

鳥取大学修士論文

相対的意味論に基づく統計機械翻訳  
における翻訳確率

令和3年2月

持続性社会創世科学研究科工学専攻  
M19J4033B 中村 勇太

# 概要

機械翻訳の手法として、パターン翻訳、機械翻訳等が研究されてきた。しかし、人手の翻訳には及ばない。この問題を解決するために村上は、新たな手法として、“相対的意味論に基づく統計機械翻訳 (RSMT : Relative Machine Translation)”[1] を提案した。この手法を本実験の従来手法とする。RSMT は出力文の選択に、翻訳確率を計算する。翻訳確率の計算方法には様々な方法が考えられる。そこで本研究では、RSMT で用いる翻訳確率の計算方法として最適な方法を調査する。実験で、RSMT を 5 つの異なる翻訳確率を用いて動作させ、結果を比較する。その結果、RSMT に最適な翻訳確率の計算方法は、自己相互情報量 (PMI : Pointwise Mutual Information)[11] と結論づけた。

# 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	GIZA++	3
2.1.4	IBM 翻訳モデル	3
2.1.5	単語に基づく統計翻訳の問題点	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	16
2.3	ニューラル機械翻訳 <sup>[9]</sup>	17
2.3.1	ニューラル機械翻訳の概要	17
2.3.2	Encoder-Decoder モデル	18
2.3.3	Attention モデル	20
2.4	相対的意味論 <sup>[1]</sup>	24
2.4.1	絶対的意味表現	24
2.4.2	相対的意味表現	25
2.4.3	2言語を用いた意味表現	26
2.4.4	2言語を用いた相対的意味表現	27
2.5	相対的意味論に基づく変換主導統計機械翻訳 (TDSMT) <sup>[10]</sup>	28
2.5.1	TDSMT の手順	29
2.5.2	学習の手順	29

2.5.3	翻訳の手順 . . . . .	31
<b>第3章</b>	<b>相対的意味論に基づく</b>	
	<b>統計機械翻訳 (RSMT)</b>	<b>33</b>
3.1	RSMT の概要 . . . . .	33
3.2	RSMT の手順 . . . . .	33
3.2.1	学習文対の日本語文と入力文の類似度を計算 . . . . .	33
3.2.2	学習文対の英語文と各候補文の類似度を計算 . . . . .	34
3.2.3	入力文と候補文に類似した学習文対を抜粋 . . . . .	34
3.2.4	DB 内の学習文対の翻訳確率を計算 . . . . .	35
3.2.5	入力文から各候補文への翻訳確率を計算 . . . . .	35
3.2.6	出力文の決定 . . . . .	35
<b>第4章</b>	<b>翻訳確率</b>	<b>36</b>
4.1	翻訳確率とは . . . . .	36
4.2	各翻訳確率の説明 . . . . .	36
4.2.1	自己相互情報量 (PMI : Pointwise Mutual Information) <sup>[11]</sup> . . . . .	36
4.2.2	Joint Probability(Joint) <sup>[12]</sup> . . . . .	37
4.2.3	Jaccard 係数 (Jaccard) <sup>[13]</sup> . . . . .	37
4.2.4	Dice 係数 (Dice) <sup>[14]</sup> . . . . .	37
4.2.5	Simpson 係数 (Simpson) <sup>[15]</sup> . . . . .	38
<b>第5章</b>	<b>Open-NMT の第1 候補を用いる実験</b>	<b>40</b>
5.1	実験目的と方法 . . . . .	40
5.2	実験条件 . . . . .	41
5.2.1	実験データ . . . . .	41
5.2.2	翻訳精度の調査の実験条件 . . . . .	42
5.3	実験結果 . . . . .	43
5.3.1	自動評価の結果 . . . . .	43
5.3.2	人手評価の結果 . . . . .	43
5.3.3	人手評価の実例 . . . . .	44

<b>第 6 章</b>	<b>Open-NMT の第 8 候補を用いる実験</b>	<b>47</b>
6.1	実験目的と方法 . . . . .	47
6.2	実験条件 . . . . .	47
6.2.1	実験データ . . . . .	47
6.3	実験結果 . . . . .	48
6.3.1	自動評価の結果 . . . . .	48
6.3.2	人手評価の結果 . . . . .	48
6.3.3	人手評価の実例 . . . . .	49
<b>第 7 章</b>	<b>実験のまとめ</b>	<b>52</b>
7.1	第 5 章のまとめ . . . . .	52
7.2	第 6 章のまとめ . . . . .	52
7.3	総評 . . . . .	52
<b>第 8 章</b>	<b>考察</b>	<b>53</b>
8.1	Open-NMT の 1best と RSMT(第 5 章) の性能比較 . . . . .	53
8.1.1	比較結果 . . . . .	54
8.1.2	人手評価の実例 . . . . .	55
8.2	Open-NMT の 1best と RSMT(第 6 章) の性能比較 . . . . .	57
8.2.1	比較結果 . . . . .	57
8.2.2	人手評価の実例 . . . . .	58
8.3	単語数の多さが原因である誤りの解決法について . . . . .	60
8.4	過剰生成を含む候補文の選択について . . . . .	61
8.5	類似度の計算方法 . . . . .	62
8.6	出力文の決定に用いる計算式 . . . . .	62
<b>第 9 章</b>	<b>おわりに</b>	<b>63</b>

# 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 . . . . .	16
2.3	Encoder-Decoder モデルの枠組み . . . . .	19
2.4	Attention モデルの枠組み . . . . .	22
2.5	TDSMT の流れ図 . . . . .	32
3.1	類似度の計算の実例 . . . . .	34
3.2	RSMT の流れ図 . . . . .	35
4.1	真部分集合の例 . . . . .	39

# 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 . . . . .	29
2.10	作成される対訳単語 . . . . .	29
2.11	単語レベル文パターンの作成例 . . . . .	30
2.12	変換テーブルの作成例 . . . . .	30
2.13	日本語側変換テーブルの適用例 . . . . .	31
2.14	英語変換テーブルの適用例 . . . . .	31
4.1	図 4.1 の状態での各種係数の数値 . . . . .	39
5.1	実験データ . . . . .	41
5.2	学習文対の例 . . . . .	41
5.3	入力文の例 . . . . .	41
5.4	自動評価の調査結果 (1000 文) . . . . .	43
5.5	人手評価結果 (100 文) . . . . .	43
5.6	人手評価の例 1 . . . . .	44
5.7	人手評価の例 2 . . . . .	44
5.8	人手評価の例 3 . . . . .	45
5.9	人手評価の例 4 . . . . .	45
5.10	人手評価の例 5 . . . . .	46

5.11	人手評価の例 6 . . . . .	46
6.1	自動評価の調査結果 (1000 文) . . . . .	48
6.2	人手評価結果 (100 文) . . . . .	48
6.3	人手評価の例 1 . . . . .	49
6.4	人手評価の例 2 . . . . .	49
6.5	人手評価の例 3 . . . . .	50
6.6	人手評価の例 4 . . . . .	50
6.7	人手評価の例 5 . . . . .	51
6.8	人手評価の例 6 . . . . .	51
8.1	自動評価結果 (1000 文) . . . . .	54
8.2	人手評価結果 (100 文) . . . . .	54
8.3	人手評価の例 1 . . . . .	55
8.4	人手評価の例 2 . . . . .	55
8.5	人手評価の例 3 . . . . .	56
8.6	人手評価の例 4 . . . . .	56
8.7	自動評価結果 (1000 文) . . . . .	57
8.8	人手評価結果 (100 文) . . . . .	57
8.9	人手評価の例 1 . . . . .	58
8.10	人手評価の例 2 . . . . .	58
8.11	人手評価の例 3 . . . . .	59
8.12	人手評価の例 4 . . . . .	59
8.13	過剰生成を含む出力の例 . . . . .	61
8.14	過剰生成が理由で人手評価 × とした数 . . . . .	61

# 第1章 はじめに

機械翻訳において様々な手法が提案されている。しかし、それらの手法で得られる出力が異なる場合が多い。

そこで、異なる機械翻訳手法によって得られた複数の出力(候補文)から、最尤の出力を選択する手法である“相対的意味論に基づく統計機械翻訳(RSMT)”が提案されている。この手法を用いることで、各手法の出力から最尤の出力を得られる。

RSMTでは出力を選択するために、学習文対の日本語側(A)、学習文対の英語側(B)、入力文(C)、候補文(D)について文類似度及び文翻訳確率を計算する。文類似度は同一言語の文章(AとCまたは、BとD)がどれだけ類似しているかを示す。文翻訳確率はCがDに翻訳される確率を示す。さらに、文類似度及び文翻訳確率には様々な計算方法が存在する。

本稿の目的は、RSMTで用いる文翻訳確率の計算方法として最適な方法を調査することである。5つの文翻訳確率の計算方法を用いてRSMTを動作させ結果を比較する。この実験によって、最適な文翻訳確率の計算方法を調査する。

実験として、1000文を入力とした自動評価による翻訳精度の調査を行った。また、100文を入力とした人手評価を行った。実験の結果、RSMTに最適な翻訳確率の計算方法は、自己相互情報量(PMI: Pointwise Mutual Information)[11]と結論づけた。

本論文の構成は以下の通りである。第2章で従来の研究について説明する。また、第3章でRSMTについて説明し、第4章で用いる文翻訳確率について説明する。第5章でOpen-NMTの第1候補を候補文に用いた実験の、実験データ、実験結果と評価を示す。さらに、第6章でOpen-NMTの第8候補を候補文に用いた実験の、実験データ、実験結果と評価を示す。そして、実験のまとめを第7章に示す。第8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 [3] 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 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。

### 2.1.4 IBM 翻訳モデル

IBM 翻訳モデルを以下に示す。これは、カ久ら [5] 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  $\epsilon$  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  $\lambda_e$  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  $\lambda_e$  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 .

