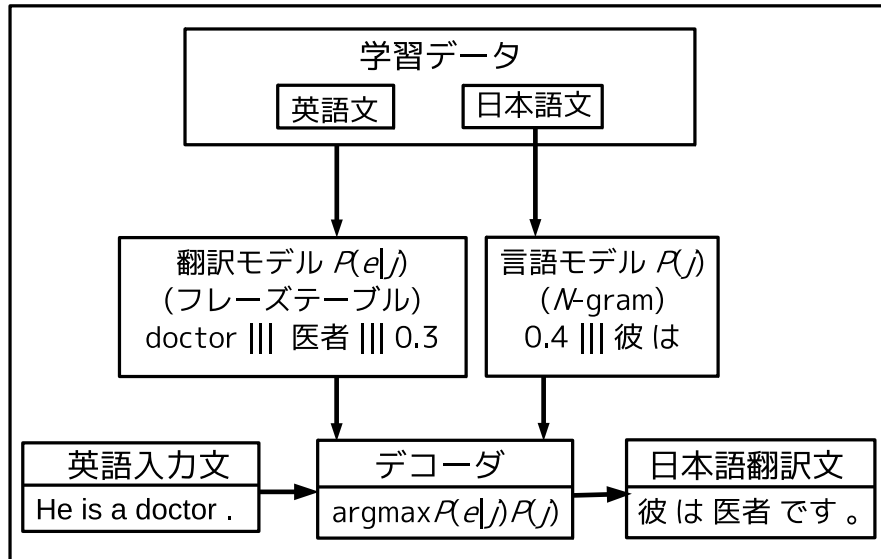


図 2.1: 日英統計翻訳の枠組み

$$E = \operatorname{argmax}_j P(e|j) \quad (2.21)$$

$$\simeq \operatorname{argmax}_j P(j|e)P(e) \quad (2.22)$$

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(英語文と日本語文の対)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.

2.2.1 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英語のフレーズが日本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 —

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

左から英語フレーズ, 日本語フレーズ, 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.

2.2.2 フレーズテーブル作成法

まず, GIZA++を用いて学習文から英日, 日英方向の双方向で最尤な単語アライメントを得る. 英日方向の単語対応の例を表 2.3, 日英方向の単語対応の例を表 2.4 に示す. 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

	He	went	to	kyoto	on	business
彼						
は						
仕事						
で						
京都						
に						
行っ						
た						

表 2.4: 英日方向の単語対応

	He	went	to	kyoto	on	business
彼						
は						
仕事						
で						
京都						
に						
行っ						
た						

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 2.3 と表 2.4 を用いた “intersection” の例を表 2.5, に “union” の例を表 2.6 に示す.

表 2.5: intersection の例

	He	went	to	kyoto	on	business
彼						
は						
仕事						
で						
京都						
に						
行っ						
た						

表 2.6: union の例

	He	went	to	kyoto	on	business
彼						
は						
仕事						
で						
京都						
に						
行っ						
た						

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-diag” の例を表 2.7 に示す .

表 2.7: grow-diag の例

	He	went	to	kyoto	on	business
彼						
は						
仕事						
で						
京都						
に						
行っ						
た						

“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-and”の例を表 2.8 に示す．

表 2.8: grow-diag-final-and の例

	He	went	to	kyoto	on	business
彼						
は						
仕事						
で						
京都						
に						
行っ						
た						

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．

2.2.3 言語モデル

言語モデルとは、人間が用いる言葉の自然な並びを確率としてモデル化したものであり、膨大な量の単言語データを用いて単語の列や文字の列が起こる遷移確率を付与したものである。言語モデルには以下のようなものがある。

N-gram(2.23)

統計翻訳では主に *N*-gram を用いる。tri-gram の式を式 2.23 に示す。

$$\sum_{i=0}^{N-1} \log_2 \frac{\text{count}(E_{i-2}, E_{i-1}, E_i)}{\text{count}(E_{i-2}, E_{i-1})} \quad (2.23)$$

E_i : 英語単語 N : 英文の単語数
 C : 対訳学習文の頻度

実際の計算例を (2.24) に示す。

$$\begin{aligned} & \log_2 P(I \text{ have a dog.}) \\ &= \log_2 \frac{\text{count}(I \text{ have a})}{\text{count}(I \text{ have})} \\ &+ \log_2 \frac{\text{count}(have a \text{ dog})}{\text{count}(have a)} \\ &+ \log_2 \frac{\text{count}(a \text{ dog.})}{\text{count}(a \text{ dog})} \\ &= \log_2 \frac{140}{1,007} + \log_2 \frac{2}{465} + \log_2 \frac{14}{31} \\ &= -11.8545 \end{aligned} \quad (2.24)$$

High order Joint Probability(2.25)

本研究では, 言語モデルに Tri-gram の代わりに High order Joint Probability を使用する. High order Joint Probability を式 2.25 に示す .

$$\sum_{j=0}^{M-1} \sum_{i=0}^{N-1} \text{count}(J_{j-2}, J_{j-1}, J_j, E_{i-2}, E_{i-1}, E_i) \times \log_2 \frac{\text{count}(J_{j-2}, J_{j-1}, J_j, E_{i-2}, E_{i-1}, E_i)}{\text{count}(J_{j-2}, J_{j-1}, J_j) \text{count}(E_{i-2}, E_{i-1}, E_i)} \quad (2.25)$$

J_j : 日本語単語 M : 日本語文の単語数

E_i : 英語単語 N : 英文の単語数

P : 出現確率

実際の計算例を (2.26) に示す . また, 計算式が長くに及ぶため, 第 1 項のみ計算例を示す .

$$\begin{aligned} & P(\text{ぶんごが揺れている。 } The \ swing \ is \ swinging.) \\ & = \text{count}(\text{ぶんごが } The \ swing) \log_2 \frac{\text{count}(\text{ぶんごが } The \ swing)}{\text{count}(\text{ぶんごが})P(The \ swing)} + \dots \\ & = \frac{1}{100,000} \log_2 \frac{\frac{1}{100,000}}{\frac{2}{100,000} \frac{1}{100,000}} + \dots \end{aligned} \quad (2.26)$$

