概要

機械翻訳の一種であるパターン翻訳は，人手作成した単語辞書と対訳文パターンを用いて翻訳を行う．この翻訳方式は，入力文が適切な文パターンに適合した場合，翻訳精度の高い文を得やすい傾向にある．しかし，単語辞書と対訳文パターンは人手で作成するため，開発にコストがかかる．

そこで江木らは，単語辞書と対訳文パターンを統計的手法で自動的に作成する “パターンに基づく統計翻訳” を提案した．この手法は，対訳句の抽出において，対訳文と単語レベル文パターンを照合して得る方法が用いられる．しかし，単語レベル文パターンから全ての可能な対訳句を出力するために，不適切な対応をとる対訳句が多く出力されていた．そこで本研究では，対訳句の変数全体の確率値を利用して最適な対訳句を出力する，PATTERN 法と SENTENCE 法の 2 手法を用いる．これにより，対訳句の出力数を抑制し，不適切な対応をとる対訳句の数を減らすことで対訳句の精度向上を目指す．また，出力した対訳句を用いて日英統計翻訳を行い，翻訳精度の調査を行う．

実験の結果，対訳句の出力数が大幅に削減できており，対訳句の精度の向上が確認できた．また，翻訳文 100 文より提案手法の対比較評価を行い，PATTERN 法○が 14 文，SENTENCE 法○が 19 文となり，わずかに SENTENCE 法の翻訳精度が高い結果となった．

目 次

図 目 次

表 目 次

第1章　はじめに

パターン翻訳 [1] は，機械翻訳における一種の翻訳方式である．入力文が適切な文パターンと照合した場合，翻訳精度の高い文を出力するが，適合しない場合は翻訳ができない．また，翻訳に用いる単語辞書と対訳文パターンを人手で作成するために，開発にコストがかかる．

そこで江木らは，単語辞書と対訳文パターンを統計的手法を用いて自動的に作成する“パターンに基づく統計翻訳 [2]”(以下, 従来手法) を提案した．この従来手法は，対訳句の抽出において，対訳文と単語レベル文パターンを照合して得る方法が用いられる．しかし，単語レベル文パターンから全ての可能な対訳句を出力するために，不適切な対応をとる対訳句が多く出力されていた．そこで本研究では，対訳句の変数全体の確率値を利用して最適な対訳句を出力する，PATTERN 法と SENTENCE 法の 2 手法を用いる．これにより，対訳句の出力数を抑制し不適切な対応をとる対訳句の数を減らすことで，対訳句の精度向上を目指す．また，出力した対訳句を用いて日英統計翻訳を行い，翻訳精度の調査を行う．

実験の結果，対訳句の出力数が大幅に削減できており，対訳句の精度の向上が確認できた．また，翻訳文 100 文より提案手法の対比較評価を行い，PATTERN 法○が 14 文，SENTENCE 法○が 19 文となり，わずかに SENTENCE 法の翻訳精度が高い結果となった．

第2章　翻訳システム

2.1　概要

本章は，江木の論文 [2] を参照して既述している．
現在，最も主流となっている翻訳システムとして “句に基づく統計翻訳” がある．句に基づく統計翻訳は，学習データとして対訳文を与えるだけで翻訳ができる．このため，翻訳にかかるコストが低い．さらに，対訳文から単語辞書と単語翻訳確率を自動的に得ることが可能である．

一方，翻訳システムの一手法として “パターン翻訳” がある．パターン翻訳は大量の対訳文パターンと単語辞書を用いて，翻訳文を出力する手法である．パターン翻訳は，入力文が適切な対訳文パターンに適合した場合に，翻訳精度の高い翻訳文が得られやすいという特徴がある．しかし，パターン翻訳に用いる単語辞書と対訳文パターンは人手で作成するため，開発コストが高くなる．

そこで江木らは，単語辞書と対訳文パターンを統計的手法で自動的に作成し翻訳する従来手法を提案した．従来手法は，句に基づく統計翻訳の特徴である対訳文から単語辞書と単語翻訳確率を自動的に取得できる点に着目し，翻訳に用いる単語辞書と対訳文パターンを統計的手法を用いて自動的に作成する．

本章ではまず，現在最も主流の翻訳システムである句に基づく統計翻訳について説明する．次に，パターン翻訳について説明し，最後に江木らによって提案された従来手法について説明する．

2.2　句に基づく統計翻訳

句に基づく統計翻訳は，機械翻訳の一手法である．最初は，“単語に基づく統計翻訳” が用いられた．しかし，単語に基づく統計翻訳より句に基づく統計翻訳の方が精度が高いことから，現在は句に基づく統計翻訳が主流となっている．この句に基づく統計翻訳は，学習データとして大量の対訳文を用いることで，自動的に翻訳規則を生成し翻訳を

行う.

2.2.1 句に基づく統計翻訳の概要

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．日英統計翻訳は，日本語入力文 j が与えられたとき，翻訳モデルと言語モデルの組み合わせの中から確率が最大となる英語翻訳文 $\hat{e}$ を検索することで翻訳を行う．基本モデルを (2.2) 式に示す.

$$e = \arg\max_{e} P(e|j) \tag{2.1}$$

$$\simeq \arg\max_{e} P(j|e)P(e) \tag{2.2}$$

ここで，$P(j|e)$ は翻訳モデル，$P(e)$ は言語モデルを表す．$P(e)$ 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．翻訳モデルは対訳学習文，言語モデルは目的言語の単言語学習文から学習する．そして，デコーダを用いて $P(j|e)P(e)$ が最大となる英語翻訳文 $\hat{e}$ を検索する．日英統計翻訳の手順を図 2.1 に示す.

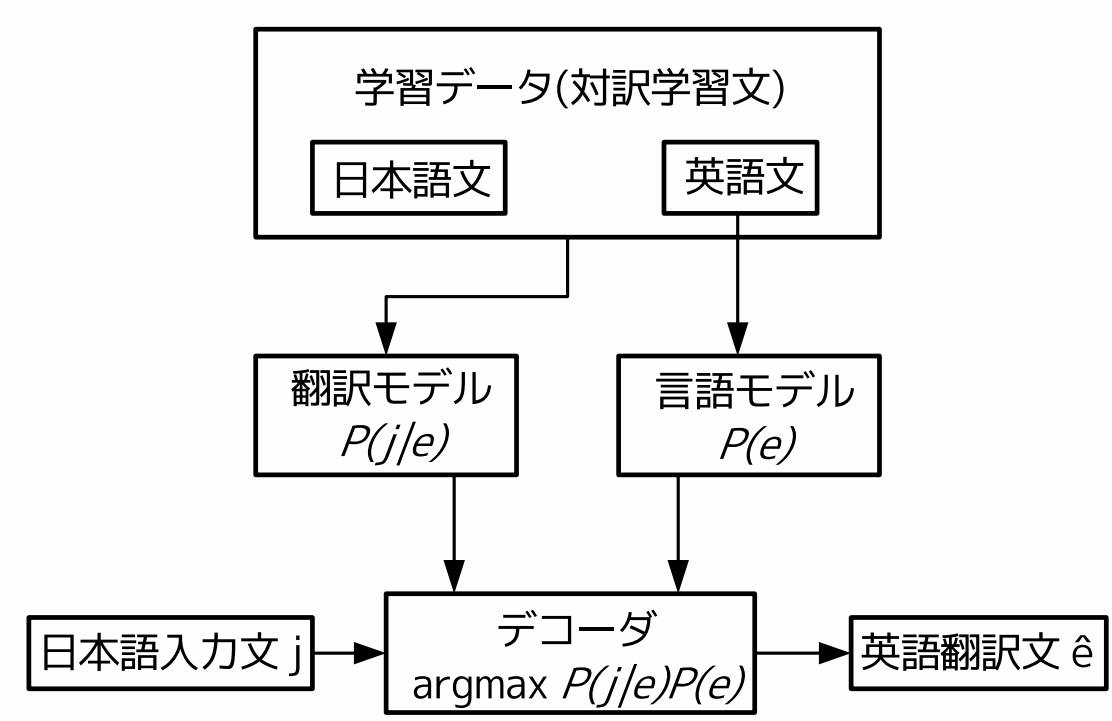

図 2.1: 日英統計翻訳の手順

2.2.2 翻訳モデル

翻訳モデルは，日本語単語列から英語単語列へ確率的に翻訳を行うためのモデルである．翻訳モデルには，単語に基づく翻訳モデルと句に基づく翻訳モデルが存在する．初期の統計翻訳では，単語に基づく翻訳モデルを用いていた．しかし，翻訳精度の高さから現在は句に基づく翻訳モデルが主流となっている．句に基づく翻訳モデルは，一般的にフレーズテーブルで管理される．句に基づく翻訳モデルの作成手順を以下に示す．

手順 1 後述する IBM モデルを用いて，単語対応づけをする．

手順 2 ヒューリスティックなルールを用いて，句に基づく対応づけをする．

手順 3 手順 2 で求めた句に基づく対応づけから，フレーズテーブルを作成する．

表 2.1 にフレーズテーブルの例を示す．

表 2.1: フレーズテーブルの例

日本語句	英語句	$P(j\|e)$	$\Pi P(j\|e)$	$P(e\|j)$	$\Pi P(e\|j)$
あの 人	That person	0.716	0.182	0.157	0.012
自由 に	free	0.014	0.011	0.153	0.101
見 に 行く	go to see	0.333	0.003	0.500	0.004

ここで，フレーズテーブルは左から順に日本語句，英語句，英日方向の翻訳確率 $P(j|e)$，英日方向の単語翻訳確率の積 $\Pi P(j|e)$，日英方向の翻訳確率 $P(e|j)$，日英方向の単語翻訳確率の積 $\Pi P(e|j)$ である．

2.2.3 IBM 翻訳モデル

統計翻訳における単語対応を得るための代表的なモデルとして，IBM 翻訳モデル [5] がある．IBM 翻訳モデルは，仏英翻訳を前提としている．しかし，本研究では日英翻訳を扱うため，原言語文を日本語文 J，目的言語文を英語文 E と定義する．IBM 翻訳モデルにおいて，日本語文 J と英語文 E の翻訳モデル $P(J|E)$ を計算するため，アライメント a と呼ばれる概念を導入する．アライメントはある日本語単語 j と英単語 e の対応関係を意味する．IBM 翻訳モデルの基本的な計算式を (2.3) 式に示す．

$$P(J|E) = \sum P(J, a|E) \tag{2.3}$$

IBM 翻訳モデルにおいて，各日本語単語に対応する英単語は 1 つであるのに対し，各英単語に対応する日本語単語は 0 から n 個あると仮定する．また，日本語単語に対応する適切な英単語がない場合，英語文の先頭に特殊文字 e_0 があると仮定し，日本語単語と対応させる．