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

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

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  $\phi$  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  $\phi_0$  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.

$$P(F|E) = \sum_{a_1=0}^l \dots \sum_{a_m=0}^l P(F, a|E) \quad (2.17)$$

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

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

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

### 2.1.5 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， 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．

## 2.2 句に基づく統計翻訳

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。

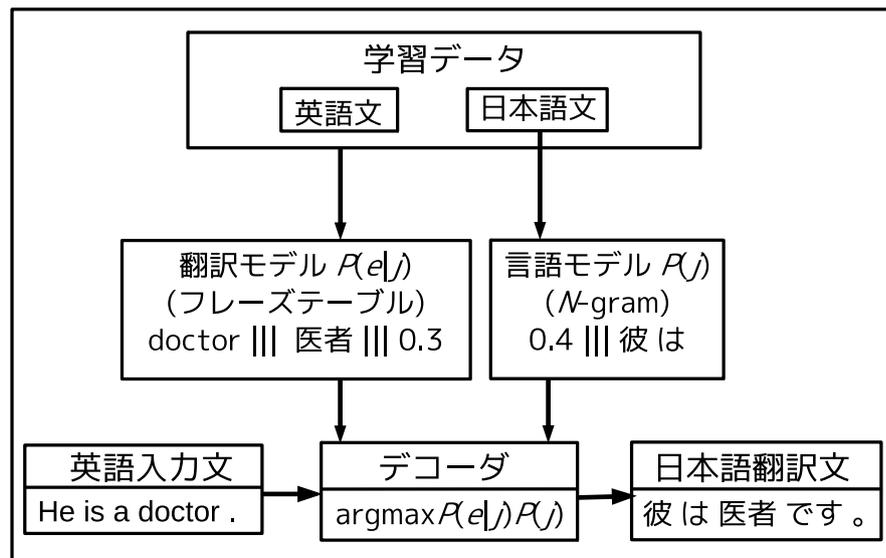


図 2.1: 日英統計翻訳の枠組み

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

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

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 Log Linear(2.25)

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

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．

$$\begin{aligned} & P(\text{ぶんこが揺れている。} \quad \textit{The swing is swinging.}) \\ = \log_2 & \left\{ \frac{\text{count}(\text{ぶんこが} \quad \textit{The swing})}{\text{count}(\text{ぶんこが})} \times \frac{\text{count}(\textit{The swing} \quad \text{ぶんこが})}{\text{count}(\textit{The swing})} \right\} \\ & = \log_2 \left\{ \frac{\frac{1}{100,000}}{\frac{2}{100,000}} \times \frac{\frac{1}{100,000}}{\frac{1}{100,000}} \right\} \end{aligned} \quad (2.26)$$

## 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” [7] 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。

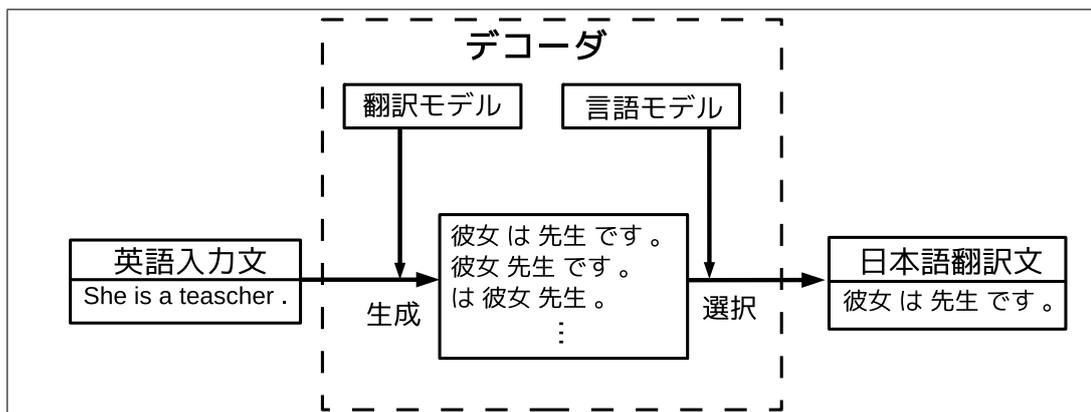


図 2.2: デコーダの動作例

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 ニューラル機械翻訳<sup>[9]</sup>

### 2.3.1 ニューラル機械翻訳の概要

ニューラル機械翻訳 (Neural Machine Translation; 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、出力系列と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。

## 2.3.2 Encoder-Decoder モデル

### Encoder-decoder モデルの概要

Encoder-Decoder モデルは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

図2.3にEncoder-Decoderモデルの枠組みを示す。Encoder-Decoderモデルでは、Encoderは入力文をベクトル系列  $x = (x_1, \dots, x_{T_x})$  として読み込み、ベクトル  $c$  に符号化する。一般的なEncoder-Decoderモデルは、2.27式および2.28式で表されるRNNを用いて構成される。

$$h_t = f(x_t, h_{t-1}) \quad (2.27)$$

$$c = q(\{h_1, \dots, h_{T_x}\}) \quad (2.28)$$

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、基本的なEncoder-Decoderモデルにおいて、活性化関数  $f$  にはLSTMが用いられ、 $q(\{h_1, \dots, h_{T_x}\}) = h_{T_x}$  として定義されている。Decoderは文脈ベクトル  $c$  と既に生成された単語列  $\{y_1, \dots, y_{t-1}\}$  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2.29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。

$$p(y) = \prod_{t=1}^{T_y} p(y_t | \{y_1, \dots, y_{t-1}\}, c) \quad (2.29)$$

2.27式および2.28式で表されるRNNを用いて、それぞれの条件付き確率は2.30式によりモデル化される。ここで、 $g$  は  $y_t$  の確率を生成する非線形の多層関数であり、 $s_t$  はRNNの隠れ層の状態となる。