High order Dice(2.27)

$$\sum_{j=0}^{M-1} \sum_{i=0}^{N-1} \log_2 \frac{2 \cdot \text{count}(J_{j-2}, J_{j-1}, J_j, E_{i-2}, E_{i-1}, E_i)}{\text{count}(J_{j-2}, J_{j-1}, J_j) + \text{count}(E_{i-2}, E_{i-1}, E_i)} \quad (2.27)$$

実際の計算例を (2.28) に示す . また, 計算式が長くに及ぶため, 第 1 項のみ計算例を示す .

$$\begin{aligned} & P(\text{ぶらんこが揺れている。 } \textit{The swing is swinging.}) \\ &= \log_2 \frac{2 \cdot \text{count}(\text{ぶらんこが } \textit{The swing})}{\text{count}(\text{ぶらんこが}) + \text{count}(\textit{The swing})} + \dots = \frac{2 \cdot \frac{1}{100,000}}{\frac{2}{100,000} + \frac{1}{100,000}} + \dots \end{aligned} \quad (2.28)$$

High order Log Linear(2.29)

$$\sum_{j=0}^{M-1} \sum_{i=0}^{N-1} \log_2 \left\{ \frac{\text{count}(J_{j-2}, J_{j-1}, J_j, E_{i-2}, E_{i-1}, E_i)}{\text{count}(J_{j-2}, J_{j-1}, J_j)} \times \frac{\text{count}(E_{i-2}, E_{i-1}, E_i, J_{j-2}, J_{j-1}, J_j)}{\text{count}(E_{i-2}, E_{i-1}, E_i)} \right\} \quad (2.29)$$

実際の計算例を (2.30) に示す．また，計算式が長くに及ぶため，第 1 項のみ計算例を示す．

$$\begin{aligned} & P(\text{ぶんこが揺れている。} \quad \textit{The swing is swinging.}) \\ &= \log_2 \left\{ \frac{\text{count}(\text{ぶんこが} \quad \textit{The swing})}{\text{count}(\text{ぶんこが})} \times \frac{\text{count}(\textit{The swing} \quad \text{ぶんこが})}{\text{count}(\textit{The swing})} \right\} \\ &= \log_2 \left\{ \frac{\frac{1}{100,000}}{\frac{2}{100,000}} \times \frac{\frac{1}{100,000}}{\frac{1}{100,000}} \right\} \end{aligned} \quad (2.30)$$

2.2.4 デコーダ

デコーダは、翻訳モデルと言語モデルを用いて、確率が最大となる翻訳候補を探索し、出力を行う変換器のことである。代表的なデコーダとして、“Moses” [13] がある。

入力文として“She is a teacher .” が与えられたときの翻訳例を図 2.2 に示す。

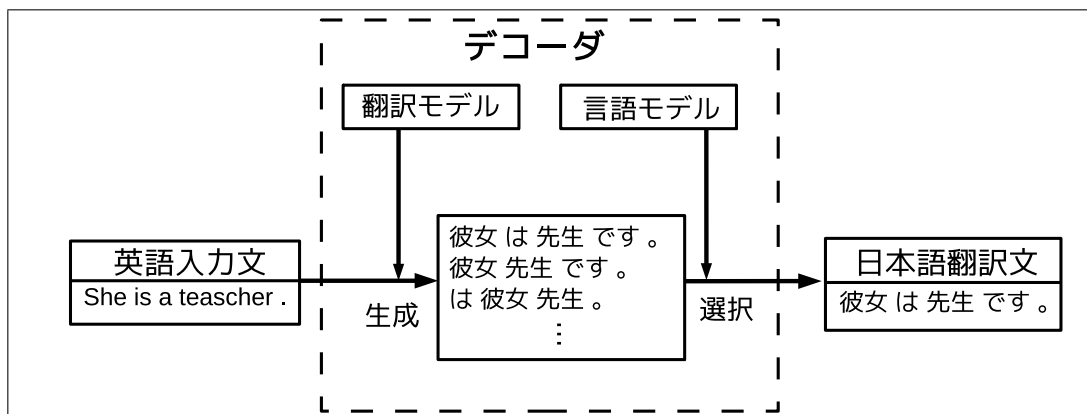


図 2.2: デコーダの動作例

日英統計翻訳において、 $\operatorname{argmax}_e P(e|j)P(j)$ の確率が最大となる英語文を出力するために、適切な順序で日本語と英語の単語対応を得る必要がある。しかし、適切な日本語文を決定するためには、計算量が膨大となり、かつ莫大な時間が必要となる。そこで計算量を削減するために、ビームサーチ法を用いる。

ビームサーチ法とは、翻訳候補の探索において、翻訳確率の低い翻訳候補を枝刈りし、探索範囲を減退する方法である。探索領域の中で一定の確率以上の翻訳候補のみを残し、それ以外の翻訳候補は除外する。

ただし、ビームサーチ法は、切り捨てられた翻訳候補が文章全体で見たときに、最大の確率を持つ翻訳候補であったという可能性がある。そのため選択した翻訳文が最適解であるとは限らないという問題がある。

2.3 相対的意味論に基づく変換主導統計機械翻訳 (TDSMT)^[1]

本節は中村ら [9] の抜粋である. “相対的意味論に基づく変換主導型統計機械翻訳 (TDSMT)” とは, 安場らが提案した機械翻訳の手法の一種である. TDSMT は, 学習文対と, 変換テーブルを用いて, 原言語文を入力とし, 目的言語文を出力する. 変換テーブルは “ A が B ならば C は D ” で表現する. A は学習文対中の原言語句, B は学習文対中の目的言語句, C は入力文中の原言語句, D は出力文中の目的言語句である.

原言語入力文が, 学習文対の原言語側と一致するまで, 入力文と変換テーブル中の AC を照合する. 次に, 一致した学習文対の目的言語側を, 照合した変換テーブルの BD に従って変換し, 目的言語翻訳文を出力する.

TDSMT は適切な学習文対及び, 変換テーブルが存在した場合, 翻訳精度の高い出力文を得ることができる. しかし, TDSMT は変換テーブルを適用した, 入力文が学習文対に完全に一致しない場合は翻訳ができない. 従って, 問題点として, 入力文に対するカバー率が低い.