モデル 1

(2.3) 式は，以下の式に分解することができる．m は日本語文の長さ，a_1^{i-1} は日本語文における，1 番目から $i-1$ 番目までのアライメント，j_1^{i-1} は日本語文における，1 番目から $i-1$ 番目まで単語を表している．

$$P(J,a|E) = P(m|E)\prod_{i=1}^{m} P(a_i|a_1^{i-1}, j_1^{i-1}, m, E)P(j_i|a_1^{i}, j_1^{i-1}, m, E) \tag{2.4}$$

(2.4) 式は，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．

- 日本語文の長さの確率 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

$$P(a_i|a_1^{i-1}, j_1^{i-1}, m, E) = (l+1)^{-1}$$

- 日本語の翻訳確率 $t(j_i|e_{a_i})$ は，日本語単語 j_i に対応する英単語 e_{a_i} に依存する

$$P(j_i|a_1^{i}, j_1^{i-1}, m, e) = t(j_i|e_{a_i})$$

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

$$P(J,a|E) = \frac{\epsilon}{(l+1)^m}\prod_{i=1}^{m} t(j_i|e_{a_i}) \tag{2.5}$$

$$P(J|E) = \frac{\epsilon}{(l+1)^m}\sum_{a_1=0}^{l}\cdots\sum_{a_m=0}^{l}\prod_{i=1}^{m} t(j_i|e_{a_i}) \tag{2.6}$$

$$= \frac{\epsilon}{(l+1)^m}\prod_{i=1}^{m}\sum_{k=0}^{l} t(j_i|e_{a_i}) \tag{2.7}$$

モデル1では翻訳確率 $t(j|e)$ 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 $t(j|e)$ の初期値を設定する．

手順 2 日英対訳対 $(J^{(s)}, E^{(s)})$(但し，$1 \leq s \leq S$) において，日本語単語 j 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．

$$c(j|e; J, E) = \frac{t(j|e)}{t(j|e_0) + \cdots + t(j|e_l)} \sum_{i=1}^{m} \delta(j, j_i) \sum_{k=0}^{l} \delta(e, e_k) \tag{2.8}$$

$\delta(j, j_i)$ は日本語文 J 中で日本語単語 j が出現する回数，$\delta(e, e_j)$ 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 $t(j|e)$ を計算する．

1. 定数 λ_e を以下の式により計算する．

$$\lambda_e = \sum_{j} \sum_{s=1}^{S} c(j|e; J^{(s)}, E^{(s)}) \tag{2.9}$$

2. (2.9) 式より求めた λ_e を用いて，翻訳確率 $t(j|e)$ を再計算する．

$$\begin{aligned} t(j|e) &= \lambda_e^{-1} \sum_{s=1}^{S} c(j|e; J^{(s)}, E^{(s)}) \\ &= \frac{\sum_{s=1}^{S} c(j|e; J^{(s)}, E^{(s)})}{\sum_{j} \sum_{s=1}^{S} c(j|e; J^{(s)}, E^{(s)})} \end{aligned} \tag{2.10}$$

手順 4 翻訳確率 $t(j|e)$ 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，i 番目の日本語単語 j_i と対応する英単語の位置 a_i は英語文の長さ l に加えて，i と，日本語文の長さ m に依存し，以下のような関係とする．

$$a(a_i|i,m,l) \equiv P(a_i|a_1^{i-1}, j_1^{i-1}, m, l) \tag{2.11}$$

この関係からモデル 1 における (2.6) 式は，以下の式に変換できる．

$$\begin{aligned} P(J|E) &= \epsilon \sum_{a_1=0}^{l} \cdots \sum_{a_m=0}^{l} \prod_{i=1}^{m} t(j_i|e_{a_i}) a(a_i|i,m,l) && (2.12) \\ &= \epsilon \prod_{i=1}^{m} \sum_{k=0}^{l} t(j_i|e_{a_i}) a(a_i|i,m,l) && (2.13) \end{aligned}$$

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

$$\begin{aligned} c(j|e;J,E) &= \frac{t(j|e)}{t(j|e_0)+\cdots+t(j|e_l)} \sum_{i=1}^{m} \delta(j,j_i) \sum_{k=1}^{l} \delta(e,e_k) && (2.14) \\ &= \sum_{i=1}^{m} \sum_{k=0}^{l} \frac{t(j|e)a(k|i,m,l)\delta(j,j_i)\delta(e,e_k)}{t(j|e_0)a(0|i,m,l)+\cdots+t(j|e_l)a(l|i,m,l)} && (2.15) \\ c(k|i,m,l;J,E) &= \sum_{a} P(a|E,J)\delta(k,a_i) && (2.16) \\ &= \frac{t(j_i|e_k)a(k|i,m,l)}{t(j_i|e_0)a(0|i,m,l)+\cdots+t(j_i|e_l)a(l|i,m,l)} && (2.17) \end{aligned}$$

$c(j|e;J,E)$ は対訳文中の英単語 e と日本語単語 j が対応付けされる回数の期待値，$c(k|i,m,l;J,E)$ は英単語の位置 k が日本語単語の位置 i 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 $a(k|i,m,l) = (l+1)^{-1}$ 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 $P(j|e)$
  英単語 e が日本語単語 j に翻訳される確率
- 繁殖確率 $n(\phi|e)$
  英単語 e が ϕ 個の日本語単語と対応する確率
- 歪み確率 $d(i|k,m,l)$
  英語文の長さ l，日本語文の長さ m のとき，k 番目の英単語 e_k が i 番目の日本語単語 j_i に翻訳される確率

さらに，英単語が日本語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．

$$P(\phi_0|\phi_1^l, E) = \left(\begin{array}{c} \phi_1 + \cdots + \phi_l \\ \phi_0 \end{array} \right) p_0^{\phi_1+\cdots+\phi_l-\phi_0} p_1^{\phi_0} \tag{2.18}$$

したがって，モデル 3 は以下の式で求められる．

$$\begin{aligned} P(J|E) &= \sum_{a_1=0}^{l} \cdots \sum_{a_m=0}^{l} P(J,a|E) && (2.19) \\ &= \sum_{a_1=0}^{l} \cdots \sum_{a_m=0}^{l} \left(\begin{array}{c} m-\phi_0 \\ \phi_0 \end{array} \right) p_0^{m-2\phi_0} p_1^{\phi_0} \prod_{k=1}^{l} \phi_k! n(\phi_k|e_k) \\ &\times \prod_{i=1}^{m} t(j_i|e_{a_i}) d(i|a_i,m,l) && (2.20) \end{aligned}$$

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 $d(i|k.m,l)$ 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日本語単語の中で，最も文頭に近い場合

$$P(\Pi_{[k]1} = i|\pi_1^{[k]-1}, \tau_0^l, \phi_0^l, E) = d_1(i - \odot_{k-1}|\mathcal{A}(e_{[k-1]}), \mathcal{B}(j_i)) \tag{2.21}$$

  $\odot_{k-1}$ は k-1 番目の英単語が対応する日本語単語の位置を表している．

- それ以外の場合

$$P(\Pi_{[k]x} = i|\pi_{[k]1}{}^{x-1}, \pi_1^{[k]-1}, \tau_0^l, \phi_0^l, E) = d_{>1}(i - \pi_{[k]x-1}|\mathcal{B}(j_i)) \tag{2.22}$$

  $\pi_{[k]x-1}$ は同じ英単語に対応している直前の日本語単語を表している．

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日本語単語の中で，最も文頭に近い場合

$$\begin{aligned} P(\Pi_{[k]1} &= i|\pi_1^{[k]-1}, \tau_0^l, \phi_o^l, E) \\ &= d_1(v_i|\mathcal{B}(j_i), v_{\odot_{k-1}}, v_m - \phi_{[k]} + 1)(1 - \delta(v_i, v_{i-1})) \end{aligned}$$

  v_i は i 番目までの空白数，$\mathcal{A}$ は英語の単語クラス $\mathcal{B}$ は日本語の単語クラスを表している．

- それ以外の場合

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

2.2.4 GIZA++

GIZA++[3] とは，統計翻訳のために作られた単語の確率の計算を行うツールである．IBM 翻訳モデルに基づいて，単語の対応関係の確率である単語翻訳確率を計算する．

2.2.5 フレーズテーブルの作成

GIZA++より IBM 翻訳モデルを推定することで最尤な単語対応を得る．これを日英，英日の両方向に対して行う．日本語文 “私は海を見に行く” とその対訳英語文 “I go to see the sea” を例に挙げ，日英方向の単語対応の例を表 2.2 に，英日方向の単語対応の例を表 2.3 に示す．また，●は単語が対応した箇所を示す．

表 2.2: 日英方向の単語対応の例

	I	go	to	see	the	sea
私	●					
は				●		
海						●
を				●		
見				●		
に				●		
行く				●		

表 2.3: 英日方向の単語対応の例

	I	go	to	see	the	sea
私	●					
は						
海						●
を					●	
見		●	●	●		
に						
行く						

次に，両方向の対応付けからヒューリスティックスなルールにより，1 対多の対応を認めた単語対応の計算を行う．ここで，ヒューリスティックとは人間の日々の意思決定類似した直感的かつ発見的な思考方法である．基本のヒューリスティックとして “intersection” と “union” がある．intersection は，日英方向と英日方向の両方向に単語対応が存在する場合にその単語対応を残す．union は日英方向と英日方向のどちらか一方に単語対応が存在する場合にその単語対応を残す．intersection の例を表 2.4 に，union の例を表 2.5 に示す．

表 2.4: intersection の例

	I	go	to	see	the	sea
私	●					
は						
海						●
を						
見				●		
に						
行く						

表 2.5: union の例

	I	go	to	see	the	sea
私	●					
は				●		
海						●
を				●	●	
見		●	●	●		
に				●		
行く				●		