$$p(y_t | \{y_1, \dots, y_{t-1}\}, c) = g(y_{t-1}, s_t, c) \quad (2.30)$$

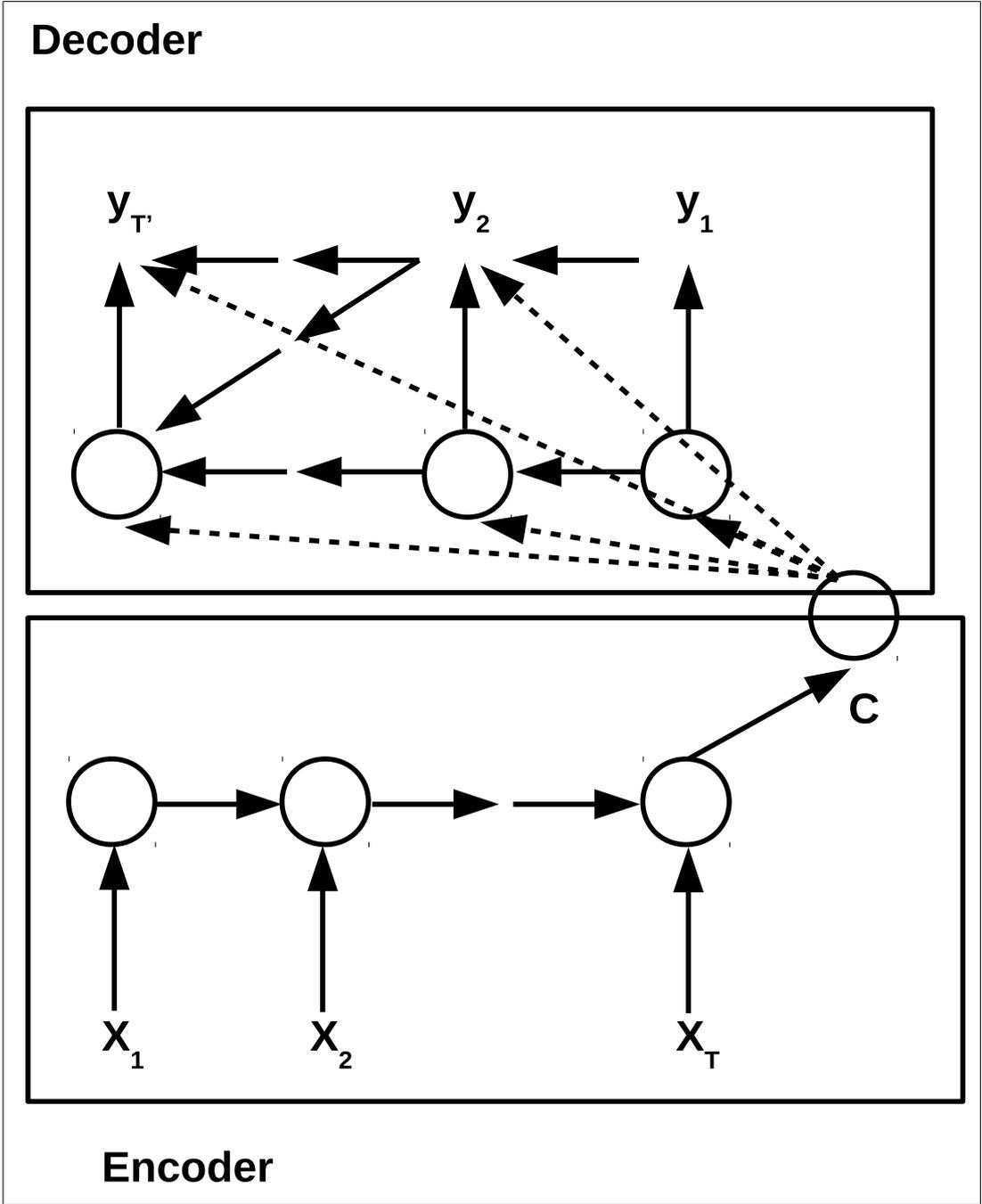


図 2.3: Encoder-Decoder モデルの枠組み

### 2.3.3 Attention モデル

#### Attention モデルの概要

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 (bidirectional RNN) 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、その単語が最も相関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

図 2.4 に Attention モデルの枠組みを示す。Encoder-Decoder モデルにおいて 2.29 式で定義されていた条件付き確率に対して、Attention モデルでは時刻  $i$  の時の条件付き確率を 2.31 式により定義する。Encoder-Decoder モデルにおける条件付き確率  $p$  は、既に出力された単語列と全ての入力系列  $\mathbf{x}$  を変換した文脈ベクトル  $\mathbf{c}$  により定義されていたが、Attention モデルにおいては入力系列を一つの文脈ベクトルのみに圧縮することなく、全ての入力系列  $\mathbf{x}$  より導かれる。

$$p(y_i | y_1, \dots, y_{i-1}, \mathbf{x}) = g(y_{i-1}, s_i, c_i) \quad (2.31)$$

また、 $s_i$  は時刻  $i$  での隠れ層の状態であり、2.32 式により計算される。

$$s_i = f(s_{i-1}, y_{i-1}, c_i) \quad (2.32)$$

ここで、2.29 式の Encoder-Decoder モデルと異なり、各目的言語単語  $y_i$  は文脈ベクトル  $c_i$  により状態が与えられる。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。入力文の前後両方向より得られる各アノテーション  $h_i$  は全入力文の情報を保持しており、特に入力文中の  $i$  番目の単語付近の情報を強く対応付けている。また、2.33 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。

$$c_i = \sum_{j=1}^{T_x} \alpha_{ij} h_j \quad (2.33)$$

各アノテーション  $h_j$  の重み  $\alpha_{ij}$  は

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{T_x} \exp(e_{ik})} \quad (2.34)$$

により計算される。ここで

$$e_{ij} = a(s_{i-1}, h_j) \quad (2.35)$$

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  $s_{i-1}$  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。

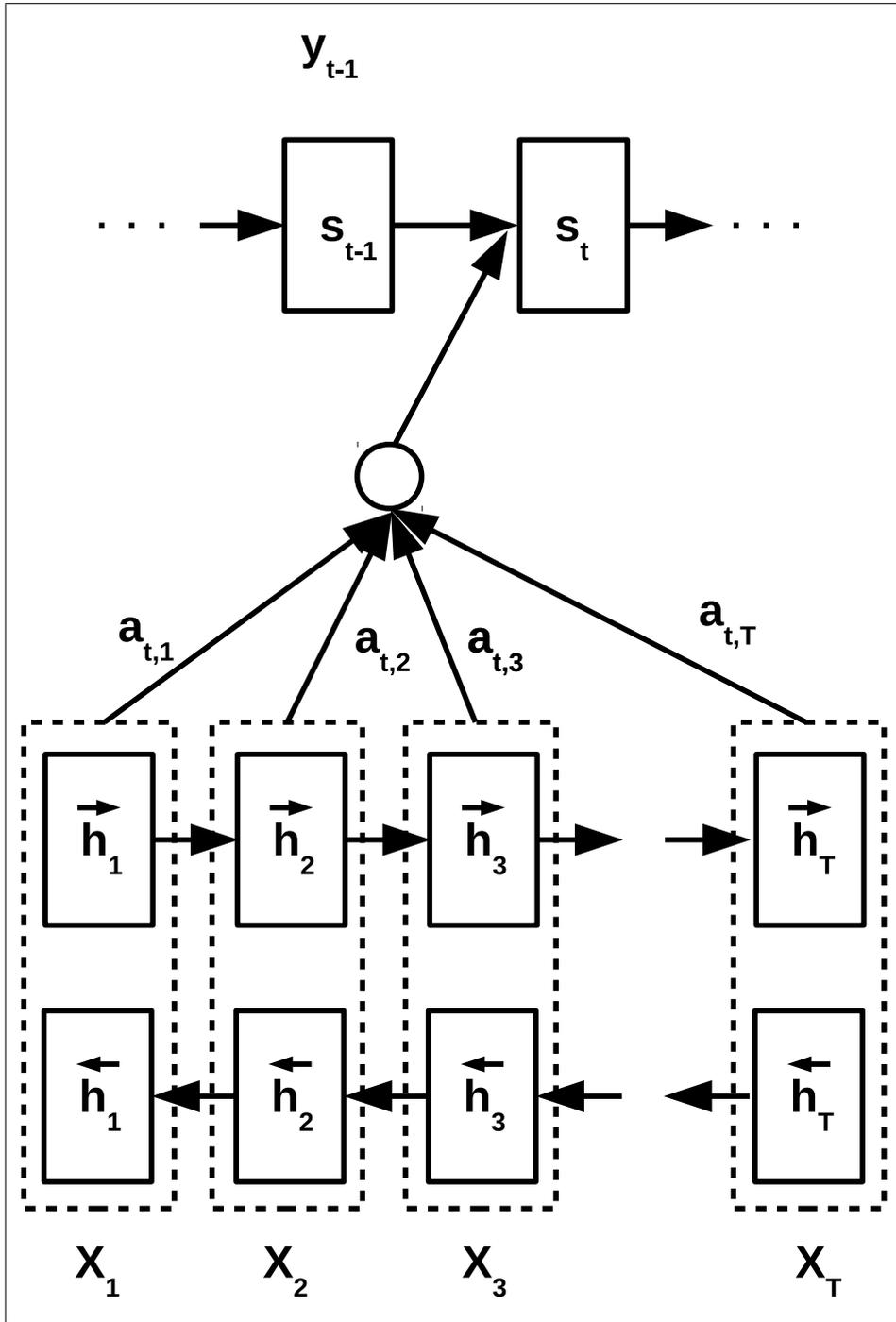


図 2.4: Attention モデルの枠組み

## NMTの学習における乱数の影響

今回の実験では、NMTの一種であるOpen-NMTを用いる。Open-NMTは、学習で乱数を用いている。よって同一のコーパスで学習を行っても、翻訳結果が異なるモデルが生成される。本論ではOpen-NMTのモデルを複数生成し、同一の入力文を翻訳することで、RSMTの候補文を獲得する。

## Open-NMTの*N*-best出力

Open-NMTは第*n*候補まで、複数の出力文を出力できる。また、一般的に第1候補が最も優れている。しかし、第2候補以降に優れた出力が存在することも考えられる。本論ではRSMTを用いることで、第1候補より優れた出力文が選択可能か調査する。

## 2.4 相対的意味論<sup>[1]</sup>

本節では、語彙の意味を説明するために、相対的意味表現と絶対的意味表現について述べる。次に、これらの方法でも説明できない語彙に対し、2言語を利用した方法(相対的意味論)を示す。

### 2.4.1 絶対的意味表現

絶対的意味表現とは、単語の語彙の意味を具体的な他の語彙で表現する。この表現方法は、基本的には、理解しやすい。以下は、絶対的意味表現の例である。

青 - 波長 400nm 近辺の光

赤 - 波長 800nm 近辺の光

多くの語彙は、絶対的意味表現で、表現できる。しかし、意味の表現が困難な語彙も多い。以下は、その例である。

右 - 南を向いたとき、西にあたる方(広辞苑)