2.3.1 TDSMT の手順

TDSMT の手順を示す．手順は“学習”と“翻訳”の二部からなる．

2.3.2 学習の手順

TDSMT における学習は“変換テーブルの作成”のみである．本節で作成手順を示す．

手順1 対訳単語の作成

学習文対と対訳単語確率 (IBM Model 1[14]) を利用して、対訳単語を作成する．このとき付与される対訳単語確率を P_w とする．例として、表 2.9 に示す学習文対を使用して、表 2.10 に示す対訳単語を作成する．表 2.10 の値は例であり、実際の数値とは異なる．

表 2.9: 対訳単語作成に用いる学習文対

学習文対 (日本語側)	彼の弟は学生だ。
学習文対 (英語側)	His brother is a student.

表 2.10: 作成される対訳単語

	日本語単語	英語単語	P_w
対訳単語 1	彼	His	0.4
対訳単語 2	弟	brother	0.7
対訳単語 3	学生	student	0.6

手順2 単語レベル文パターンの作成

学習文対内で対訳単語に当たる部分を変数化し、単語レベル文パターンを作成する。例を表 2.11 に示す。

表 2.11: 単語レベル文パターンの作成例

学習文対 (日本語側)	彼の兄は医者だ。
学習文対 (英語側)	His brother is a doctor.
単語レベル文パターン (日本語側)	$X0$ の $X1$ は $X2$ だ
単語レベル文パターン (英語側)	$X0$ $X1$ is a $X2$

手順3 変換テーブルの作成

変換テーブルは「A が B ならば C は D」という形式である。「A が B」には文パターンの作成の際に利用した対訳単語が当たる。そして、「C は D」には文パターン原文とは別の対訳文に文パターンを当てることによって抽出した日本語句と英語句の対が当たる。表 2.12 では変数 $X2$ の部分から変換テーブル“「学生」が「student」ならば「教師」は「teacher」”が得られる。

表 2.12: 変換テーブルの作成例

学習文対 (日本語側)	彼の弟は学生だ。
学習文対 (英語側)	His brother is a student.
単語レベル文パターン (日本語側)	$X0$ の $X1$ は $X2$ だ。
単語レベル文パターン (英語側)	$X0$ $X1$ is a $X2$.
照合する学習文対 (日本語側)	私の母は教師だ。
照合する学習文対 (英語側)	My mother is a teacher.
変換テーブル ($X2$)	A:学生 B:student C:教師 D:teacher

手順4

計算時間短縮のため、作成した変換テーブルの枝刈りを行う。具体的には、作成した変換テーブルが以下の条件に当てはまる場合に削除する。

- 「A は B」の共起頻度が 1 回以下
- 「C は D」の共起頻度が 1 回以下
- 「A は B」の英日翻訳の単語翻訳確率の順位が 4 位以下 (IBM model 1 による確率を参照)

- 「A は B」の日英翻訳の単語翻訳確率の順位が 4 位以下 (IBM model 1 による確率を参照)
- 「C は D」の英日翻訳の単語翻訳確率の順位が 64 位以下 (IBM model 1 による確率を参照)
- 「C は D」の日英翻訳の単語翻訳確率の順位が 16 位以下 (IBM model 1 による確率を参照)

手順 4 変換テーブルに確率を付与

対訳単語確率 P_w を利用し, 変換テーブルに確率を付与する . この確率を変換テーブル確率 P_v とする .

1. 変換テーブルの CD に存在する全ての日英単語の組み合わせを確認する .
2. 日本語単語に対応する英語単語の中で, 対訳単語確率 P_w の最大値を得る .
3. 各日本語単語について得られた値と, 変換テーブルの AB の対訳単語確率 P_w について, 対数の総和を求める .

2.3.3 翻訳の手順

本節で TDSMT における翻訳の手順を示す．入力文を「私の姉は教師だ。」とする．

手順 1 入力文に日本語側の変換テーブルを適用

変換テーブルの C と A を利用して, 入力文を学習文対の日本語側と一致させる．表 2.13 では入力文中の「教師」を「生徒」に変換する．

表 2.13: 日本語側変換テーブルの適用例

入力文	私の姉は教師だ。
変換テーブル: C	教師
変換テーブル: A	生徒
一致する学習文対(日本語側)	私の姉は生徒だ。

手順 2 学習文対に英語側の変換テーブルを適用

手順 1 と同じ変換テーブルの B と D を学習文対の英語側に適用し, 出力候補文を作成する．表 2.14 では学習文対中の「student」を「teacher」に変換している．

表 2.14: 英語変換テーブルの適用例

一致した学習文対(日本語側)	私の姉は生徒だ。
一致した学習文対(英語側)	My sister is a student.
変換テーブル: B	student
変換テーブル: D	teacher
出力候補文	My sister is a teacher.

手順 3 最終的な出力文の決定

複数の出力候補文が得られた場合, 計算式 (2.31) に従って, 最終的な出力文を決定する．ここで P_m は言語モデルの確率である．

$$\log P = \log P_v + \log P_m \quad (2.31)$$

図 2.3 に TDSMT の流れ図を示す.