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 の例を表 2.6 に，grow-diag の例を表 2.7 に示す．

表 2.6: grow の例

	I	go	to	see	the	sea
私	●					
は						
海						●
を				●		
見				●		
に						
行く						

表 2.7: grow-diag の例

	I	go	to	see	the	sea
私	●					
は						
海						●
を				●	●	
見			●	●		
に				●		
行く						

grow-diag の最終処理として “final” と “final-and” がある．final は少なくとも一方の言語に対応がない場合に，union の単語対応を追加し，final-and は両言語単語に対応がない場合に，union の単語対応を追加する方法である．grow-diag-final の例を表 2.8 に，grow-diag-final-and の例を表 2.9 に示す．

表 2.8: grow-diag-final の例

	I	go	to	see	the	sea
私	●					
は				●		
海						●
を				●	●	
見		●	●	●		
に				●		
行く				●		

表 2.9: grow-diag-final-and の例

	I	go	to	see	the	sea
私	●					
は						
海						●
を				●	●	
見		●	●	●		
に				●		
行く						

得られた単語対応うち，矛盾しない全ての対訳句を得る．このとき，対訳句に対して翻訳確率を計算し，対訳句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．

2.2.6 言語モデル

言語モデルは，単語列に生成確率を付与するモデルである．言語モデルは単言語学習文から学習される．統計翻訳では，N-gram モデルを用いる．

N-gram モデルは，“単語列 $w_1^n = w_1, w_2, ..., w_n$ の i 番目の単語 w_i の生起確率 $P(w_i)$ は直前の $(n-1)$ 単語に依存する” という仮説に基づくモデルである．単語列 w_1^n の生起確率 $P(w_i)$ の計算式を (2.23) 式に示す．

$$P(w_1^n) = \Pi_{i=1}^{n} P(w_i \mid w_{i-(N-1)}^{i-1}) \tag{2.23}$$

ここで $P(w_1^n)$ は，i から j 番目までの単語列を表す．例えば，“She is a teacher” という単語列に対して 2-gram モデルを適応した場合，単語列の生起確率は (**??**) 式で計算される．

$$P(\text{“}She\,is\,a\,teacher\,.\text{”}) \approx P(She) \times P(is \mid She) \times P(a \mid is) \times P(teacher \mid a) \quad (2.24)$$

2.2.7 デコーダ

デコーダは，翻訳モデルと言語モデルの全ての組み合わせから確率が最大となる翻訳文を検索し出力する．代表的なデコーダとして，Moses[6] がある．デコーダの手順を図 2.2 に示す．

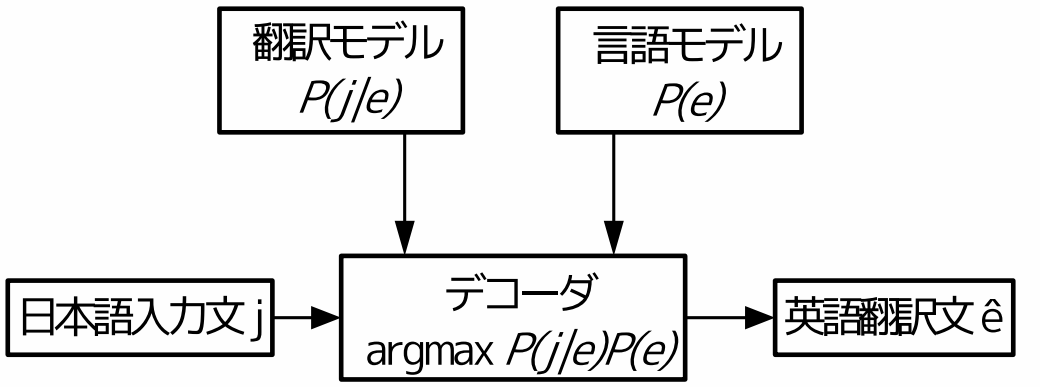

図 2.2: デコーダの手順

2.3 パターン翻訳

パターン翻訳は，機械翻訳の一手法であり，大量の単語辞書と対訳文パターンを用いて翻訳を行う．パターン翻訳は適切に対訳文パターンが適合した場合に，文全体の構造を保持した翻訳精度の高い翻訳文を出力する傾向にある．しかし，単語辞書や対訳文パターンを人手で作成するため，開発にコストがかかる．また，対訳文パターンに適合しない場合に翻訳ができない．

2.3.1 日英パターン翻訳の概要

一般的な日英パターン翻訳の手順を以下に示す.

手順 1 単語辞書と対訳文パターンを用意する．対訳文パターンとは，大量の対訳文から任意の単語やフレーズを変数化して得られる文パターンである.

手順 2 日本語入力文と日本語文パターンを照合する.

手順 3 変数部に対応する日本語単語を単語辞書を用いて，英単語に翻訳する.

手順 4 日本語文パターンに対応する英語文パターンの変数部を，翻訳した英単語に置き換える.

手順 5 手順 4 で得た英語翻訳文を出力する.

日英パターン翻訳の手順を図 2.3 に示す.

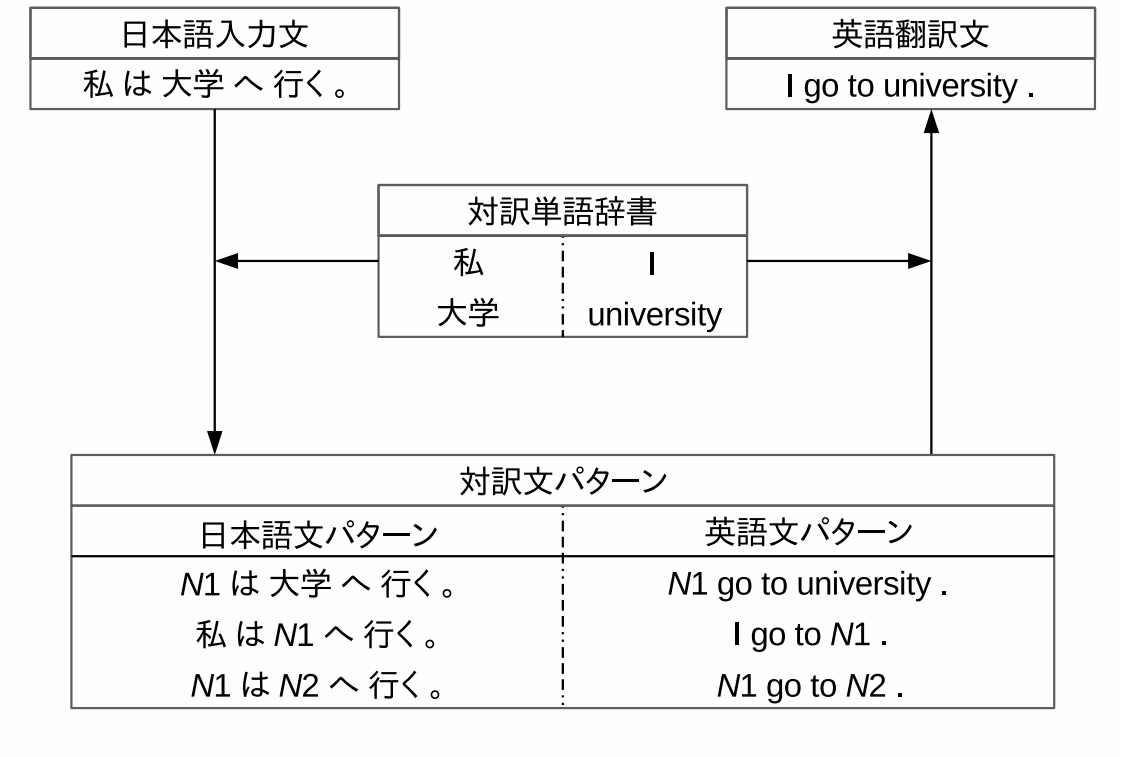

図 2.3: 日英パターン翻訳の流れ

2.4 従来手法

パターン翻訳は単語辞書と対訳文パターンを人手で作成するため，開発コストが高くなる．そこで江木らは，単語辞書と対訳文パターンを統計的手法で自動作成し翻訳を行う従来手法を提案した．この従来手法は，句に基づく統計翻訳の特徴である対訳文から対訳単語と単語翻訳確率を自動的に作成できる点に注目し，考案されている．

2.4.1 従来手法の概要

従来手法は，大きく分けて5つの手順で翻訳を行う. 日英翻訳における従来手法の概要を以下に示す.

手順 1 対訳単語の作成

GIZA++を用いて, 単語辞書を作成する.

手順 2 単語に基づく対訳文パターンの作成

対訳単語を用いて, 単語に基づく対訳文パターン (以下, 単語レベル文パターン) を作成する.

手順 3 対訳句の作成

単語レベル文パターンを用いて, 対訳句を作成する.

手順 4 句に基づく対訳文パターンの作成

対訳句を用いて, 句に基づく対訳文パターンを作成する.

手順 5 翻訳文の生成

句に基づく対訳文パターンを用いて, 翻訳文を生成する.

2.4.2 対訳単語の作成

GIZA++を用いて，対訳文の単語対応を取り，対訳単語と単語翻訳確率を得る．図 2.4 に対訳単語作成の例を示す.