- アナログ時計の1時から5時までの表示がある側(新明解国語辞典)

- この辞書を開いて読むとき、偶数のページのある側(岩波国語辞典)

- 人体を対称線に沿って二分したとき、心臓のない方。(大修館書店 明鏡国語辞典)

- 大部分の人が食事の時、箸を持つ側(大辞泉)

左 - 南に向かったとき、東にあたる方(旺文社国語辞典)

- アナログ時計の文字盤に向かった時に、七時から十一時までの表示のある側。(新明解国語辞典)

- この辞典を開いて読む時、奇数ページのある側(岩波国語辞典)

- 大部分の人にとって、はしを持つ手と反対のほう(新選国語辞典)

- 人の背骨の中心線と鼻の先端とを含む平面で空間を二つの部分に分けた時に、大部分の人の場合、心臓の搏動を感じる場所がある方の部分(新明解国語辞典)

いずれの説明も、曖昧な点が残る(例えば、“南”とは?)似た語彙として“上”“下”などがある。

## 2.4.2 相対的意味表現

相対的意味表現とは、単語の意味を、お互いに相反する単語で表現する方法である。例を以下に述べる。

右 - 左の反対

左 - 右の反対

絶対的意味表現で困難な語彙は、相対的意味表現で表現可能な場合が多い。しかし、この表現方法は、説明がループするため、“煙を蒔いたような表現”と言って嫌悪する人も多い。

また、語彙の中には、対的意味表現を用いても、説明できない語彙がある。以下は、その例である。

人間

- ひと 人類
- 遺伝的な両親が人間である人
- 考える葦

いずれの説明も、疑問点が残る。このように、“人間”の意味を説明することは、非常に困難である。

### 2.4.3 2言語を用いた意味表現

単言語では説明するのが困難な語彙は、2言語を利用して表現することが可能な場合がある。例を以下に挙げる。

- “人間” とは “human” である。

以下に、2言語を用いた例を示す。

“右” とは “right” である。

“左” とは “left” である。

“犬” とは “dog” である。

“猫” とは “cat” である。

“動物” とは “animal” である。

しかし、“犬” には別の意味がある。

“犬” とは “spy” である。

このように、2言語を用いても、語彙に複数の意味があるとき、一意に表現できない。

#### 2.4.4 2言語を用いた相対的意味表現

単言語において、複数の意味があるとき、1言語を用いた意味表現では、一意に表現できない。このような場合、2言語の相対的意味を用いて表現できる。

“aがbならばcはd”

これを“2言語を用いた相対的意味表現”と呼んでいる。

例を以下に示す。

1. “右”が“right”ならば“左”は“left”
2. “犬”が“dog”ならば “猫”は“cat”
3. “犬”が“Canis”ならば“猫”は“Felis”
4. “犬”が“spy”ならば “猫”は“sex decoy”
5. “人間”が“human”ならば “動物”は“animal”
6. “人間”が“human”ならば “猫”は“cat”

## 2.5 相対的意味論に基づく変換主導統計機械翻訳(TDSMT)<sup>[10]</sup>

“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。

## 2.5.1 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．

## 2.5.2 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[8]) 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

学習文対と単語レベル文パターンを照合する．変数化した対訳単語と，変数に当たる対訳句を変換テーブルとする．表 2.12 では変数  $N2$  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 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．

### 2.5.3 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 (2.36) 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．

$$\log P = \log P_v + \log P_m \quad (2.36)$$

図 2.5 に TDSMT の流れ図を示す.

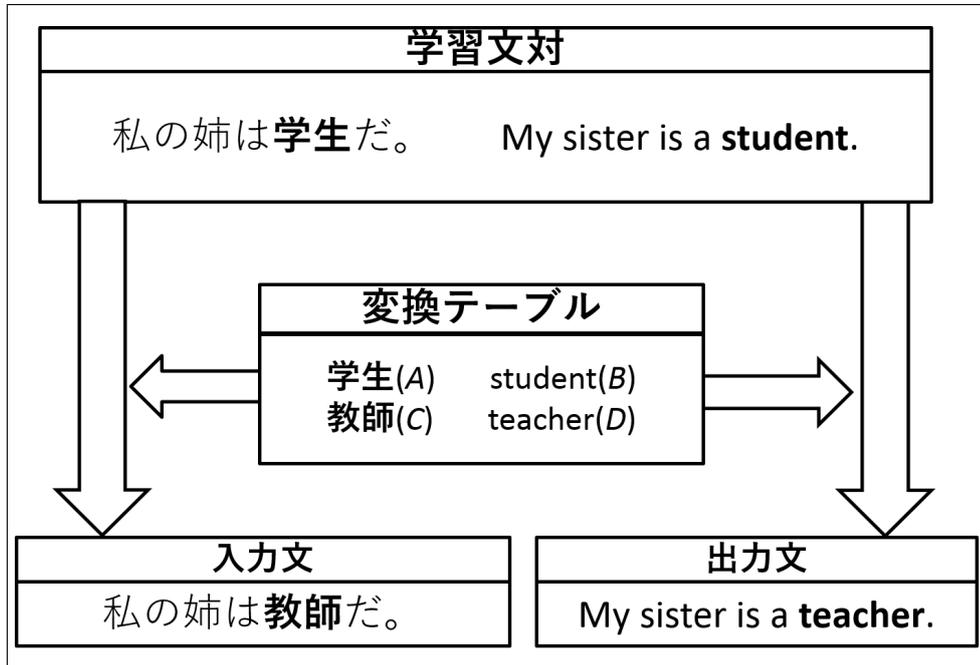


図 2.5: TDSMT の流れ図

# 第3章 相対的意味論に基づく 統計機械翻訳(RSMT)

## 3.1 RSMTの概要

“相対的意味論に基づく統計機械翻訳”では異なる機械翻訳手法によって得られた複数の出力(候補文)から、最尤の出力を選択する。そのために、学習文対の日本語側(A)、学習文対の英語側(B)、入力文(C)、候補文(D)について文類似度及び文翻訳確率を計算する。

## 3.2 RSMTの手順

本節で、RSMTの動作手順を示す。

### 3.2.1 学習文対の日本語文と入力文の類似度を計算

学習文対の日本語文と入力文の類似度を計算する。計算された値を  $sim_{AC}$  とする。

$sim_{AC}$  の計算式を式(3.1)に示す。本論では、同一言語の文章間における類似度を、単語の一致率であると定義する。式(3.1)は、学習文対の日本語文から見た入力文との単語の一致率と、入力文から見た学習文対の日本語文との単語の一致率を乗算した式である。

$$sim_{AC} = \frac{count(Match)}{count(A_{all})} \times \frac{count(Match)}{count(C_{all})} \quad (3.1)$$

$count(Match)$  : 学習文対の日本語文と入力文において共通している単語の総数

$count(A_{all})$  : 学習文対の日本語文の単語の総数

$count(C_{all})$  : 入力文の単語の総数

### 3.2.2 学習文対の英語文と各候補文の類似度を計算

学習文対の英語文と各候補文の類似度を計算する．計算された値を  $sim_{BD}$  とする． $sim_{BD}$  の計算式を式 (3.2) に示す．式 (3.2) の立式の考え方は式 (3.1) と同様である．

$$sim_{BD} = \frac{count(Match)}{count(B_{all})} \times \frac{count(Match)}{count(D_{all})} \quad (3.2)$$

$count(Match)$  : 学習文対の英語文と候補文において共通している単語の総数

$count(B_{all})$  : 学習文対の英語文の単語の総数

$count(D_{all})$  : 候補文の単語の総数

類似度の計算の実例を図 3.1 に示す．

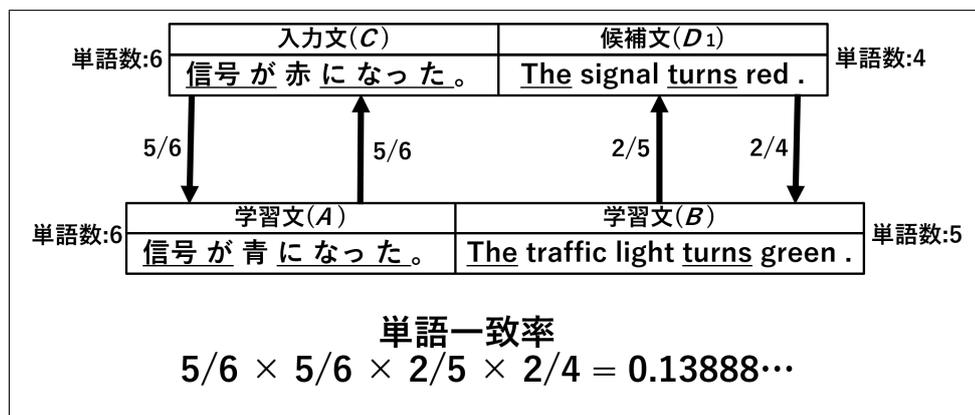


図 3.1: 類似度の計算の実例

### 3.2.3 入力文と候補文に類似した学習文対を抜粋

入力文と各候補文に類似した学習文対を一定数抜粋する．抜粋された学習文対の集まりを  $DB$  とする．

各候補文について式 (3.3) の数値が高い順に抜粋する．式 (3.3) によって，日本語文間の類似度と英語文間の類似度を乗算することで， $AB$  と  $CD$  間の類似度を計算する．

