

概要

現在、機械翻訳の分野において、対訳文対から自動的に翻訳規則を生成し、翻訳を行う統計翻訳が注目されている。統計翻訳はプログラムを用いて、大量の対訳文対から対訳フレーズ対を自動的に生成する。よって、その対訳フレーズ対のカバー率は高いが、信頼性が低いという問題がある。

対訳フレーズ対の信頼性が低いという問題に対し、鏡味らは人手で作成した対訳フレーズ対を統計翻訳に導入し、翻訳性能の向上を報告した [1]。この先行研究は人手で作成した対訳フレーズ対を導入するために、自動作成した対訳フレーズ対の翻訳確率を利用する。対訳フレーズ対の翻訳確率を利用するために、自動作成した対訳フレーズ対と人手で作成した対訳フレーズ対のマッチングを行う。そして、完全に一致した場合のみ、人手で作成した対訳フレーズ対に対して翻訳確率を付与する。よって、導入可能なフレーズ数が少ないという問題がある。

そこで、本研究では、導入可能な対訳フレーズ対が少ないという問題に対処するために、完全に一致した対訳フレーズ対だけでなく、部分的に一致した対訳フレーズ対を統計翻訳に導入し、翻訳性能の向上を試みる。

その結果、従来手法 (moses) と比較すると、提案手法の翻訳精度は向上した。しかし、先行研究と比較すると、提案手法の翻訳精度は低いという結果であった。よって、提案手法の有効性は低いという結果であった。

目次

第1章	はじめに	1
第2章	統計翻訳システム	3
2.1	基本概念	3
2.2	翻訳モデル	4
2.2.1	IBM 翻訳モデル	5
2.2.2	GIZA++	11
2.2.3	フレーズテーブルの作成法	12
2.2.4	フレーズ対応の抽出	15
2.2.5	抽出したフレーズ対応の確率付け	16
2.3	言語モデル	16
2.3.1	リオーダーリングモデル	19
2.4	デコーダ	20
2.4.1	ビームサーチ法	21
2.4.2	マルチスタック法	21
2.4.3	moses のパラメータ	22
2.4.4	パラメータチューニング	22
第3章	それぞれの手法の説明	24
3.1	提案手法	24
3.1.1	英辞郎	24
3.1.2	部分的に一致した翻訳対の翻訳確率	26
3.1.3	部分的に一致した翻訳対の確率のペナルティ	26
3.1.4	翻訳確率の付与手順	27
3.2	先行研究	28
3.3	従来手法	29
3.4	先行研究と提案手法を組み合わせた手法	30

第4章	実験データ・実験環境	31
4.1	実験データ	31
4.1.1	対訳学習文	31
4.1.2	テスト文	32
4.1.3	確率を付与した翻訳対	32
4.2	実験環境	32
4.2.1	フレーズテーブル	32
4.2.2	N -gram モデルの学習	32
4.2.3	デコーダのパラメータ	32
第5章	評価方法	33
5.1	自動評価	33
5.1.1	BLEU	33
5.1.2	NIST	34
5.1.3	METEOR	35
5.2	人手評価	35
第6章	従来手法と提案手法の比較	36
6.1	フレーズテーブルの総数の比較	36
6.2	翻訳精度の自動評価	36
6.3	翻訳精度の人手評価	36
6.3.1	翻訳結果の例	37
第7章	先行研究と提案手法の比較	38
7.1	フレーズテーブルの総数の比較	38
7.2	翻訳精度の自動評価	38
7.3	先行研究と提案手法の人手評価	39
7.3.1	翻訳結果の例	39
第8章	先行研究と追加実験の比較	41
8.1	フレーズテーブルの総数の比較	41
8.2	翻訳精度の自動評価	41
8.3	先行研究と追加実験の人手評価の結果	42
8.3.1	翻訳結果の例	42