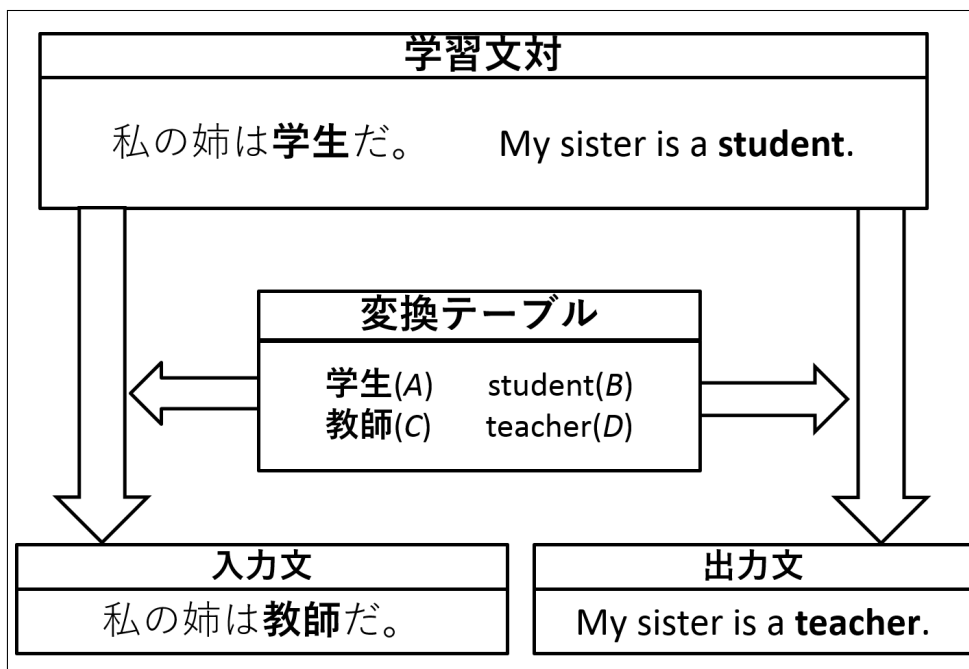


図 2.3: TDSMT の流れ図

2.3.4 問題点

従来手法の問題点は、出力文中の未知語の翻訳は行われておらず、入力文中における未知語に相当する部分の意味が読み取れない点である。出力文中の未知語が翻訳されていない例を表 2.15 に示す。

表 2.15: 未知語が翻訳されていない例

入力文	彼は麻薬所持の疑いで検挙された。
参照文	He was arrested for having drugs .
出力文	He was arrested on suspicion of <u>麻薬所持</u> .

上記のような出力がされる流れを以下に示す。

1. 表 2.15 の入力文に対して表 2.16 の文パターンが適応される。

表 2.16: 適応された文パターン

日本語側	彼は $X1$ の疑いで $X2$ された。
英語側	He was $X2$ on suspicion of $X1$.

2. 表 2.17 の学習文対を表 2.16 の文パターンにしたがって変換し、学習文対に一致させる。

表 2.17: 変換する学習文対

日本語側	彼は詐欺の疑いで逮捕された。
英語側	He was arrested on suspicion of fraud .

3. $X1$ に変換テーブル表 2.18 を適用し、学習文対と入力文の日本語側を一致させる。

表 2.18: $X1$ に適用される変換テーブル

A	逮捕	B	arrested
C	検挙	D	arrested

4. $X2$ に変換テーブルを適用する。この際、変換テーブルの A にあたる「詐欺」と C にあたる「麻薬所持」を持つ変換テーブルが存在しない。よって、未知語出力用変換テーブル表 2.19 を適用し、学習文対と入力文の日本語側を一致させる。

表 2.19: $X2$ に適用される変換テーブル

A	詐欺	B	fraud
C	麻薬所持	C	麻薬所持

以上により、出力文中の未知語が翻訳されずに出力される。

2.4 Neural Machine Translation

本節は今仁ら [10] の抜粋である.

2.4.1 概要

ニューラル機械翻訳 (NMT) とは, 近年提案された機械翻訳の手法である. 多くの細かい構成要素によって成立する従来の統計翻訳のシステムとは異なり, NMT は巨大なニューラルネットワークを一つ (もしくは少数) 用いてシステムを構築する. NMT の手法には, Encoder-Decoder モデルおよびそれを拡張した Attention モデルが提案されている. Encoder-Decoder モデルは入力の系列を固定長のベクトルに符号化 (Encode) し, 固定長のベクトルより出力の系列を復号化 (decode) するモデルである. 機械翻訳のタスクにおいて, Encoder-Decoder モデルの入出力系列の要素は単語のベクトル表現となる. Attention モデルは Encoder-Decoder モデルにおいて出力系列を生成する際に, 入力系列の参照を行う機構 (Attention) を用いた手法であり, Encoder-Decoder モデルと比較して, 長い入力に対するより妥当な出力を得られるとされている.

2.4.2 Encoder-Decoder モデル

Encoder-Decoder モデル [4] は 2 つのリカレントニューラルネットワーク (RNN) により構成されるニューラル機械翻訳のモデルである。1 つの RNN は入力系列を一つの固定長のベクトルに符号化 (encode) し、もう一方の RNN により固定長のベクトル符号を出力の系列へと復号化 (decode) する。Encoder-Decoder モデルでは、同時に 2 つの RNN の訓練を行い、入力の原言語文に対して出力の目的言語文を得る条件付き確率を最大化する。

Encoder-Decoder モデルの枠組み

Encoder-Decoder モデルでは、Encoder は入力の系列 $x = (x_1, \dots, x_{T_x})$ をベクトル c に符号化する。一般的な Encoder-Decoder モデルでは 2.32 式および 2.33 式で表される RNN を用いて構成される。

$$h_t = f(x_t, h_{t-1}) \quad (2.32)$$

$$c = q(\{h_1, \dots, h_{T_x}\}) \quad (2.33)$$

ここで h_t は各時刻 t の隠れ層の状態であり、 c は隠れ層を用いて生成されたベクトルである。 f および q は活性化関数であり、Sutskever ら [5] は f に LSTM を用いた上、 $q = (f\{h_1, \dots, h_T\}) = h_T$ としている。Decoder は文脈ベクトル c と既に生成された単語 $f(y_1, \dots, y_{t-1})$ が与えられた際の次の単語 y_t を予測するように訓練され、結合確率を 2.34 式に示す条件式に分解することで翻訳文 $y = (y_1, \dots, y_{T_y})$ を得る条件付き確率を定義している。

$$p(y) = \prod_{t=1}^{T_y} p(y_t | \{y_1, \dots, y_{t-1}\}, c) \quad (2.34)$$

2.33 式および 2.34 式で表される RNN を用いて、それぞれの条件付き確率は 2.35 式によりモデル化される。ここで、 g は非線形の多層関数であり、 y_t の確率を生成する。 s_t は RNN の隠れ層となる。

$$p(y_t | \{y_1, \dots, y_{t-1}\}, c) = g(y_{t-1}, s_t, c) \quad (2.35)$$

2.4.3 Attention モデル

Encoder-Decoder モデルの問題は、原言語文中の全ての情報を一つの固定長のベクトルに圧縮する点である。Encoder-Decoder モデルでは対訳学習文中で用いられている文よりも長い原言語文が入力された場合に、極端に精度が低下することが報告されている。これは、長い原言語文の全情報が一つの固定長のベクトルに圧縮されるために、目的言語文を生成する際に必要な情報が損失することが原因である。Attention モデル [6] はこの問題を改善するために提案された NMT の手法である。Attention モデルの Encoder では入力単語を前後両方向から RNN に渡す手法 (bidirectionalRNN) を用いている。Encoder は、入力文を前から読み込んだノードと後ろから読み込んだノードを組み合わせることで各単語を符号化 (encode) する。Decoder では、モデルが出力文中のある単語を生成する際に、その単語が最も関連する原言語文中の単語に相当する符号を探索する。その後、探索により得られた原言語単語の文中の位置情報を有した文脈ベクトル、および既に生成された全ての目的言語単語を参照し、次の目的言語単語を予測する。Attention モデルは Encoder-Decoder モデルと異なり、原言語文の情報を一つの固定長のベクトルに圧縮せずに目的言語文中の単語生成時に参照することで、より長い文における精度の向上を実現している。