$$sim_{AC} \times sim_{BD} \quad (3.3)$$

### 3.2.4 DB内の学習文対の翻訳確率を計算

DB内の学習文対の翻訳確率を計算する．計算された値を  $trans_{AB}$  とする．  
ここで用いる式は第4章で説明する．

### 3.2.5 入力文から各候補文への翻訳確率を計算

入力文から各候補文への翻訳確率を計算する．計算された値を  $trans_{CD}$  とする．  
ここで用いる式は前節と同様である．

### 3.2.6 出力文の決定

計算された値を用いて出力文を決定する．

式(3.4)の値が最も高い候補文を，出力文とする．ここまでの手順で求めた類似度及び翻訳確率を全て合算し，最終的な候補文選択に用いるスコアを算出する．

$$sim_{AC} + sim_{BD} + trans_{AB} + trans_{CD} \quad (3.4)$$

RSMTの流れ図を図3.2に示す．

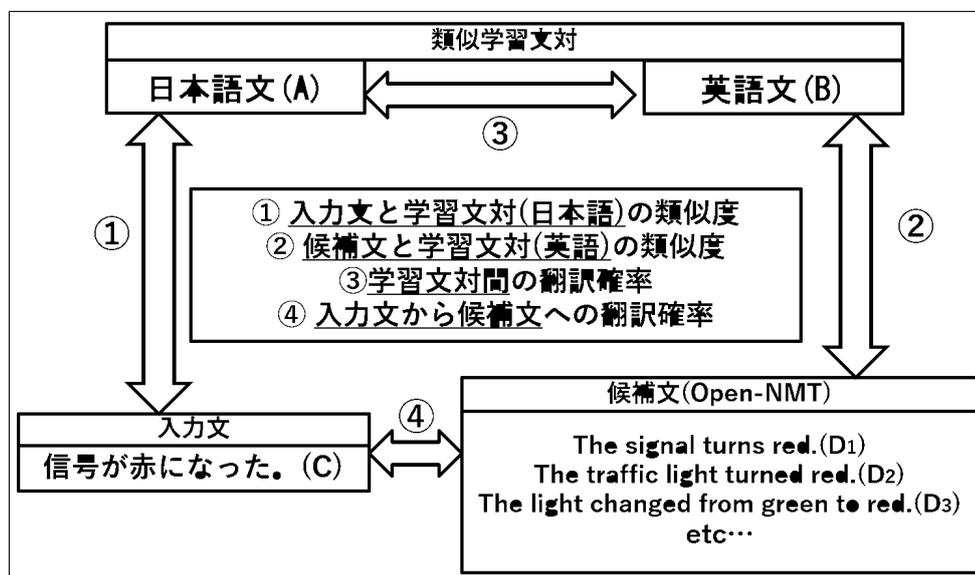


図 3.2: RSMT の流れ図

## 第4章 翻訳確率

### 4.1 翻訳確率とは

翻訳確率とは、RSMT の選択で用いる「入力文 ( $C$ ) が候補文 ( $D$ ) に翻訳される確率」を表す数値である。翻訳確率の計算方法には様々な計算方法が考えられる。本論では、RSMT で用いるにあたって最も優れた翻訳確率の計算方法を決定することを目的としている。

### 4.2 各翻訳確率の説明

本論の実験で用いる翻訳確率の計算方法について説明する。なお実際の計算では、2つの単語の連なりを1つの単語と見なす。

#### 4.2.1 自己相互情報量 (PMI : Pointwise Mutual Information)<sup>[11]</sup>

自己相互情報量とは、2つの事象の間の関連度合いを測る尺度である。日英単語の共起をこの尺度に当てはめることで翻訳確率として用いる。計算式を式 4.1 に示す。

$$\sum_{j=1}^M \log_2 \sum_{i=1}^N \frac{\text{count}(J_j, E_i)}{\text{count}(J_j)\text{count}(E_i)} \quad (4.1)$$

$\text{count}()$  : 学習文対中での登場回数

$J_j$  : 入力文中  $j$  番目の単語  $E_i$  : 候補文中  $i$  番目の単語

$M$  : 入力文の単語数  $N$  : 候補文の単語数

## 4.2.2 Joint Probability(Joint)<sup>[12]</sup>

Joint Probability とは、交差エントロピーを拡張した翻訳確率の計算方法である。計算式を式 4.2 に示す。

$$\sum_{j=1}^M \sum_{i=1}^N \frac{\text{count}(J_j, E_i)}{\text{count}(J_{all})\text{count}(E_{all})} \times \log_2 \frac{\text{count}(J_j, E_i)}{\text{count}(J_j)\text{count}(E_i)} \quad (4.2)$$

$\text{count}(J_{all})$  : 全学習文対における日本語単語の総数

$\text{count}(E_{all})$  : 全学習文対における英語単語の総数

## 4.2.3 Jaccard 係数 (Jaccard)<sup>[13]</sup>

Jaccard 係数とは 2 つの集合がどれだけ類似しているかを表す係数である。入力文と候補文を「文を構成する単語の集合」とみなすことで、Jaccard 係数を翻訳確率として用いる。計算式を式 4.3 に示す。

$$\sum_{j=1}^M \log_2 \sum_{i=1}^N \frac{\text{count}(J_j, E_i)}{\text{count}(J_j) + \text{count}(E_i) - \text{count}(J_j, E_i)} \quad (4.3)$$

## 4.2.4 Dice 係数 (Dice)<sup>[14]</sup>

Dice 係数とは Jaccard 係数と同じく、2 つの集合がどれだけ類似しているかを表す係数である。計算式を式 4.4 に示す。

$$\sum_{j=1}^M \log_2 \sum_{i=1}^N \frac{2\text{count}(J_j, E_i)}{\text{count}(J_j) + \text{count}(E_i)} \quad (4.4)$$

#### 4.2.5 Simpson 係数 (Simpson)<sup>[15]</sup>

Simpson 係数とは Jaccard 係数と同じく、2つの集合がどれだけ類似しているかを表す係数である。計算式を式 4.5 に示す。

$$\sum_{j=1}^M \log_2 \sum_{i=1}^N \frac{\text{count}(J_j, E_i)}{\min(\text{count}(J_j), \text{count}(E_i))} \quad (4.5)$$

## Simpson 係数の特徴について

Jaccard 係数, Dice 係数, Simpson 係数の中で最も大きな違いを持つのは Simpson 係数である。Simpson 係数は、「片方の集合がもう片方の集合の真部分集合である時, 類似度が最大になる」という特徴がある。この特徴について図 4.1 を例に考える。かっこの中は集合の要素数である。

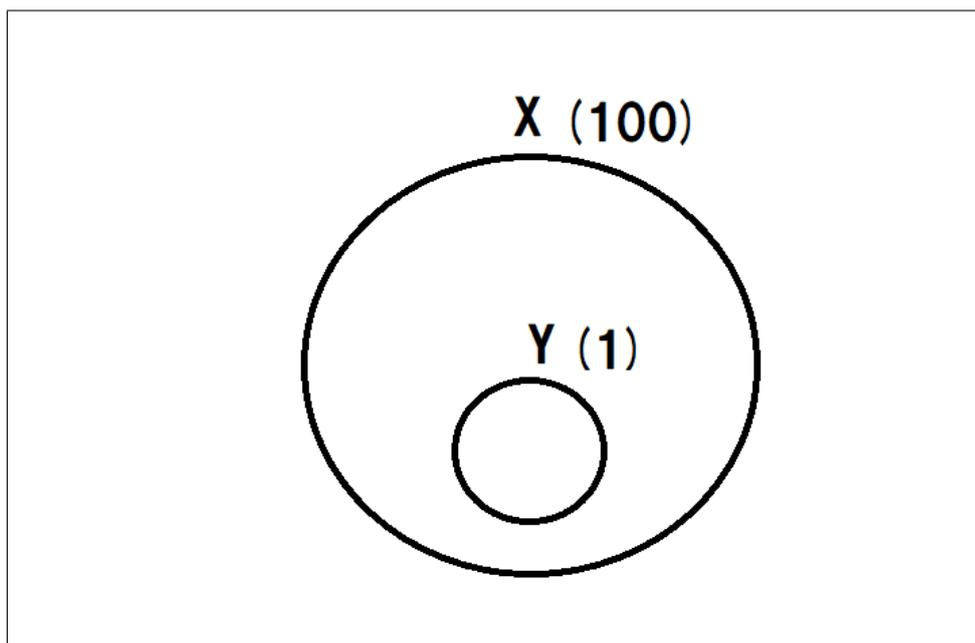


図 4.1: 真部分集合の例

図 4.1 の状態で各種係数を用いて類似度を求めると表 4.1 に示す数値となる。図 4.1 が示すように, Jaccard と Dice にほぼ差は出ないが, Simpson は類似度 1.0 となる。

表 4.1: 図 4.1 の状態での各種係数の数値

Jaccard	$1 \div 100 = 0.01$
Dice	$2 \div 101 \approx 0.02$
Simpson	$1 \div 1 = 1.0$

# 第5章 Open-NMTの第1候補を用いる 実験

## 5.1 実験目的と方法

本章では，Open-NMTの第1候補を，RSMTの候補文として実験を行う．5つの翻訳確率計算方法でRSMTを動作させ，翻訳精度を調査する．翻訳精度の調査は自動評価及び人手評価で行う．自動評価にはBLEU[16]，METEOR[17]，TER[18]を用いる．

## 5.2 実験条件

### 5.2.1 実験データ

本研究では，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 [19] .