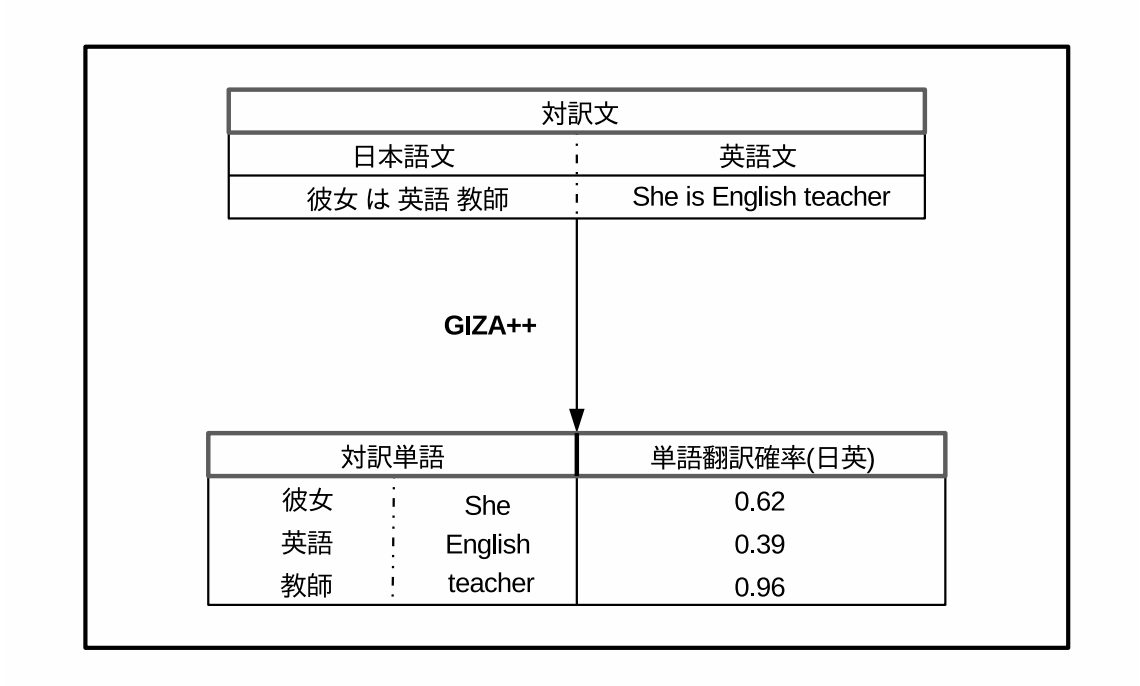

図 2.4: 対訳単語作成の例

2.4.3 単語レベル文パターンの作成

対訳単語と対訳文を用いて，単語レベル文パターンを作成する．まず，対訳単語と対訳文を照合する．そして，対訳文において，適合した対訳単語を変数化する．図 2.5 に単語レベル文パターン作成の例を示す.

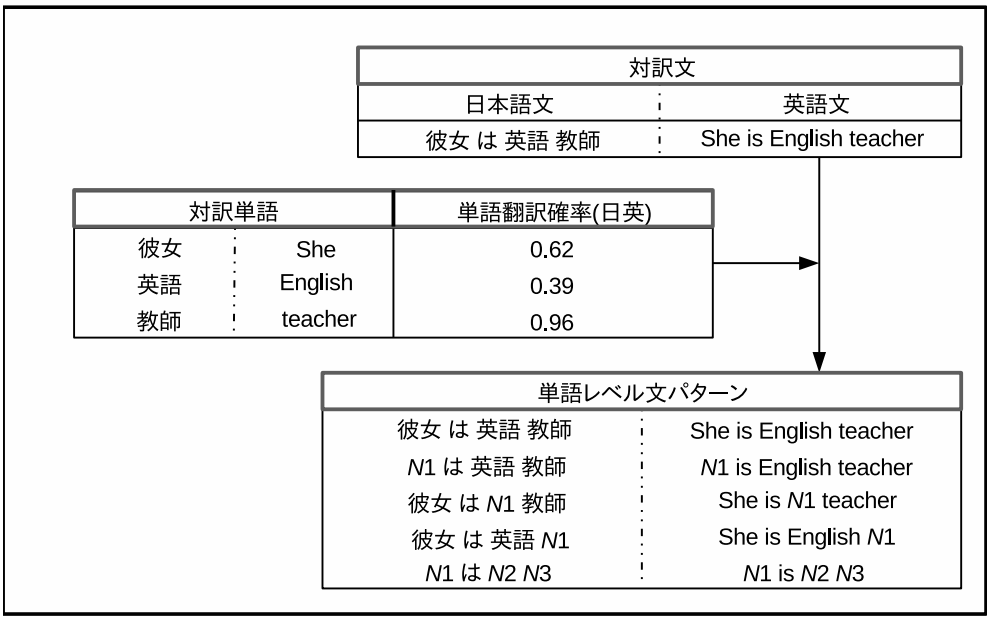

図 2.5: 単語レベル文パターン作成の例

2.4.4 対訳句の作成

a) 対訳句の抽出

まず、単語レベル文パターンと対訳文を照合する．そして適合した場合，単語レベル文パターンの変数部に対応する単語を，対訳句として対訳文より抽出する．図 2.6 に対訳句抽出の流れを示す.

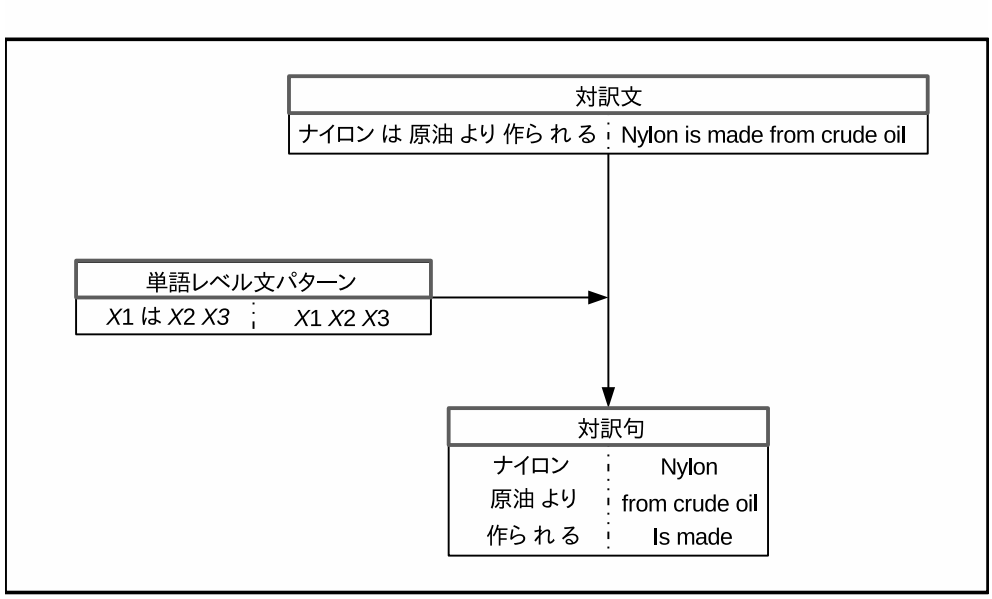

図 2.6: 対訳句の抽出例

b) 対数フレーズ確率の付与

対訳単語と単語翻訳確率を用いて，対訳句に確率を付与する．まず，対訳句において日本語句の単語と英語句の単語の全ての組み合わせを得る．次に，日本語句の単語に対応する英語句の単語の中で，対訳単語の単語翻訳確率 (日英) が最大となる値をとる．これを各日本語単語に対して行い，得られた値について対数の総和を求める．(以下，この値を対数フレーズ確率と呼ぶ)．同様に対訳句において，英単語に対応する日本語単語の中で単語翻訳確率の最大値を取得し，英日方向の対数フレーズ確率も求める．日英方向の対数フレーズ確率付与の例を図 2.7 に示す．

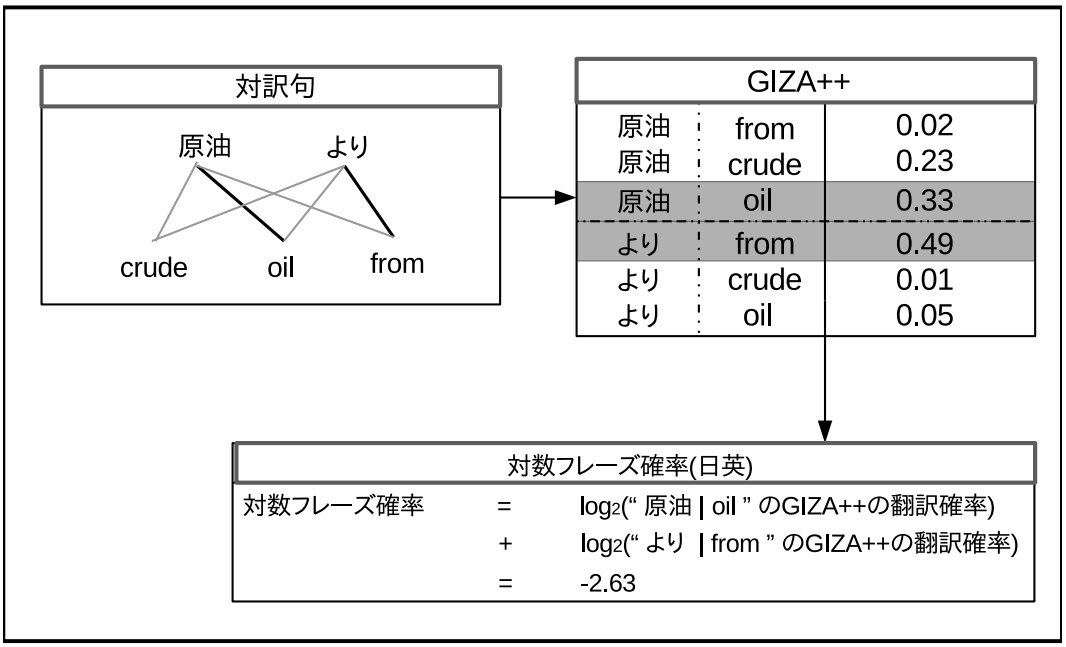

図 2.7: 対数フレーズ確率付与の例 (日英)

2.4.5 句に基づく対訳文パターンの作成

a) 対訳文パターンの作成

対訳句と対訳文を用いて，句に基づく対訳文パターン (以下，句レベル文パターン) を作成する．作成方法は，単語レベル文パターンの作成と同様に変数の組み合わせを考慮して，句レベル文パターンを可能な限り多く作成する．句レベル文パターン作成の例を図 2.8 に示す.