Attention モデルの枠組み

Attention モデルでは 2.36 式の条件付き確率を 2.37 式により定義する .

$$p(y_i | \{y_1, \dots, y_{i-1}\}, c) = g(y_{i-1}, s_i, c_i) \quad (2.36)$$

ここで, s_i は時刻 i での隠れ層の状態であり, 2.37 式により計算される .

$$s_i = f(s_{i-1}, y_{i-1}, c_i) \quad (2.37)$$

確率 s_i は, Encoder-Decoder モデルと異なり, 各目的言語単語 y_i について文脈ベクトル c_i により状態付けられる . 文脈ベクトル c_i はアノテーション系列 (h_1, \dots, h_{T_x}) に依存し, Encoder により入力文と対応付けられる . 各アノテーション h_i は特に入力文中の i 番目の単語付近の情報を強く保有しており, さらに全入力文の情報を保持している . また, 2.38 式に示す文脈ベクトル c_i はアノテーション h_i の重み付き和により計算される .

$$c_i = \sum_{j=1}^{T_x} a_{ij} h_j \quad (2.38)$$

各アノテーション h_j の重み a_{ij} は

$$a_{ij} = \frac{\exp e_{ij}}{\sum_{k=1}^{T_x} \exp e_{ik}} \quad (2.39)$$

により計算される . ここで

$$e_{ij} = a(s_{i-1}, h_j) \quad (2.40)$$

は j 番目付近の入力の対応および i 番目の出力の適合を示す値であり, アライメントモデルと呼ばれる . この値は y_i を出力する直前の RNN の隠れ層の状態 s_i および入力文中の j 番目のアノテーション h_j に基づいている .

第3章 未知語処理におけるNMTの利用

3.1 概要

本研究ではTDSMTにおいて、未知語が出現した際にNMTを利用して変換テーブルを作成する手法を提案する。具体的には、未知語が出現した際にその単語にNMTを利用して翻訳を行い、その結果を変換テーブルの D (出力文内の英語句)に格納する。翻訳時にこの未知語出力用変換テーブルを利用することで未知語が英語で出力されることが期待できる。

3.2 未知語出力用変換テーブルの作成手順

以下に提案手法により未知語出力用変換テーブルを作成し、入力文を翻訳する流れを示す。

3.2.1 NMTによる翻訳

手順1 従来手法と同様に学習文対から対訳単語を作成する。

手順2 NMTに手順1で作成した対訳単語を学習させ翻訳モデルを作成する。

手順3 入力文を連続未知単語として分割する。具体的には入力文を1単語、2単語連続、3単語連続ごとに分割する。連続未知単語の例を表3.1に示す。

手順4 手順3で作成した入力文の各連続未知単語をNMTを用いて翻訳を行う。入力文の各連続未知単語の翻訳の例を表3.1に示す。

表 3.1: 連続未知単語での分割の例

入力文	
私の父は漁師だ。	
連続未知単語	NMTの翻訳結果
私	I
私 の	my
私の父	My father
の	of
の父	father
の父は	father is

3.2.2 翻訳テーブルの作成

手順 5 NMTの翻訳結果を変換テーブルのDにあたる部分に格納し未知語出力用変換テーブルを作成する。作成したテーブルを変換時に使用するテーブルの候補に追加する。翻訳テーブル作成の例を表 3.2 に示す。

表 3.2: 未知語出力用変換テーブル

対訳単語	
医者	doctor
入力文	
私の父は漁師だ。	
未知語出力用変換テーブル	
A: 医者	B: doctor
C: 私	D: I
A: 医者	B: doctor
C: 私 の	D: My
A: 医者	B: doctor
C: 私の父	D: My father
etc...	