RSMT の候補文は Open-NMT の第 1 候補を用いる．また，使用する Open-NMT のモデル数は 32 個である．よって候補文は，入力文 1 文につき 32 文存在する．使用するデータの内訳を表 5.1 に示す．

表 5.1: 実験データ

学習文対	160,000 文対
入力文 (自動評価用)	1,000 文
候補文 (自動評価用)	32,000 文
入力文 (人手評価用)	100 文
候補文 (人手評価用)	3,200 文

学習文対および，入力文の例を表 5.2 と表 5.3 に示す．

表 5.2: 学習文対の例

学習文対	
日本語原文	英語原文
あいつは甘えている。	He is spoiled.
あの小川で釣りをしよう。	Let's fish the creek.
いつかは石油資源が枯渇する。	Oil resources will dry up someday.

表 5.3: 入力文の例

入力文	
日本語文	参照文
花は太陽の方に傾く。	Flowers bend toward the sun.
彼らの間に不満が増大した。	Discontent waxed among them.
彼はすぐ帰国の途につく。	He will soon start for home.

### 5.2.2 翻訳精度の調査の実験条件

RSMT 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、人手評価は絶対評価で行う。人手評価の基準は次の3つである。

- ○ : 入力文の意味と合致かつ文法の誤りがない
- △ : 入力文の意味と部分的に合致
- × : 入力文の意味と合致していない

類似度の計算方法は図 3.1 の方法を用いる。また、類似度の計算は2単語を1単語として用いる。

## 5.3 実験結果

### 5.3.1 自動評価の結果

自動評価の結果を表 5.4 に示す。

表 5.4: 自動評価の調査結果 (1000 文)

	BLEU	METEOR	TER
PMI	<b>0.183</b>	<b>0.484</b>	<b>0.583</b>
Joint	0.164	0.468	0.652
Jaccard	0.173	0.474	0.605
Dice	0.176	0.477	0.606
Simpson	0.174	0.471	0.615

表 5.4 によると，全ての評価指標において PMI が最も優れた結果を示す。

### 5.3.2 人手評価の結果

100 文を対象に，人手評価を行った結果を表 5.6 に示す。

表 5.5: 人手評価結果 (100 文)

			×
PMI	33	21	46
Joint	26	21	53
Jaccard	33	21	46
Dice	33	22	45
Simpson	30	23	47

表 5.5 によると，Joint が他の手法と比較して性能が劣る。また，Joint 以外の手法間に大きな差はない。

### 5.3.3 人手評価の実例

本節では，人手評価の実例をいくつか掲載する．

表 5.6: 人手評価の例 1

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 .	評価結果
PMI Jaccard Dice	He spent his late years in Venice .	
Joint Simpson	He spent his late years to Venice .	

表 5.6 の例では，“Venice” の前の前置詞の違いが評価結果に影響している．PMI , Jaccard , Dice の出力文の前置詞“in”の方が，Joint , Simpson の前置詞“to”より適切であると考えた．

表 5.7: 人手評価の例 2

入力文	私はその仕事に慣れていない。	
参照文	I am not used to the task .	評価結果
Jaccard Dice	I am not used to the work .	
Joint	I am not yet used to the work .	
PMI Simpson	I am unused to the work .	×

表 5.7 の例では，“used” 付近の違いが評価結果に影響している．Jaccard , Dice の出力文は参照文と一致している．

次に，Joint の出力文は日本語で「まだ」という意味を付与する“yet”が付随している．これによって僅かに入力文の意味から離れている．

最後に，PMI , Simpson の出力文は“unused”という動詞が“used”の代わりに用いられている．“used to”で慣用句的な扱いであるため，“unused”を用いるのは不適切である．

表 5.8: 人手評価の例 3

入力文	あの女の人は美しい顔をしている。	
参照文	That woman has a beautiful face .	評価結果
Jaccard Dice	That woman has a beautiful face .	
PMI Joint Simpson	That woman's face has a beautiful face .	

表 5.8 の例では、主語の違いが評価結果に影響している。Jaccard、Dice の出力文は参照文と一致している。

PMI、Joint、Simpson の出力文は主語が “That's womans face” となっている。これによって「女性の顔が顔を持っている」という訳となり、ニュアンスは伝わるが文章としては不適切である。

表 5.9: 人手評価の例 4

入力文	公園は川まで広がっている。	
参照文	The park reaches to the river .	評価結果
Simpson	The park reaches to the river .	
PMI Jaccard Dice	The park reaches the river .	
Joint	The park is spreading to the river .	×

表 5.9 の例では、動詞周辺の違いが評価結果に影響している。Simpson の出力文は参照文と一致している。

PMI、Jaccard、Dice の出力文は前置詞 “to” が不足している。入力文の意味は伝わるが、文法的に不敵である。

Joint の出力文は動詞 “spreading” が明らかに入力文の意味と不一致である。

表 5.10: 人手評価の例 5

入力文	賛成が 50 票、反対が 15 票あった。	
参照文	There were fifty votes for and fifteen votes against .	評価結果
PMI	There were fifty votes in favor .	×
Joint Jaccard Dice Simpson	The park reaches the river .	×

表 5.10 の例では，出力文に不足している情報があるため，全て評価 × となっている．  
 全ての出力文において，「反対が 15 票あった」の情報が不足している．  
 この入力文は第 6 章の評価例にも掲載している．

表 5.11: 人手評価の例 6

入力文	彼は彼女に慰めのことばをかけた。	
参照文	He spoke comforting words to her .	評価結果
全ての計算法	He spoke a few words of comfort to her .	

表 5.11 の例では，全て評価 ○ となっている．この入力文は第 6 章の評価例にも掲載している．

# 第6章 Open-NMTの第8候補を用いる 実験

## 6.1 実験目的と方法

ここでは、Open-NMTの第8候補までをRSMTの候補文として用いて実験を行う。候補文となる基準を引き下げることで、Open-NMTのモデルを複数生成するよりも容易に、多くの候補文を得ることができる。ここでは、第5章の実験との比較のために、候補文の数を揃える。

## 6.2 実験条件

### 6.2.1 実験データ

実験に使用するデータはRSMTの候補文以外、表5.1と同様である。使用するOpen-NMTのモデル数は4つである。よって候補文は、入力文1文につき32文存在する。類似度の計算方法は図3.1の方法を用いる。また、類似度の計算は2単語を1単語として用いる。

## 6.3 実験結果

### 6.3.1 自動評価の結果

自動評価の結果を表 6.1 に示す。

表 6.1: 自動評価の調査結果 (1000 文)

	BLEU	METEOR	TER
PMI	0.189	0.465	<b>0.613</b>
Joint	0.173	0.451	0.702
Jaccard	0.193	0.465	0.630
Dice	<b>0.194</b>	<b>0.466</b>	0.635
Simpson	0.189	0.465	0.655

表 6.1 によると，PMI と Dice が他の手法と比較すると優れた結果を示す。

### 6.3.2 人手評価の結果

100 文を対象に，人手評価を行った結果を表 6.2 に示す。

表 6.2: 人手評価結果 (100 文)

			×
PMI	36	18	46
Joint	18	17	65
Jaccard	32	21	47
Dice	29	20	51
Simpson	26	21	53

表 5.6 と表 6.2 を比較すると，PMI 以外の手法は が減少している。特に，Joint の減少数は顕著である。一方で，PMI は が増加している。このことから，PMI を用いることで，第 1 候補以外の優れた出力文を選択できることが分かる。

### 6.3.3 人手評価の実例

本節では，人手評価の実例をいくつか掲載する．

表 6.3: 人手評価の例 1

入力文	彼の話には深みがある。	
参照文	There is something deepin what he said .	評価結果
PMI	His story has a depth of meaning .	
Jaccard	There is a depth in his story .	
Joint Dice Simpson	His story has a depth of depth .	×

表 6.3 の例では，目的語の違いが評価結果に影響している．PMI の出力文を訳すると「彼の話は意味の深さを持っている。」となり入力文の意味に適している．

次に，Jaccard の出力文を訳すると「彼の話には深さがある。」となる．評価に悩んだが，“depth” だけでは「話の深さ」という意味を十全に表現していない，と判断した．

Joint , Jaccard , Simpson の出力文は，文中に同一単語 (depth) が過剰に繰り返されている．この現象を過剰生成と言う．本論の人手評価では過剰生成が見られた出力文は全て評価 × としている．

表 6.4: 人手評価の例 2

入力文	公園は川まで広がっている。	
参照文	The park reaches to the river .	評価結果
PMI Simpson	The park reaches as far as the rivers .	
Joint Jaccard Dice	The park extends as far as the rivers .	