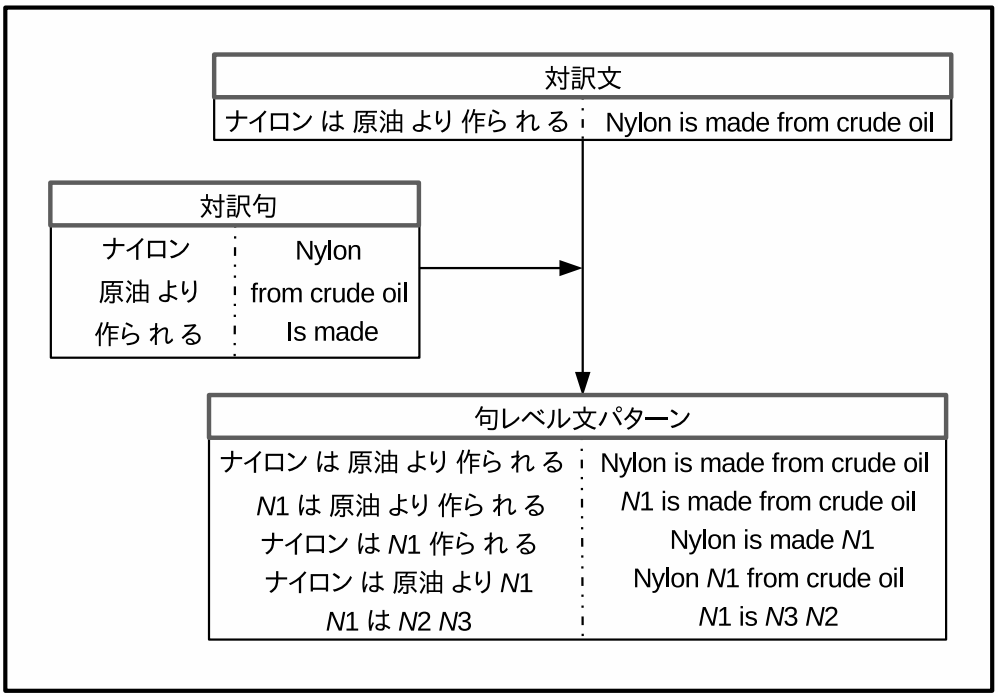

図 2.8: 句レベル文パターン作成の例

b) 対数文パターン確率の付与

対訳単語と単語翻訳確率を用いて，句レベル文パターンに確率を付与する．句レベル文パターンにおいて字面を用いて，対数フレーズ確率の付与と同様の計算手法で確率を求める．本研究では，この値を対数文パターン確率と呼ぶ．日英方向の対数文パターン確率付与の例を図 2.9 に示す.

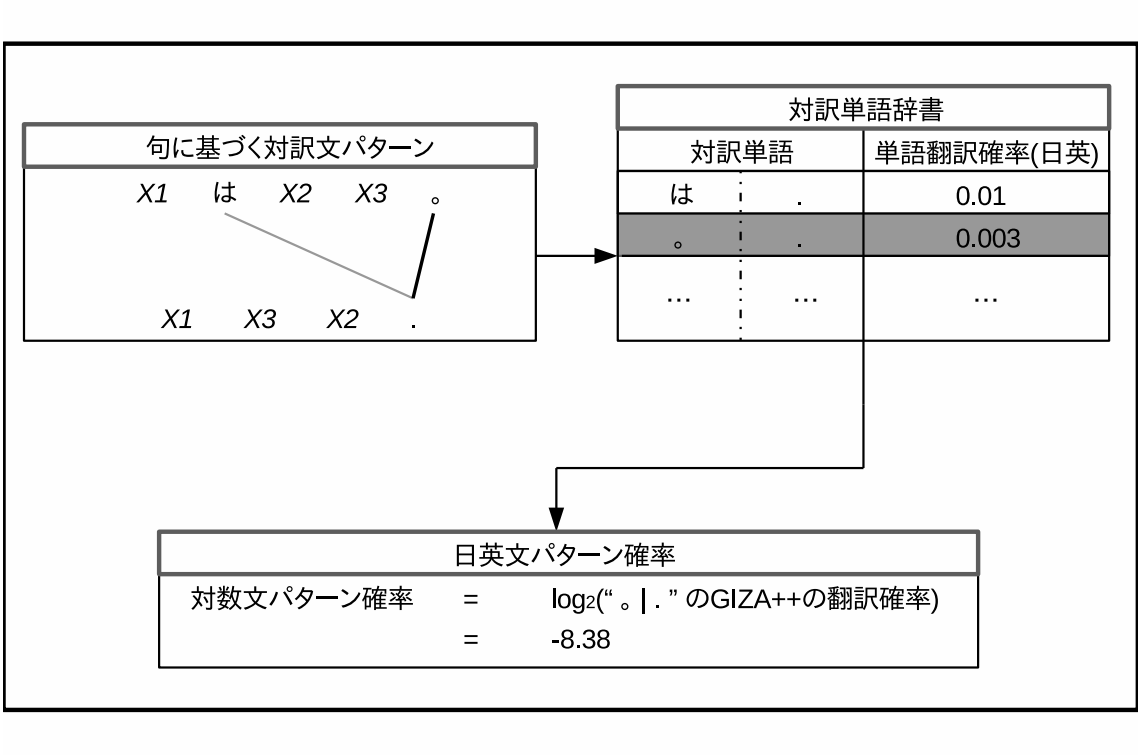

図 2.9: 対数文パターン確率付与の例 (日英)

2.4.6 翻訳文の生成

句レベル文パターンと対訳句を用いて，翻訳文を生成する．まず，日本語文パターンと入力文を照合し, 入力文に適合する日本語文パターンを選択する．なお，の本語文パターンの選択には，入力文と日本語文パターンの字面を比較し，字面が多く一致する文パターンを選択する．そして，選択した文パターンにおいて，英語文パターンの変数部に対訳句を用いて英語句を置換し，翻訳候補文を生成する．この処理を各適合する文パターンに対して同様に行う．最後に，各翻訳候補文から翻訳文を選択するために，対訳文パターンの対数文パターン確率 (α) と対訳句の対数フレーズ確率 (β)，言語翻訳確率 (trigram $=\gamma$) の総和を用いる．各翻訳候補文の対訳文パターンの対数文パターン確率と対訳句の対数フレーズ確率，言語翻訳確率 (trigram) の総和を求め，翻訳候補文の中で総和が最大となる文を翻訳文として出力する．日英翻訳における翻訳文の生成例を図 2.10

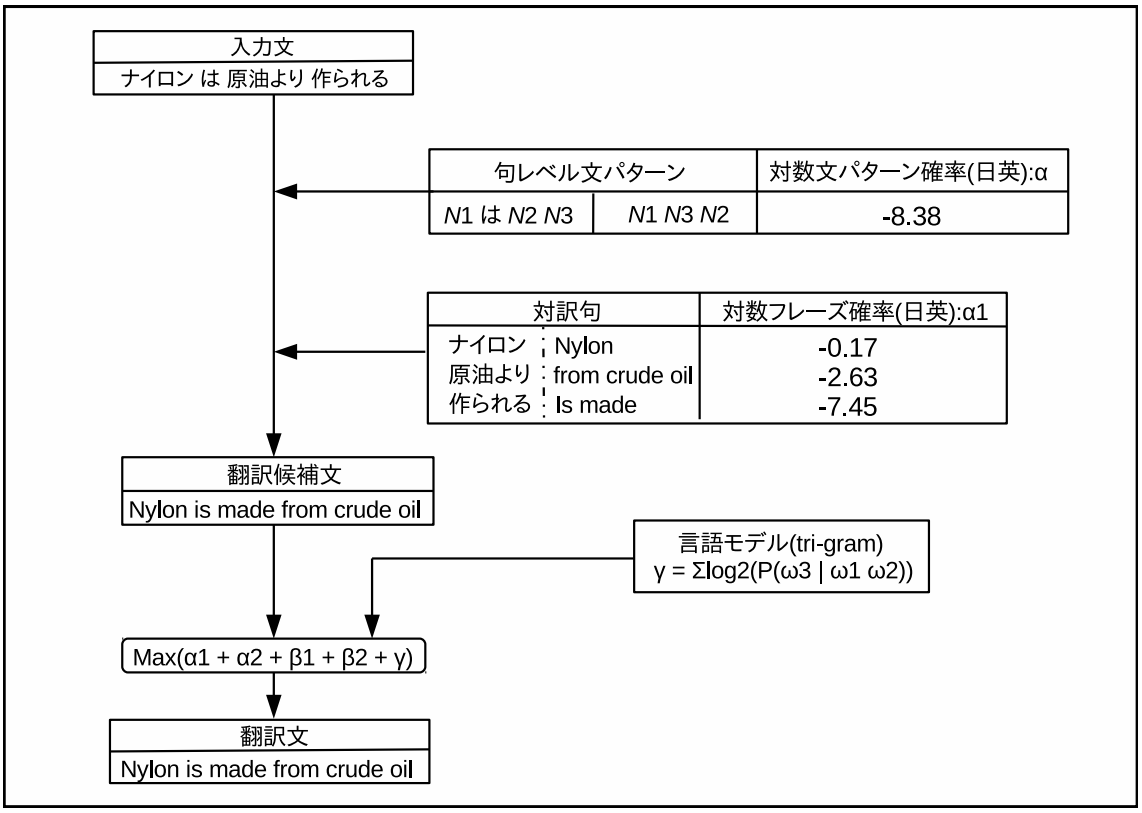

図 2.10: 対数文パターン確率付与の例 (日英)

2.5 従来手法の問題点

従来手法より出力した翻訳文を調査したところ，全体的に見て不適切な対応をとる対訳句が翻訳文に含まれていた．原因の一つとして，対訳句の抽出において対訳文 1 文につき複数パターン適合した場合に，その適合したパターンから考えられる全ての対訳句を出力するために，不適切な対応をとる対訳句が出力されたことが挙げられる．従来手法における対訳句の抽出方法を具体的に説明する．入力は，表 2.10 であると仮定する．

表 2.10: 入力例

対訳文	ナイロン は 原油 より 作ら れ る	Nylon is made from crude oil
単語レベル文パターン 1	*N1* は *N2 N3*	*N1 N3 N2*
単語レベル文パターン 2	*N1* は *N2 N3*	*N1 N2 N3*

変数 *N1*，*N2*，*N3* を 1 組と見なす．従来手法は，単語レベル文パターンから考えられる全ての対訳句を出力する．この場合，単語レベル文パターン 1 で出力される対訳句は全部で 40 組あり，句でに換算する 120 句となる．また，同様に単語レベル文パターン 2 においても 40 組対訳句が出力される．よって，表 2.10 の場合に従来手法を用いると，単語レベル文パターン 1 と 2 だけで対訳句は 240 句出力される．従来手法の具体例を図 2.11 示す．