手順 6 変換テーブルを使用して従来手法と同じ手法で翻訳を行う。

図 3.1 に未知語処理を行い、翻訳を行う流れを示す。

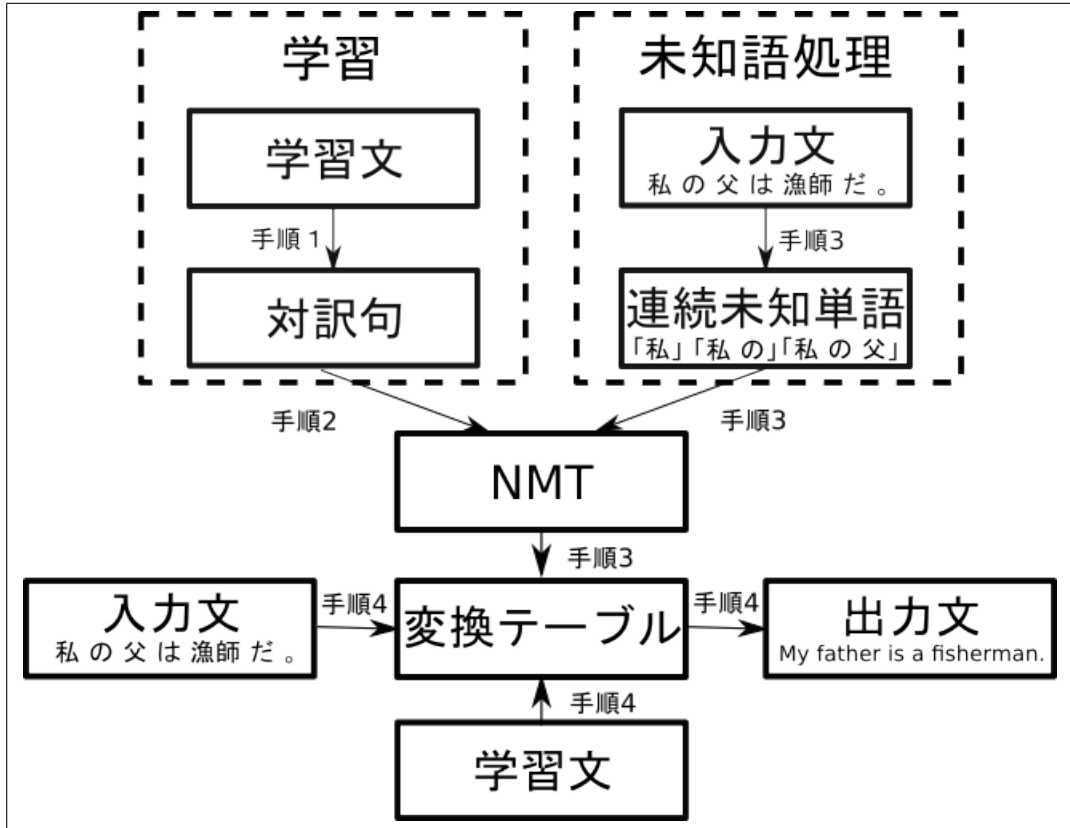


図 3.1: 出力の流れ

第4章 実験

4.1 実験目的と方法

従来手法と提案手法の比較実験により, 提案手法の有効性を調査する. 翻訳精度の調査は自動評価と人手評価で行う. 自動評価にはBLEU[15], METEOR[16], TER[17]を用いる.

4.2 実験条件

4.2.1 実験データ

本研究では, TDSMT の学習文対および翻訳実験に用いる入力文として, 電子辞書などの例文より抽出した単文データを用いる [18] . NMT の学習文対として, TDSMT の学習文対から抽出した対訳単語を用いる. データの内訳を表 4.1 に示す . 自動評価には BLEU[15], METEOR[16], TER[17] を用いる .

表 4.1: 実験データ

	TDSMT の学習	NMT の学習
学習文対	159,998 文	対訳単語 1,065,615 対
ディベロップ文	/	対訳単語 10,871 対
テスト文	100 文	

学習文対および, 入力文の例を表 4.2 と表 4.3 に示す .

表 4.2: 学習文対の例

学習文対	
日本語原文	英語原文
ペンが椅子の下を転がった。	The pen rolled under the chair .
原告は敗訴した。	The plaintiff was cast .
せわしい毎日を過ごしています。	My days have been busy

表 4.3: 入力文の例

入力文	
日本語文	参照文
ピアノの勉強にヨーロッパに行く。	Go to Europe to study the piano .
この季節としてはよい天気だ。	It's good weather for this time of year .
彼はその白馬を追いかけた。	He went after the white horse .

4.2.2 カバー率の調査の実験条件

従来手法と提案手法を用いて, 入力文を 100 文として, 得られた出力文数及び, 出力文に日本語を含む文章の数を調査する.

4.2.3 翻訳精度の調査の実験条件

従来手法と提案手法を用いて, 得られた出力文を自動評価と人手評価で評価する. また, 人手評価は対比較評価で行う. 人手評価の基準は次の 3 つである.

- : 提案手法の方が優れている.
- : どちらが優れているか一概に判断できない.
- × : 従来手法の方が優れている.

4.3 実験結果

4.3.1 カバー率調査の結果

カバー率調査の結果を 4.5 に示す.

表 4.4: カバー率

	入力文	出力文	未知語（日本語）を 含む出力文
従来手法	100	84	44
提案手法	100	84	11

提案手法により未知語（日本語）を含む文章が 44 文から 11 文に減少した. この結果から, 提案手法により未知語（日本語）を含む出力を削減できた.

4.3.2 翻訳精度の調査の結果

4.3.2.1 自動評価の結果 (100 文)

TDSMT で出力を得られた文の内, 100 文を対象に自動評価を行った結果を表 4.5 に示す .

表 4.5: 実験結果

	BLEU	RIBES	TER
従来手法	0.059	0.637	0.818
提案手法	0.076	0.677	0.756

表 4.5 より従来手法と提案手法の翻訳精度にほぼ差はない .

4.3.2.2 人手評価の結果 (100 文)

TDSMT で出力を得られた文の内, 100 文を対象に対比較評価を行った結果を表 4.6 に示す .

表 4.6: 人手評価の結果

	×		
	14	57	12

表 4.6 より従来手法と提案手法の翻訳精度にほぼ差はない. 評価例を表 4.7, 表 4.8, 表 4.9, 表 4.10, 表 4.11, 表 4.12 に示す . 表中の下線部は評価の根拠とした部分である .

表 4.7: とした例 1

入力文	彼は黄疸の症状が出た。
参照文	He had symptoms of jaundice .
従来手法	He became 黄疸の症状 .
提案手法	He became <u>an icteric</u> .

表 4.7 の従来手法と提案手法の出力を比較すると、従来手法では「黄疸の症状」は日本語で出力されている。一方で、提案手法では「an icteric」という正しい英語に翻訳されて出力されている。また、それ以外の部分については同一である。従って、提案手法が従来手法よりも優れていると判断した。

表 4.8: とした例 2

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 .
従来手法	He spent his late years in ベニス .
提案手法	He spent his late years in <u>to Venice</u> .

表 4.8 の従来手法と提案手法の出力を比較すると、従来手法では「ベニス」は日本語で出力されている。一方で、提案手法では「to Venice」という英語で出力されており、文法的には誤っているがベニスという地名が英語で表現されており、提案手法が従来手法よりも優れていると判断した。

表 4.9: とした例 1

入力文	献身的な夫であり父親だった。
参照文	He was a devoted husband and father .
従来手法	He <u>あり 父親 だっ a dedication husband</u> .
提案手法	He <u>was father unfortunate husband her husband</u> .

表 4.9 の従来手法と提案手法の出力を比較すると、従来手法では「あり 父親 だっ」が日本語で出力された。一方で、提案手法では、「was father」という英語で出力された。しかし、「献身的な夫」の翻訳が、従来手法では「dedication husband」となっており、正しい翻訳が為されているのに対して、提案手法では「unfortunate husband her husband」となっており、同じ意味が連続して出現する形になっている。従って、従来手法と提案手法、どちらが優れているかは一概に判断できないものとした。

表 4.10: とした例 2

入力文	油圧 系統 が 漏れる。
参照文	There is a leak in the hydraulic system .
従来手法	The 油圧 系統 escapes .
提案手法	The 油圧 系統 escapes .