表 6.4 の例では，動詞の違いが評価結果に影響している．PMI の出力文を訳すると「公園は川と同じくらい遠く届いている。」となる．評価に悩んだが，入力文の「広がっている」は「届いている」と同義であると判断した．よって，出力文は入力文の意味を表現できていると判断した．

次に，Joint , Jaccard , Dice の出力文は「伸ばす」という意味の“extends” が用いられている．よって，入力文の意味を十全に表現していない．

表 6.5: 人手評価の例 3

入力文	わが国の刑務所制度はとんでもなく時代遅れなものとなっている。	
参照文	Our prison system has become an absurd anachronism .	評価結果
Joint Simpson	The prison system of this country is out of date .	
PMI Jaccard Dice	Our prison system is in a sad tone .	×

表 6.5 の例では、「時代遅れ」の意味を表現できているかが評価結果に影響している。Joint, Simpson の出力文を訳すると「この国の刑務所制度は時代遅れた。」となる。入力文との細かい違いはあるものの、入力文の意味をほぼ表現できているため、評価 として。

PMI, Jaccard, Dice の出力文は「時代遅れ」が “in a sad tone” となっている。“in a sad tone” は「悲しそうな口調」という意味で用いられる連語である。よって、入力文の意味と大きく離れている。

表 6.6: 人手評価の例 4

入力文	彼女の結婚は一大行事となるだろう。	
参照文	Her marriage will be quite an event .	評価結果
全ての計算法	Her marriage would be a great event .	

表 6.6 の例では、“would” の時制が評価結果に影響している。入力文の時制は未来形となっている。入力文との細かい違いはあるものの、入力文の意味をほぼ表現できているため、評価 として。

PMI, Jaccard, Dice の出力文は「時代遅れ」が “in a sad tone” となっている。“in a sad tone” は「悲しそうな口調」という意味で用いられる連語である。よって、入力文の意味と大きく離れている。

表 6.7: 人手評価の例 5

入力文	賛成が 50 票、反対が 15 票あった。	
参照文	There were fifty votes for and fifteen votes against .	評価結果
全ての計算法	Fifty votes in favor , and fifteen votes opposed .	

表 6.7 の例は，表 5.10 と同一の入力文である．表 5.10 の結果と比較すると，人手評価で とした出力文が存在する．この例のように，Open-NMT の第 1 候補より優れた出力文を，出力できた入力文が存在する．

表 6.8: 人手評価の例 6

入力文	彼は彼女に慰めのことばをかけた。	
参照文	He spoke comforting words to her .	評価結果
Jaccard Dice	He put a few words of comfort to her .	
Joint	He spoke a few words of comfort to her , he said .	×
PMI Simpson	He spoke a few words of comfort to him .	×

表 6.8 の例は，表 5.11 と同一の入力文である．表 5.11 の結果と比較すると，人手評価で とした出力文が存在しない．8.2 節において，PMI の出力文が出力された原因を考察している．

## 第7章 実験のまとめ

本章では第5章と第6章の結果から判明した事項を述べる．最後に総評として，本論で述べた5つの翻訳確率の計算方法から，RSMTで用いる最適な方法を決定する．

### 7.1 第5章のまとめ

5章の実験では，翻訳確率の各計算方法の性能を比較するための実験を行った．その結果，Jointは他4つの計算方法より性能が若干劣ることが分かった．また，Joint以外の計算方法間に性能差はほぼなかった．

### 7.2 第6章のまとめ

6章の実験では，候補文に用いるOpen-NMTの出力の順位を引き下げることによって，第1候補以外の優れた候補文を選択できるか調査した．その結果，PMI以外の計算方法は一律に性能が低下した．しかし，PMIは他4つと異なり性能が向上した．

### 7.3 総評

総評すると，第5章の実験でJoint以外の計算方法にはほぼ差がなく，第6章の実験でPMIが第1候補以外の優れた候補文を選択できることが判明した．よって，RSMTで用いる翻訳確率の計算方法は，PMIが最も適している．

## 第8章 考察

### 8.1 Open-NMT の 1best と RSMT(第5章) の性能比較

RSMT(第5)の性能を評価するために Open-NMT の 1best と性能を比較した。また、人手評価の結果から RSMT が誤った出力をする原因を考察した。実験に使用するデータは第6章と同様とする。比較のために用いる RSMT の結果は、人手評価で最もが多かった PMI の結果を用いる。また、第8章の人手評価は対比較評価で行う。人手評価の基準を下記に示す。

- RSMT : RSMT の出力文が、NMT の出力文より優れている
- NMT : NMT の出力文が、RSMT の出力文より優れている
- 差なし : 出力文の優劣に差がつけがたい
- 同一 : 出力文が同一である

### 8.1.1 比較結果

Open-NMT の 1best と RSMT の , 自動評価と人手評価の結果をそれぞれ表 8.1 と表 8.2 に示す .

表 8.1: 自動評価結果 (1000 文)

	BLEU	METEOR	TER
Open-NMT	<b>0.191</b>	0.453	0.628
RSMT	0.183	<b>0.484</b>	<b>0.583</b>

表 8.2: 人手評価結果 (100 文)

RSMT	21
NMT	17
差なし	49
同一	13

表 8.1 , 表 8.2 によると , RSMT の方が NMT より優れた結果を示した .

### 8.1.2 人手評価の実例

人手評価の例を表 8.3 から表 8.6 に示す。また，RSMT が誤った候補文を選んだ原因を考察する。

表 8.3: 人手評価の例 1

入力文	この本にそう書いてある。
参照文	This book says so .
Open-NMT	This book is written in the book .
RSMT	This book says so .
類似学習文 (日本語)	この本は易しく書いてある。
類似学習文 (英語)	This book is simply written .
人手評価の結果	RSMT

表 8.3 の例では，過剰生成が RSMT によって改善された。RSMT の出力文は，参照文と一致している。NMT の出力文は，“the book” の過剰生成によって評価を下げている。

表 8.4: 人手評価の例 2

入力文	地球は太陽の周囲を回転する。
参照文	The earth revolves around the sun .
Open-NMT	The earth turns around the sun .
RSMT	The earth revolves around the sun .
類似学習文 (日本語)	地球は自転する。
類似学習文 (英語)	The earth rotates .
人手評価の結果	RSMT

表 8.4 の例では，動詞が RSMT によって改善された。RSMT の出力文は，参照文と一致している。

NMT の出力文は，“revolves” ではなく “turns” を用いている。“turns around” は「振り向く」という意味となる。よって，入力文の意味から離れている。

表 8.5: 人手評価の例 3

入力文	彼はゴルバチョフに次ぐ実力者だ。
参照文	He is second in power to Gorbachev .
Open-NMT	He is next person to Gorbachev .
RSMT	He is the best donor .
類似学習文 (日本語)	彼は音に対して鈍感だ。
類似学習文 (英語)	He is dull to sound .
人手評価の結果	NMT

表 8.5 では，RSMT の出力文中に，固有名詞である “Gorbachev” が出力文中に登場していない．固有名詞は学習文対における登場回数が少なく，RSMT で正しく評価がしにくい．RSMT のスコア順位では 5 番目に “Gorbachev” が登場する候補文が出現した．

表 8.6: 人手評価の例 4

入力文	彼はおこづかいの大半を本に使う。
参照文	He spends most of his pocket money on books .
Open-NMT	He use most of his allowance to book .
RSMT	He use most of his pocket money .
類似学習文 (日本語)	彼は金をどんどん使う。
類似学習文 (英語)	He is prodigal with money .
人手評価の結果	NMT

表 8.6 では，RSMT は “book” が出力文中に登場していない．これは類似学習文対との類似度が高かったことで，翻訳確率によるスコア付けの重みが低かったことが原因である．RSMT の評価では 1 位から 8 位までの候補中，7 文が表 8.5 に示した類似学習文との類似度を用いていた．

## 8.2 Open-NMT の 1best と RSMT(第 6 章) の性能比較

RSMT(第 6) の性能を評価するために Open-NMT の 1best と性能を比較した。また，人手評価の結果から RSMT が誤った出力をする原因を考察した。実験に使用するデータは第 5 章と同様とする。比較のために用いる RSMT の結果は，人手評価で最もが多かった PMI の結果を用いる。

### 8.2.1 比較結果

Open-NMT の 1best と RSMT の，自動評価と人手評価の結果をそれぞれ表 8.7 と表 8.8 に示す。

表 8.7: 自動評価結果 (1000 文)