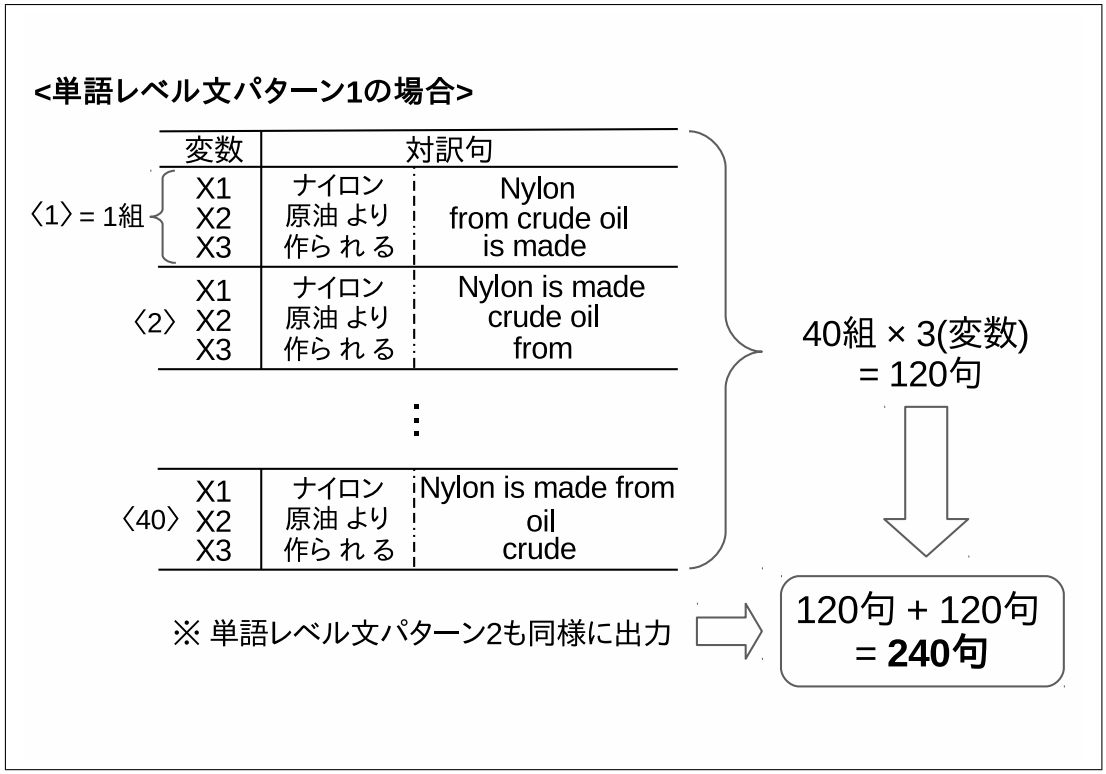

図 2.11: 従来手法の具体例

第3章　提案手法

3.1　提案手法の概要

江木らの従来手法は，単語辞書と対訳文パターンを統計的手法で自動的に作成し翻訳する手法を提案した．しかし，この手法では対訳句の抽出において単語レベル文パターンから全ての可能な対訳句を出力するために，不適切な対応をとる対訳句が多く出力されていた．

そこで本研究では，対訳句の変数全体の確率値を利用して最適な対訳句を出力する，PATTERN 法と SENTENCE 法の 2 手法を提案する．提案手法により，対訳句の出力数を抑制し不適切な対応をとる対訳句の数を減らすことで，対訳句の精度向上を目指す．また，出力した対訳句を用いて日英統計翻訳を行い，翻訳精度の調査を行う．

3.2　PATTERN 法

PATTERN 法は，その単語レベル文パターンごとに対数フレーズ確率の総和が最大の対訳句を出力する手法である．PATTERN 法の対訳句の抽出手順を以下に示す．

手順 1 対訳文と単語レベル文パターンを表 2.10 とし，パターンを照合する．

手順 2 適合した場合，単語レベル文パターンの変数部に対応する全ての組み合わせの対訳句を抽出する．

手順 3 GIZA++の単語確率を用いて，各組み合わせの中から最大となる単語確率を得る

手順 4 得られた単語確率を用いて，対数フレーズ確率を計算する．
ここまでは，従来手法と同様の手順．

手順 5 各単語レベル文パターンごとに，手順 5 で計算した対数フレーズ確率の総和の最大値をとる対訳句を 1 つずつ選出する．

手順 6 手順 6 で選出した対訳句を最終的に抽出される対訳句として出力する.

図 3.1 に，PATTERN 法の具体例を示す．

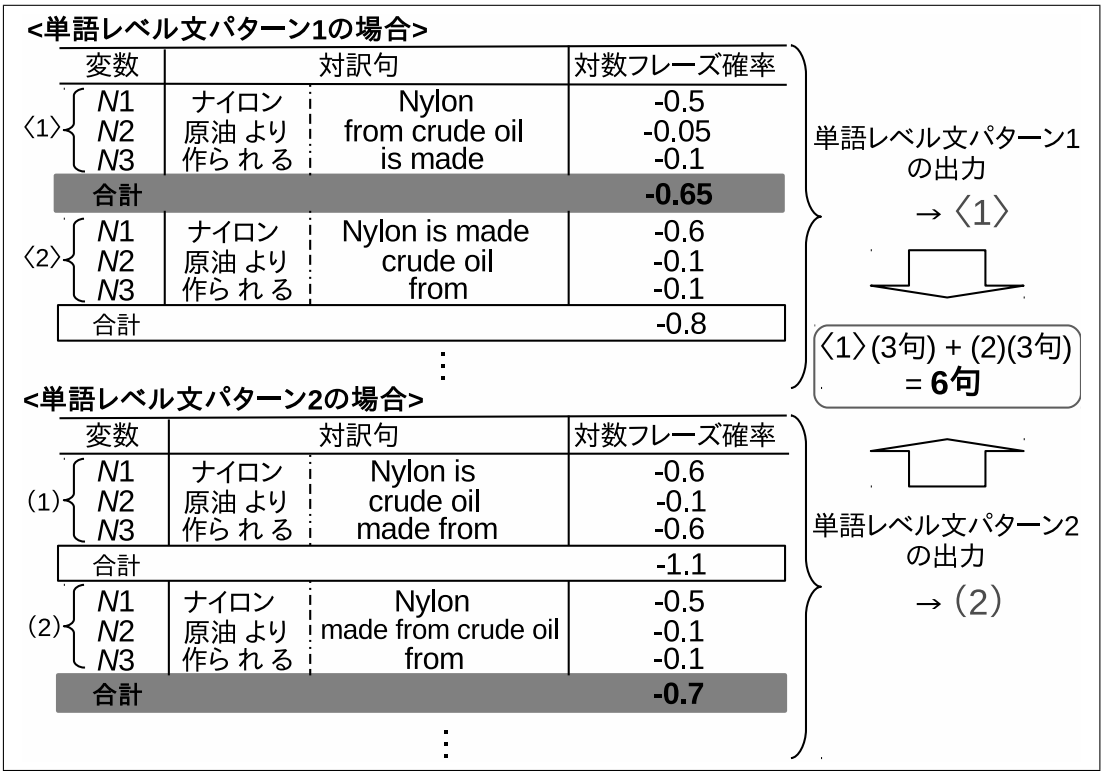

図 3.1: PATTERN 法の具体例

従来手法と同様，単語レベル文パターン 1 で対訳句の候補が 40 組存在する．PATTERN 法は，各組ごとに対数フレーズ確率の総和を計算し，その中から総和が最大となる 1 組を出力する．つまり、単語レベル文パターン 1 で出力される対訳句は，対数フレーズ確率の総和が-0.65 と一番高い ⟨1⟩ となる．同じ方法を，単語レベル文パターン 2 でも行い (2) を出力する．そして，各単語レベル文パターンごとに選ばれた対訳句の数を合わせ出力する．つまり，表 2.10 の場合に PATTERN 法を用いると，対訳句は単語レベル文パターン 1 と 2 を合わせた 6 句のみが出力される．

3.3 SENTENCE法

SENTENCE法は，対訳文1文につき全ての単語レベル文パターンの中から対数フレーズ確率の総和が最大の対訳句を出力する．つまり，対訳文に適合する全ての単語レベル文パターンにPATTERN法を適用し，その中から対数フレーズ確率の総和が最大となる対訳句を出力する．SENTENCE法の対訳句の抽出手順を以下に示す．

手順1 対訳文と単語レベル文パターンを表2.10とし，パターンを照合する．

手順2 適合した場合，単語レベル文パターンの変数部に対応する全ての組み合わせの対訳句を抽出する．

手順3 GIZA++の単語確率を用いて，各組み合わせの中から最大となる単語確率を得る

手順4 得られた単語確率を用いて，対数フレーズ確率を計算する．
ここまでは，従来手法と同様の手順．

手順5 全ての単語レベル文パターンの対訳句の中から，手順5で計算した対数フレーズ確率の総和が一番高い対訳句を1つだけ選出する．

手順6 手順6で選出した対訳句を最終的に抽出される対訳句として出力する．

図 3.2 に SENTENCE 法の具体例を示す.