表 4.10 の出力は, 従来手法と提案手法の出力が同一であった.

表 4.11: ×とした例 1

入力文	最後に 皆で 校歌 を うたった。
参照文	At the end we all sang our school song together .
従来手法	He sang his 皆で 校歌 <u>in the end</u> .
提案手法	I was the school in on a last to .

表 4.11 の従来手法と提案手法の出力を比較すると, 従来手法では「最後に を歌った」という文章の大まかな流れは理解できる. それに対して, 提案手法では文意が全く判断できない. このため, 従来手法の方が優れていると判断した.

表 4.12: ×とした例 2

入力文	スキー シーズン には 臨時 列車 が 運転 される。
参照文	Extra trains run during the ski season .
従来手法	スキー シーズン に <u>trains are run extraordinary</u> .
提案手法	skiing season trains are run is driven .

表 4.12 の従来手法と提案手法の出力を比較すると, 従来手法では「スキーシーズン」の意味は読み取れないが「臨時列車が運転される」という文章の大まかな意味は読み取れる. これに対して, 提案手法では「スキーシーズン」の部分は翻訳されているが全体的に文法が誤っており, 意味が読み取れない. 従って, 提案手法の方が優れていると判断した.

第5章 考察

5.1 翻訳精度の自動評価結果

表 4.5 より, 提案手法では従来手法よりも高い精度となった. 特に BLEU は上昇幅が大きくなった. これは従来手法では日本語単語だった部分が提案手法によって英単語に翻訳されたことで単語の対応が取れるようになったためだと考えられる.

5.2 翻訳精度の人手評価

人手評価は, 提案手法では従来手法との精度の違いは無かった.(表 4.6)

5.2.1 提案手法の方が優れていた出力

提案手法の方が優れていた出力の例を表 5.1 に示す. この出力の流れを解析し, 考察する.

表 5.1: ○の出力例

入力文	この機会に徹底的な論議を期待したい。
参照文	It is to be hoped that the panel members will take this opportunity to thoroughly debate the issue .
従来手法	This want hope 徹底的な論議 the in my opportunity .
提案手法	This want hope thorough discussion in my opportunity .

適応された文パターン及び変換に使用された学習文対を表 5.2 に示す. 従来手法では表 5.3 の変換テーブルによって未知語「徹底的な議論」が日本語のまま出力されている. 提案手法では表 5.4 の変換テーブルによって英語で出力されている. したがって, 提案手法で追加されたテーブルによって出力の精度が向上した.

表 5.2: 適用された文パターンと変換に使用された学習文対

学習文対	日本語側	私は胸に鋭い痛みを感じた。
	英語側	I felt a sharp pain in my chest .
文パターン	日本語側	X1 X2 に X3 X4 X5 X6 。
	英語側	X1 X6 X5 X3 X4 in my X2 .

表 5.3: 従来手法において適応された変換テーブル

X3			
A	鋭い	B	sharp
C	徹底的な議論	C	徹底的な議論
X4			
A	痛み	B	pain
C	を	D	the

表 5.4: 提案手法において適応された変換テーブル

X3			
A	鋭い	B	sharp
C	徹底的な	D	thorough
X4			
A	痛み	B	pain
C	議論を	D	discussion

5.2.2 提案手法と従来手法のどちらが優れているか一概に判断のできない出力

提案手法と従来手法のどちらが優れているか一概に判断のできない出力の例を表 5.5 に示す。従来手法で未知語だった「あり父親 だっ」は提案手法で「was father」へと翻訳されている。しかし、「献身的な父」は従来手法で「dedication husbnd」という正しい出力が為されているが提案手法では、「unfortunate husband husband」という誤った出力をしている。よって と評価した。

表 5.5: の出力例

入力文	献身的な夫であり父親だった。
参照文	He was a devoted husband and father .
従来手法	He あり 父親 だっ a dedication husband .
提案手法	He was father unfortunate husband her husband .

5.2.3 従来手法の方が優れていた出力

従来手法の方が優れていた出力の例を表 5.6 に示す。この出力の流れを解析し、考察する。

表 5.6: ×の出力例

入力文	良心が彼女を苦しめた。
参照文	Her conscience stung her .
従来手法	She suffered from a conscience .
提案手法	She suffered from a worried her .

適用された文パターン及び変換に使用された学習文対を表 5.7 に示す。従来手法では「良心」を表 5.8 の変換テーブルにより、「conscience」という正しい翻訳が成されていた。しかし、提案手法では、表 5.9 の提案手法による未知語出力用変換テーブルを使用して翻訳を行なっている。その結果「worried her」という誤った出力がされている。つまり、提案手法で追加したテーブルにより精度が低下している。

表 5.7: 適用された文パターンと変換に使用された学習文対

学習文対	日本語側	頭痛が彼女を苦しめた。
	英語側	She suffered from a headache .
文パターン	日本語側	<i>XI</i> が彼女を苦しめた。
	英語側	She suffered from <i>XI</i> .

表 5.8: 従来手法において *XI* に適応された変換テーブル

<i>A</i>	頭痛	<i>B</i>	headache
<i>C</i>	良心	<i>D</i>	conscience

表 5.9: 提案手法において *XI* に適応された変換テーブル

<i>A</i>	頭痛	<i>B</i>	headache
<i>C</i>	良心	<i>D</i>	worried her

5.3 未知語の出力

表 4.5 より, 日本語の単語を含む出力は従来手法では, 44 文であった. 提案手法では 11 文となった. このため出力文中での日本語単語の数は削減されたことが確認できる.

提案手法で日本語単語が出力された原因を考察する. 今回の実験では従来手法でも変換テーブルを作成し, それを追加している. そのため提案手法で作成した未知語出力用変換テーブルよりも従来手法で作成した未知語出力用変換テーブルの方が適用確率が高い場合, 出力文中に日本語単語が出力される. 従来手法でのテーブル作成を行わなければ, 出力文中に原言語は出現されなくなるがカバー率が低下すると考えられる.