	BLEU	METEOR	TER
Open-NMT	<b>0.191</b>	0.453	0.628
RSMT	0.189	<b>0.458</b>	<b>0.613</b>

表 8.8: 人手評価結果 (100 文)

RSMT	24
NMT	15
差なし	49
同一	12

表 8.7，表 8.8 によると，RSMT の方が NMT より優れた結果を示した。

## 8.2.2 人手評価の実例

人手評価の例を表 8.9 から表 8.12 に示す。また、RSMT が誤った候補文を選んだ原因を考察する。

表 8.9: 人手評価の例 1

入力文	鞆が肩に掛かっている。
参照文	The bag is hung over his shoulders .
Open-NMT	My bag is stuck in my shoulder .
RSMT	The bag is on my shoulder .
類似学習文 (日本語)	箱を肩にかついで運んだ。
類似学習文 (英語)	I carried the box on my shoulder .
人手評価の結果	RSMT

表 8.9 の例では、前置詞の使い分けが RSMT によって改善された。Open-NMT の出力文は “stuck in” が入力文の意味と不一致である。RSMT の出力文は “on” が用いられており、Open-NMT と比較すると入力文の意味と合致している。

表 8.10: 人手評価の例 2

入力文	彼も一端の音楽家になった。
参照文	He has become a pretty good musician .
Open-NMT	He was given a great musical at random .
RSMT	He became a musician .
類似学習文 (日本語)	音楽家に生まれついた。
類似学習文 (英語)	He was meant for a musician .
人手評価の結果	RSMT

表 8.10 の例では、目的語が RSMT によって改善された。Open-NMT の出力文は「ミュージカルを与えられた」となり、入力文の意味と不一致である。RSMT の出力文は「音楽家になった」となり、Open-NMT と比較すると入力文の意味と合致している。

表 8.11: 人手評価の例 3

入力文	良心が彼女を苦しめた。
参照文	Her conscience stung her .
Open-NMT	Her conscience worried her .
RSMT	My conscience worried her .
類似学習文 (日本語)	良心がとがめる。
類似学習文 (英語)	My conscience pricksme .
人手評価の結果	NMT

表 8.11 では、RSMT は所有格が誤っている。これは日本語と英語の所有格の明記の差異が原因で、誤っている候補文の類似度が正しい候補文より高くなっている。

表 8.12: 人手評価の例 4

入力文	彼は彼女に慰めのことばをかけた。
参照文	He spoke comforting words to her .
Open-NMT	He spoke a few words of comfort to her .
RSMT	He spoke a few words of comfort to him .
類似学習文 (日本語)	彼は懐かしさを隠しきれぬ表情だった。
類似学習文 (英語)	He looked as if everything were indescribably dear to him .
人手評価の結果	NMT

表 8.12 では主語と目的語が明記してあるにも関わらず、RSMT が誤っている。これは、入力文及び RSMT の出力文が、学習文対「彼は懐かしさを隠しきれぬ表情だった。」“He looked as if everything were indescribably dear to him .” との類似度が高かったことが原因である。

計算方法の都合、学習文対及び候補文の単語数が多いほど、翻訳確率が高くなる。よって、単語数が多い候補文が選択されやすい。

### 8.3 単語数の多さが原因である誤りの解決法について

本節では，単語数の多さが原因である誤りを解決する方法を提案する．

解決方法の1つは，文の単語数によって正規化を行うことである．正規化を行うことで，単語数が多いほど選択されやすい問題を解決できる．

2つめの解決方法として，翻訳確率の計算式に  $\text{argmax}$  を用いる方法を提案する．各日本語単語について最多の共起回数を持つ英語単語のみの組み合わせ情報を用いることで，問題を解決する．式 8.1 に，PMI の計算式に  $\text{argmax}$  を導入した計算式を示す．

$$\sum_{j=1}^M \log_2 \sum_{i=1}^N \frac{\max(Pair)}{\text{count}(\arg \max_{J_j} Pair) \text{count}(\arg \max_{E_i} Pair)} \quad (8.1)$$

$$Pair : \text{count}(J_j, E_i)$$

## 8.4 過剰生成を含む候補文の選択について

NMT の出力文の特徴の 1 つとして「過剰生成」がある。これは表 8.13 に示すように、同じ単語が複数回、出力に現れることを指す。過剰生成は NMT の出力数を増やすほど、現れる傾向がある。多くの場合、RSMT は過剰生成を含む候補文を選択しないことが望ましい。

表 8.13: 過剰生成を含む出力の例

入力文	左舷に塔が見えてきた。
参照文	We opened a tower to the port .
出力	The tower was sighted in sight in sight .

しかし第 6 章の実験において、過剰生成を含む候補文が選択される入力文が増加した。また、過剰生成が原因で人手評価を  $\times$  とした数を表 8.14 に示す。

表 8.14: 過剰生成が理由で人手評価  $\times$  とした数

PMI	6
Joint	26
Jaccard	7
Dice	14
Simpson	20

それぞれの手法で過剰生成が精度低下の原因となった入力が存在する。その中でも、Joint と Simpson は過剰生成が精度低下の原因として多かったことを表 8.14 が示している。この問題は式 (8.1) を用いることで解決できる可能性がある。

## 8.5 類似度の計算方法

本論では、翻訳確率の計算方法の調査を行った。しかし8.1節で示したように、RSMTの誤りの理由の1つとして類似度の計算方法が挙げられる。また、RSMTにおいて類似度の計算方法も翻訳確率と同様に、様々な方法が考えられる。今後の研究では、類似度の計算方法についても調査が必要である。

## 8.6 出力文の決定に用いる計算式

本論では、出力文の決定の時に式(3.4)を用いた。しかし、この時に用いる式によって結果が変化する。今回用いたパラメータは4種類あるので、それぞれのパラメータの使用または不使用の組み合わせだけでも16通りの計算式が考えられる。また、8.1節の問題を解決するならば、翻訳確率に重みを付与することも考えられる。今後の研究では、式(3.4)を用いた時以上の性能を得られる式を調査する。

## 第9章 おわりに

相対的意味論に基づく統計機械翻訳 (RSMT) に最適な翻訳確率の計算方法の調査を目的として、5つの計算方法を用いた翻訳結果を比較した。実験により、自己相互情報量 (PMI) が他の計算方法とほぼ同等の性能であると分かった。さらに PMI は、Open-NMT の第1候補以外の優れた候補文を選択できることが分かった。よって、今回述べた計算方法の中では、PMI が RSMT で用いるのに最も適している。今後は、第8章で述べた課題の解決を目指す。

# 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様々に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] 村上仁一, “相対的意味論と機械翻訳の応用” 言語処理学会第 27 回年次大会, 2021 年.
- [2] Franz Josef Och, Hermann Ney, “A Systematic Comparison of Various Statistical Alignment Models”, *Computational Linguistics*, 29(1), pp.299-314, 1996.
- [3] 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 鳥取大学卒業論文, pp.3-16, 2016.
- [4] 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”, 言語処理学会第 20 回年次大会予稿集, pp.951-954, 2014.
- [5] 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, 2015.
- [6] Vladimir Iosifovich Levenshtein, “Binary codes capable of correcting deletions, insertions, and reversals”, *Soviet Physics Doklady*, 10(8), pp.707-710, 1966.
- [7] Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, 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, 2007.
- [8] 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.
- [9] 今仁 優希, “ニューラル機械翻訳における低頻度語の処理”, 平成 30 年度 鳥取大学大学院修士論文, pp.2-5, 2017 .

- [10] 安場裕人, 村上仁一, “変換主導型翻訳の提案”, 言語処理学会第 24 年次大会, 2018.
- [11] Church K., Hanks P., “Word association norms, mutual information, and lexicography” *Computational Linguistics*, 16(1):22-29, 1990 .
- [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”, 言語処理学会第 25 回年次大会, 2019.
- [13] Salton G., McGill M., “Introduction to Modern Information Retrieval”, McGraw-Hill, 1983.
- [14] Smadja F., McKeown K. R., Hatzivassiloglou V., “Translating collocations for bilingual lexicons: a statistical approach” *Comput Linguist.*, 22(1):1-38, 1996.
- [15] Bollegala D., Matsuo Y., Ishizuka M., “A web search engine-based approach to measure semantic similarity between words” *IEEE Trans. on Knowl. and Data Eng.*, 23(7):977-990, 2011.
- [16] Kishore Papineni, Salim Roukos, Todd Ward, Wei-Jing Zhu, “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.
- [17] Satanjeev Banerjee, Alon Lavie, “METEOR:An Automatic Metric for MT Evaluation with High Levels of Correlation with Human Judgemnts”, *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, pp.65-72. 2005.
- [18] Richard Schwartz, Linnea Micciulla, John Makhoul, “A Study of Translation Edit Rate with Targeted Humman Annotation”, *AMTA*, 2006.
- [19] 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.