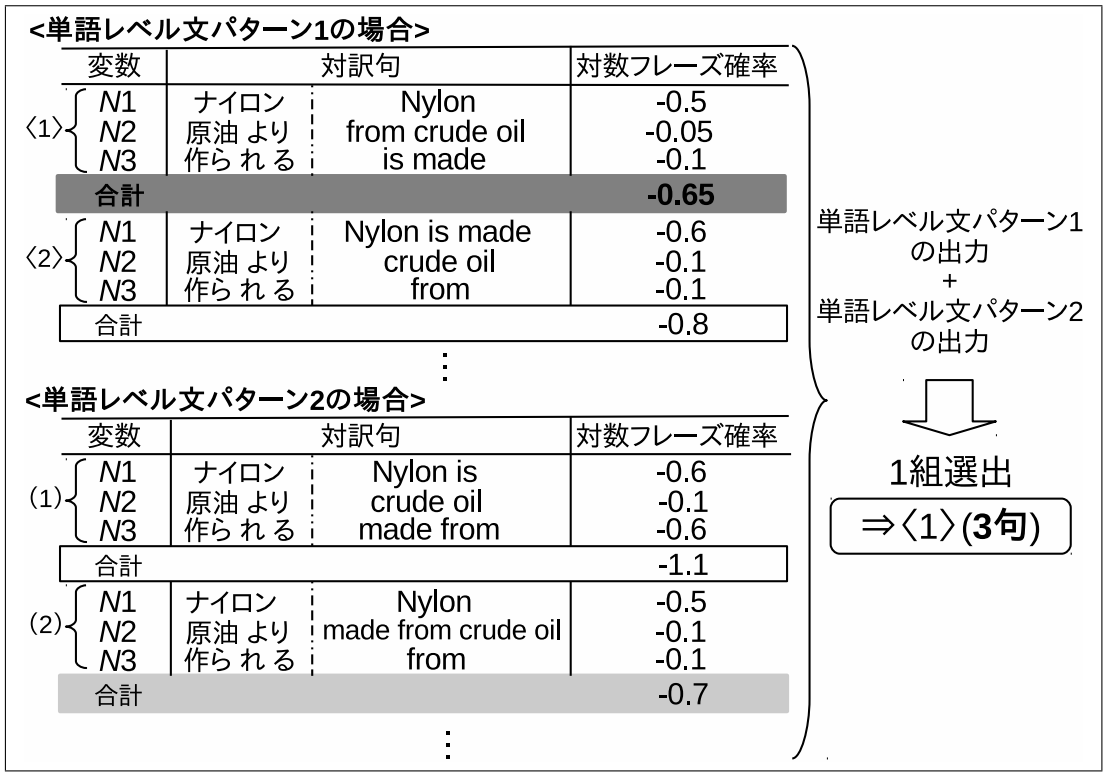

図 3.2: SENTENCE 法の具体例

単語レベル文パターン 1 において対訳句の候補は 40 組，同様に単語レベル文パターン 2 も対訳句の候補が 40 組存在する．まず，それらの各候補に PATTERN 法を適用させ対数フレーズ確率の総和を計算する．SENTENCE 法は，全ての単語レベル文パターンの対訳句の候補から最適な対訳句を 1 組出力する．つまり，全 80 組の対訳句の候補の中から対数フレーズ確率の総和が-0.65 と一番高い ⟨1⟩ が出力される．よって，表 2.10 の場合に SENTENCE 法を用いると，対訳句は 80 組の中から対数フレーズ確率の総和が一番高い 3 句だけが出力される．

第4章　実験

4.1　実験データ

対訳文および翻訳実験に用いるテスト文は，電子辞書から抽出した単文データを用いる [4]．なお，単文データは日本語文が単文であるのに対し，英語文は単文とは限らず，重文・複文が含まれる．対訳文および翻訳実験に用いるテスト文の例を表 4.1 に，使用するデータの内訳を表 4.2 に示す.

表 4.1: 対訳文および翻訳実験に用いるテスト文の例

日本語文	今朝 は 早く 目 が 覚め た 。
英語文	I woke up early this morning .
日本語文	見知らぬ 女性 が 私 に 微笑み かけ た 。
英語文	A woman who was a total stranger smiled at me .
日本語文	どんな スポーツ で も 君 に 負け ない 。
英語文	He will match you in any sport .

表 4.2: 実験データの内訳

対訳文	100,000 文
テスト文	200 文

4.2　評価方法

本研究は，対訳句の精度評価と翻訳文の精度評価のため，実験を 2 回行う．まず 1 回目の実験は，従来手法と PATTERN 法，SENTENCE 法を用いて対訳句を抽出し，対訳句の数の比較と精度を人手評価する．翻訳実験は行わない．2 回目の実験では，提案手法のみを用いて統計翻訳を行い，対訳句の精度とともに翻訳文の精度を人手評価する．翻訳文の精度評価には，対比較評価を用いる．また，実験に使用する対訳文は，1 回目に 1,000 文，2 回目に 100,000 文とする．

第5章　実験結果

5.1　対訳句の出力数 (1)

従来手法と両提案手法において，対訳文 1,000 文を用いて対訳句の抽出を行った．この結果より，従来手法と両提案手法を用いて，対訳句の出力数がどれ程抑制できているか調査する．各手法により抽出した対訳句の数を表 5.1 に示す．また，同一の単語を一語とする (異なり語数) 条件を適用した場合の対訳句の数も示す.

表 5.1: 従来手法と両提案手法の対訳句の出力数

	従来手法	PATTERN 法	SENTENCE 法
総出力数	90,858,412	25,888	3,488
異なり語数	1,003,589	13,131	2,783

表 5.1 の結果より，両提案手法の方が従来手法よりも対訳句の数が大きく減少していることが確認できた,

5.2　対訳句の精度評価 (1)

各手法より抽出した対訳句からランダムに 100 句取り出し，対訳句の精度を調査した．評価基準を以下に示す.

○：適切な対応をとる対訳句

△：部分的に対応している対訳句

×：不適切な対応をとる対訳句

5.2.1 総出力の精度評価

対訳句の総出力数からランダムに 100 句取り出し，人手精度評価を行った．人手評価結果を表 5.2 に示す.

表 5.2: 総出力数における対訳句の人手評価

	従来手法	PATTERN 法	SENTENCE 法
○	5	36	50
△	21	14	20
×	74	50	30

表 5.2 より，両提案手法とも従来手法より対訳句の精度が良く、特に SENTENCE 法は半数以上が適切な対応が取れてた.

また，各手法の判断基準における対訳句の例を示す．従来手法の例を表 5.3 に，PATTERN 法の例を表 5.4 に，SENTENCE 法の例を表 5.5 に示す.

表 5.3: 従来手法の対訳句の例

○	△	×
町 の 人:townspeople	自分 の 成果:comment	テスト:higt degree
大きな リンゴ:big apple	事件 を 法廷:case to court	父:the new job
た:was	時間 半:a half	くらん だ:by

表 5.4: PATTERN 法の対訳句の例

○	△	×
あの 店:That store	賛美 を:his praise	難しい 問題:solved
この 試験:This test	裕福 に:enriched the country	に は:all
微風:breeze	私 に は:my understanding	準備 に:We

表 5.5: SENTENCE 法の対訳句の例

○	△	×
伴う:accompanied	この:I'm saving this	強化 する:will
深い 意味:deep meaning	倍 の 量:quantity	評価:work
道路 工事:road construction	話:talking about	逆立っ:end

5.2.2 異なり語数の精度評価

対訳句の異なり語数からランダムに 100 句取り出し，人手精度評価を行った．人手評価結果を表 5.6 に示す.

表 5.6: 異なり語数における対訳句の人手評価

	従来手法	PATTERN 法	SENTENCE 法
○	2	27	45
△	19	14	22
×	79	59	33

表 5.6 より表 5.2 と同様，両提案手法とも従来手法より対訳句の精度が良い．しかし、全体的に対応が取れている数が減ってた.

また，各手法の判断基準における対訳句の例を示す．従来手法の例を表 5.7 に，PATTERN 法の例を表 5.8 に，SENTENCE 法の例を表 5.9 に示す.

表 5.7: 従来手法の対訳句の例

○	△	×
考え:idea	試合 で:in an important game	の 雰囲気 が:New
メーカー:manufacturer	その:The hill	を お願い:king to
	いろいろ の:various forms	黒 塗り の:front of

表 5.8: PATTERN 法の対訳句の例

○	△	×
若者 たち:young people	雨 が:of rain in June	泥棒:police
彼 の 父:His father	苦しい 戦い:struggle	ない:weapons
外交 官:diplomat	青:blue background	彼女:Indira

表 5.9: SENTENCE 法の対訳句の例

○	△	×
稲妻:Lightning	各 時代 の:of the ages	月面:Apollo
白い スカーフ:white scafe	何 年 も:for years	評価:work
町中:whole town	改正 ラッシュ:revised	雨:pelting

5.3 対訳句の数 (2)

両提案手法において，対訳文 100,000 文を用いて対訳句の抽出を行った．この結果より，PATTERN 法と SENTENCE 法のどちらがより対訳句の出力数を抑制できるか調査する．両提案手法により抽出した対訳句の数を表 5.10 に示す．また，同一の単語を一語とする (異なり語数) 条件を適用した場合の対訳句の数も示す.

表 5.10: 両提案手法の対訳句の出力数

	PATTERN 法	SENTENCE 法
総出力数	2,607,687	353,468
異なり語数	1,013,708	214,259

表 5.10 の結果より，SENTENCE 法の方が大幅に対訳句の数を減少できていることが確認できた,

5.4 対訳句の精度評価 (2)

両提案手法より抽出した対訳句からランダムに 100 句取り出し，対訳句の精度を調査した．評価基準は，4.2 節と同様である．

5.4.1 総出力の精度評価

対訳句の総出力数からランダムに 100 句取り出し，人手精度評価を行った．人手評価結果を表 5.11 に示す．

表 5.11: 総出力数における対訳句の人手評価

	PATTERN 法	SENTENCE 法
○	29	37
△	17	13
×	54	50

表 5.11 より，SENTENCE 法の方が対訳句の精度が良いことが確認できたが，両手案手法とも不適切な対応をとる対訳句の数が半数を占めている．

また，両提案手法の判断基準における対訳句の例を示す．PATTERN 法の例を表 5.12 に，SENTENCE 法の例を表 5.13 に示す．

表 5.12: PATTERN 法の対訳句の例

○	△	×
慈善:The charity	の 提案:proposal had	ポケット : table
まとめ の 段階 : stage of conclusion	眺め た:The view	住ん で:in
批評 家 たち:The critics	動き出し:move on its own	の:reached

表 5.13: SENTENCE 法の対訳句の例

○	△	×
たわいない : nonsense	引き下げ:pulled	痛感 : democracy
湯 を わかす:boils water	彼 の 言葉:by his remarks	沈没を : The
彼 の 声:His voice	天王星 に:Uranus at long	心臓:My

5.4.2 異なり語数の精度評価

対訳句の異なり語数からランダムに 100 句取り出し，人手精度評価を行った．人手評価結果を表 5.14 に示す.

表 5.14: 異なり語数における対訳句の人手評価

	PATTERN 法	SENTENCE 法
○	28	21
△	26	31
×	46	48

表 5.14 より，表 5.11 とは異なり PATTERN 法の方が若干ではあるが対訳句の精度が良かった．また，不適切な対応をとる対訳句の数は，両提案手法とも減少していた.