5.4 従来手法の出力において, 未知語を含んでいた文章における提案手法の評価

入力文において, 従来手法で翻訳を行った際に出力文に日本語が含まれていた文に着目し, 本手法の未知語処理における有用性を検証する. 人手評価の結果を表 5.10 に示す. 表 5.11 に と判断した出力の具体例を示す.

表 5.10: 人手評価の結果

	×
10	32
	2

表 5.11: ○の出力例

入力文	彼は麻薬所持の疑いで検挙された。
参照文	He was arrested for having drugs .
従来手法	He was arrested on suspicion of 麻薬 所持 .
提案手法	He was arrested on suspicion of accepting drugs .

表 4.6 と比較すると, 評価 × の出力文は 2 文まで大きく減少している. したがって, 未知語処理における本手法の有用性が証明された. また, 本手法で正しい出力がされない文章は従来手法の出力で日本語を含まない文章に多いことがわかる. つまり, 未知語が存在しない文章の翻訳の際に本手法は精度が低くなる.

以上より, 事前に未知語出力用変換テーブルを作成するのではなく, 予め従来手法で翻訳を行った後, 翻訳できなかった部分のみを取り出して NMT によって翻訳を行うという手法ならば精度が向上すると考えられる.

第6章 おわりに

本研究では, TDSMT の未知語が原言語のまま出力されるという問題に着目し, 未知語処理に NMT を使用する手法を提案した. TDSMT は, 未知語処理のために, 未知語出力用変換テーブルを作成する. 未知語出力用変換テーブルは, 「 A が B ならば C は C 」と表現できる. これにより, カバー率を向上するが, 出力文に原言語の句が出現する.

そこで, 本研究では, 新たな未知語出力用変換テーブルを作成する手法を提案した. 提案手法では未知語 C を NMT で翻訳した結果を D とした. つまり, 提案手法で作成される未知語出力用変換テーブルは「 A が B ならば C は D (C を NMT で翻訳した結果)」と表現できる. 本研究の提案手法では, 入力文の全ての句が未知語となる可能性があると考えた. まず, 入力文から連続未知単語を作成し, それを NMT により翻訳した. そして, 翻訳結果を変換テーブルの D に格納し, 未知語出力用変換テーブルを作成した.

実験より, 入力文 100 文に対する人手評価では, \times が 14 文に対して, \times が 12 文であり, 全体の翻訳精度は向上しなかった. また, 従来手法で未知語(日本語)を含んでいた文章のみに着目して, 人手評価を行った. 人手評価の結果, \times が 12 文に対して, \times が 2 文であり, 未知語を含む文章は翻訳精度が向上していることが確認できた.

本研究の精度を向上させる手法として, 未知語のみに対して未知語処理を行う手法が考えられる. 具体的には, 事前に未知語処理を行うのではなく, 一度従来手法で翻訳を行った後に, 出力文に出現した日本語の部分を NMT で翻訳する手法が考えられる.

謝辞

本研究を進めるにあたり、研究の説明や論文の書き方など本研究を進めるにあたり、研究の説明や論文の書き方など様々のご指導を頂きました鳥取大学工学部電気情報系工学科自然言語処理研究室の村上仁一准教授に心から御礼申し上げます。また、本研究を進めるにあたり、御指導、御助言を頂きました、村田真樹教授に心から御礼申し上げます。また、自然言語処理研究室の皆様へ心から感謝の気持ちと御礼を申し上げたく、謝辞にかえさせていただきます。

参考文献

- [1] 安場裕人, 村上仁一, “変換主導型翻訳の提案”, 自然言語処理学会第 24 年次大会, March, 2018.
- [2] 西尾聡一郎, “パターンに基づく統計翻訳における文パターン確率の考察”, 平成 27 年度 卒業論文, pp.3-16, February 2016 .
- [3] Franz Josef Och, Hermann Ney, “A Systematic Comparison of Various Statistical Alignment Models”, *Computational Linguistics*, 29(1), pp.299-314, 1996.
- [4] KyungHyun Cho, “Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation”, 2014.
- [5] Ilya Sutskever, Oriol Vinyals, Quoc V. Le “Sequence to Sequence Learning with Neural Networks”, 2014
- [6] Minh-Thang Luong, Hieu Pham, and Christopher D. Manning. “Effective approaches to attention-based neural machine translation”, *In Proceedings of EMNLP*, pp.1412-1421, 2015.
- [7] 江木孝史, 村上仁一, 徳久雅人, “句に基づく対訳文パターンの自動作成と統計的手法を用いた英日パターン翻訳”, 自然言語処理学会第 20 回年次大会予稿集, pp.951-954, 2014.
- [8] カ久 剛士, “レーベンシュタイン距離を用いた翻訳精度の向上”, 平成 26 年度 卒業論文, pp.3-15, February 2015 .
- [9] 中村 勇太, 村上 仁一, “相対的意味論に基づく変換主導型パターンベース統計機械翻訳 (TDPBSMT) の提案”, 言語処理学会第 25 回年次大会 , P1-21 , 2019.
- [10] 今仁 優希, 村上 仁一, “ニューラル機械翻訳における低頻度語処理”, 言語処理学会 第 24 回年次大会, P4-18, 2018.

- [11] Vladimir Iosifovich Levenshtein, “Binary codes capable of correcting deletions, insertions, and reversals”, *Soviet Physics Doklady*, 10(8), pp.707-710, 1966.
- [12] 松本大輝, 村上仁一 . “翻訳における分野依存性を軽減する言語モデルの調査” 自然言語処理学会第 25 回年次大会, March 2019 . (予定)
- [13] Philipp Koehn, Marcello Federico, Brooke Cowan, Richard Zens, Chris Dyer, Ondřej 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, June 2007.
- [14] 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.
- [15] BLEU: “a Method for Automatic Evaluation of Machine Translation” , *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics(ACL)*, pp.311-318. 2002.
- [16] METEOR; Lavie Alon, and Denkowski Michael “An Automatic Metric for MT Evaluation with High Levels of Correlation with Human Judgements”, *Proceedings of the Second Workshop on Statistical Machine Translation*, pp.228-231. 2007.
- [17] Richard Schwartz, Linnea Micciulla, John Makhoul: “A Study of Translation Edit Rate with Targeted Human Annotation”, *AMTA*, 2006.
- [18] 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予稿集, pp.119-130, 2012.