また，両提案手法の判断基準における対訳句の例を示す．PATTERN 法の例を表 5.15 に，SENTENCE 法の例を表 5.16 に示す.

表 5.15: PATTERN 法の対訳句の例

○	△	×
この 川:this river	公明党:New Koumeito	ところ:I'm
隕石:meteor	うまく 答え:neat answer	に 負え:are
我々 の 希望:our hopes	廊下 で 本:downstairs	君 の:in this

表 5.16: SENTENCE 法の対訳句の例

○	△	×
まだ 不十分:still unsatisfactory	花 の 蜜:flower	で 歩い た:the
広い 野原:wide field	彼 の 最近 の:recent	ヨーロッパ:supremacy
日本 の 国花:The national flower of Japan	彼 は 私:call on me	件:Let

5.5 翻訳文の精度評価

両提案手法より抽出した対訳句と対訳文 100,000 文を用いて統計翻訳を行い，翻訳文の精度を調査した．人手評価には，PATTERN 法と SENTENCE 法の翻訳文の対比較評価を行う．PATTERN 法と SENTENCE 法との対比較評価結果を表 5.17 に示す.

表 5.17: PATTERN 法と SENTENCE 法の対比較評価結果

PATTERN 法○	SENTENCE 法○	差なし	同一出力
14	19	57	10

PATTERN 法○の例を表 5.19 に，SENTENCE 法○の例を表 6.1 に示す.

表 5.18: PATTERN 法と SENTENCE 法の対比較評価結果:PATTERN 法○の例

入力文	私 は 正反対 の こと を 考え て い た 。
参照文	I thought quite the opposite .
PATTERN 法	I thought quite the opposite .
SENTENCE 法	I thought 正反対 never .
入力文	わたし は その 事故 の 犠牲 者 の 冥福 を 祈っ た 。
参照文	I prayed for the peaceful repose of the victims of the accident .
PATTERN 法	I prayed 冥福 of accident victims .
SENTENCE 法	I prayed for 冥福 victims of the accident .
入力文	いなか の 空気 は 私 たち を 元気づけ て くれ た 。
参照文	The country air refreshed us .
PATTERN 法	The country air was a good tonic to us .
SENTENCE 法	We air country has a good tonic to us .

表 5.19: PATTERN 法と SENTENCE 法の対比較評価結果:SENTENCE 法○の例

入力文	その 報道 は かなり の 不安 を 生じ させ た 。
参照文	The report aroused considerable anxiety .
PATTERN 法	The news caused a considerable number of financial unrest .
SENTENCE 法	The news caused a considerable apprehension .
入力文	彼 は 人々 から 非難 を 浴び た 。
参照文	He was subjected to criticism by the people .
PATTERN 法	He was showered with denounced by the people .
SENTENCE 法	He received criticism from the people .
入力文	未熟 な 技術 者 は 雇え ない 。
参照文	We don't employ inexperienced engineers .
PATTERN 法	He immaturity 雇え engineers .
SENTENCE 法	We don't employ inexperienced engineers .

表 5.17 の結果より，PATTERN 法と SENTENCE 方を比較して，SENTENCE 法が優れていることがわかる．よって，本研究で用いた入力文における SENTENCE 法の有効性が確認された.

第6章　考察

6.1　提案手法の有効性

対訳句を全て出力した場合の数と，同一の単語を一語とする (異なり語数) という条件を適用した場合に，従来手法と PATTERN 法，SENTENCE 法でどれ程対訳句の数を抑制できているか，またチュ出した対訳句を用いて精度調査を行った．表 5.1 より，両提案手法とも従来手法に比べ大幅に数を削減できていることが確認でき，表 5.2 と表 5.6 より，従来手法に比べ精度が高いことがわかった．これより，提案手法の有効性が確認できる．しかし，両提案手法ともに対応が取れている対訳句の数が半数以下という結果であったことより，対訳句の抽出において誤ったパターン適合しているか，または，対数フレーズ確率の総和が低い可能性がある．

6.2　追加実験

本研究では，従来手法と提案手法における翻訳実験が出来なかった．そこで，従来手法の代わりに Moses[6] を用いて翻訳実験を行った．対訳文 100,000 文，テスト文 100 文使用し，翻訳文に対して対比較評価を行う．

6.2.1　Moses と PATTERN 法の対比較評価結果

Moses と PATTERN 法の対比較評価結果を表 6.1 に示す．

表 6.1: Moses と PATTERN 法の対比較評価結果

Moses ○	PATTERN 法○	差なし	同一出力
16	22	62	0

Moses ○の例を表 6.2 に，PATTERN 法○の例を表 6.3 に示す.

表 6.2: Moses と PATTERN 法の対比較評価結果:Moses 法○の例

入力文	世間 の 同情 が 彼 に 集まっ た 。
参照文	Public sympathies were centered on him .
Moses	The Public sympathy gathered around him .
PATTERN 法	Public sympathy was centered onhim .
入力文	その 報道 は かなり の 不安 を 生じ させ た 。
参照文	The report aroused considerable anxiety .
Moses	The English from Latin 借用 a lot of words .
PATTERN 法	My English borrowed lot of people from Latin .
入力文	その 部屋 は 食堂 と 居間 を 兼ね て いる 。
参照文	That room serves as both dining room and living room .
Moses	The room with the dining room doubles as a living room .
PATTERN 法	The room serves both as a study and in the living room .

表 6.3: Moses と PATTERN 法の対比較評価結果:PATTERN 法○の例

入力文	バス が えんこ し て しまっ た 。
参照文	The bus broke downon the road .
Moses	The bus was えんこ .
PATTERN 法	The bus broke down .
入力文	自動車 が 動か ない 。
参照文	The car will not start .
Moses	The car is quiet .
PATTERN 法	The car is not working .
入力文	彼 の 沈黙 を 拒絶 と 解釈 し た 。
参照文	The soldiers died in a heap .
Moses	His silence was interpreted a flat refusal .
PATTERN 法	His silence was interpreted as a flat refusal .

表 6.1 の結果より，Moses と PATTERN 法を比較して，PATTERN 法が若干ではあるが翻訳精度が高いことが確認できた．しかし，ほぼ差がないと言っていい．

6.2.2 Moses と SENTENCE 法の対比較評価結果

Moses と SENTENCE 法の対比較評価結果を表 6.4 に示す.

表 6.4: Moses と SENTENCE 法の対比較評価結果

Moses ○	SENTENCE 法○	差なし	同一出力
14	27	58	1

Moses ○の例を表 6.5 に，SENTENCE 法○の例を表 6.6 に示す.

表 6.5: Moses と SENTENCE 法の対比較評価結果:Moses 法○の例

入力文	地動説 は 、 コペルニクス によって 証明 さ れ た 。
参照文	The heliocentric theory was proven by Copernicus .
Moses	The 地動説 , was proved by コペルニクス .
SENTENCE 法	地動説 by コペルニクス prove it .
入力文	ボール が 彼 の 頭上 に 落ち た 。
参照文	A ball fell on his head .
Moses	The ball fell over the heads of him .
SENTENCE 法	The ball fell in the head .
入力文	世間 の 同情 が 彼 に 集まっ た 。
参照文	Public sympathies were centered on him .
Moses	The Public sympathy gathered around him .
SENTENCE 法	Public sympathy was centered onhim .

表 6.6: Moses と SENTENCE 法の対比較評価結果:SENTENCE 法○の例

入力文	その 試合 は テレビ で 見 まし た 。
参照文	I watched the game on television .
Moses	The game I saw on television .
SENTENCE 法	I watched the match on television .
入力文	その 責任 が 私 の 肩 に 重く のしかかっ て いる 。
参照文	The responsibility rests heavily on my shoulders .
Moses	The responsibility fell upon my shoulder .
SENTENCE 法	The burden of his responsibility weighs heavily on my shoulders .
入力文	仕事 が 山積 し て いる 。
参照文	I have too much work to do .
Moses	The business to attend to .
SENTENCE 法	I have a lot of business to attend to .

表 6.4 の結果より，Moses と SENTENCE 法を比較して，SENTENCE 法が翻訳精度が高いことが確認できた．ある程度差があることより，本研究で用いた入力文における SENTENCE 法の有効性が確認できる．

第7章　おわりに

従来手法において，対訳句の抽出では単語レベル文パターンから全ての可能な対訳句を出力するために，不適切な対応をとる対訳句が多く出力されていた．そこで，対訳句の変数全体の確率値を利用して最適な対訳句を出力する，PATTERN 法と SENTENCE 法の 2 手法を提案した．提案手法により，対訳句の出力数を抑制し不適切な対応をとる対訳句の数を減らすことで，対訳句の精度向上を目指し，また翻訳精度の調査を行った．従来手法と両提案手法の対訳句の精度評価では，対訳句の数の大幅さ削減ができており，精度調査についても従来手法よりも適切な対応がとれている対訳句が多かった．これより，対訳句の精度の向上が確認できる．

また，PATTERN 法と SENTENCE 法より翻訳実験を行った．対訳句の数については，SENTENCE 法が大きく数を削減した．しかし，対訳句の精度を調査したところ，両提案手法に差があまりなかった．翻訳文の精度評価においても，差がなかった．このことより，大幅に対訳句の数を抑制して対訳句の精度，翻訳の精度に影響がないことがわかった．今後は，なぜこのような結果になったのか原因究明を行い，さらなる精度向上を目指したい．

謝辞

最後に，一年間に渡り，本研究のご指導をいただきました鳥取大学工学部知能情報工学科計算機工学 C 講座研究室の村上仁一准教授，村田真樹教授に深く感謝すると共に，厚く御礼申し上げます．そして，日常の議論を通じて多くの知識や示唆を頂いた同研究室の皆様に深謝いたします．また，参考にさせていただいた著書の著者の方々に，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] 石上真理子, 水田理夫, 徳久雅人, 村上仁一, 池原悟: “関数・符号付き文型パターンを用いた機械翻訳の試作と評価”, 言語処理学会第 13 回年次大会予稿集, pp.67-70, 1997.

[2] 江木 孝史: “句に基づく文パターンを用いた英日翻訳”, 修士論文, 2014.

[3] Franz Josef Och,Hermann Ney: “A Systematic Compar-ison of Various Statistical Alignment Models”, Computa-tional Linguistics, 29(1), pp.19-51, 2003.

[4] 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.

[5] 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.

[6] 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.