第 9 章 考察	43
9.1 翻訳精度が向上した文の解析	43
9.1.1 適切なフレーズ対の増加	44
9.1.2 未知語の減少	54
9.2 翻訳精度が低下した文の解析	56
9.2.1 不適切なフレーズ対の選択	57
9.2.2 未知語の影響	59
9.2.3 マッチングの対象の問題	61
9.3 翻訳候補文に使用された翻訳対の違い	62
9.4 ペナルティの影響	62
9.5 未知語の影響	62
第 10 章 おわりに	64
第 11 章 付録	68
11.1 実験データ	68
11.1.1 学習データ	68
11.1.2 テストデータ	68
11.2 フレーズテーブルの総数の比較	68
11.3 翻訳精度の自動評価	68
11.4 使用された翻訳対の違い	69

目 次

2.1	日英統計翻訳の枠組	4
2.2	フランス語英語翻訳における、アライメントの例	5
2.3	日英、英日方向の最尤な単語 alignment の例	12
2.4	最尤な単語 alignment の例 (intersection)	13
2.5	最尤な単語 alignment の例 (union)	13
2.6	最尤な単語 alignment の例 (grow-diag)	14
2.7	最尤な単語 alignment の例 (grow-diag-final)	14
2.8	フレーズ対応の抽出例	15
2.9	reordering における swap の例	19
2.10	入力文「言語はコミュニケーションの道具である」の探索例	20
2.11	マルチスタック・ビームサーチ法の適用例 (入力文「言語はコミュニケーションの道具である」)	22
3.1	提案手法の翻訳確率の付与手順	27
3.2	先行研究の翻訳確率の付与手順	28
3.3	従来手法の手順	29
3.4	追加実験の翻訳確率の付与手順	30

表 目 次

2.1	フレーズテーブルの例	4
2.2	作成されたフレーズテーブルの例 (grow-diag-final)	15
2.3	図 2.8 から抽出した“矛盾しない”フレーズ対応の例	16
2.4	言語モデルの例	17
2.5	言語モデル作成に用いた英文の例	18
2.6	スムージングパラメータ別の言語モデルの例	18
3.1	英辞郎の例	25
3.2	翻訳対の例	25
3.3	部分的に一致した翻訳対の例	26
3.4	翻訳確率を付与した翻訳対の例	27
4.1	単文の対訳学習文の例	31
6.1	フレーズテーブルの総数と追加した翻訳対の比較	36
6.2	単文における翻訳精度の評価	36
6.3	従来手法と提案手法の対比較評価結果	37
6.4	提案手法が優れていると判断した例	37
7.1	フレーズテーブルの総数と追加した翻訳対の比較	38
7.2	単文における翻訳精度の評価	38
7.3	先行研究と提案手法の対比較評価結果	39
7.4	提案手法が優れていると判断した例	39
7.5	提案手法が劣っていると判断した例	40
8.1	フレーズテーブルの総数と追加した翻訳対の比較	41
8.2	単文における翻訳精度の評価	41
8.3	先行研究と追加実験の対比較評価結果	42

8.4	単文において提案手法が優れていると判断した例	42
9.1	翻訳精度向上の理由	43
9.2	提案手法が優れていると判断した例 1(従来手法との比較)	44
9.3	提案手法が優れていると判断した例 2(従来手法との比較)	45
9.4	提案手法が優れていると判断した例 3(従来手法との比較)	46
9.5	提案手法が優れていると判断した例 4(従来手法との比較)	47
9.6	提案手法が優れていると判断した例 5(先行研究との比較)	48
9.7	提案手法が優れていると判断した例 6(先行研究との比較)	49
9.8	提案手法が優れていると判断した例 7(先行研究との比較)	50
9.9	追加実験が優れていると判断した例 8(追加実験と先行研究の比較)	51
9.10	追加実験が優れていると判断した例 9(追加実験と先行研究の比較)	52
9.11	追加実験が優れていると判断した例 10(追加実験と先行研究の比較)	53
9.12	提案手法が優れていると判断した例 11(従来手法との比較)	54
9.13	追加実験が優れていると判断した例 12(追加実験と先行研究の比較)	55
9.14	翻訳精度低下の理由	56
9.15	提案手法が劣っていると判断した例 1(先行研究との比較)	57
9.16	提案手法が劣っていると判断した例 2(先行研究との比較)	58
9.17	提案手法が劣っていると判断した例 3(先行研究との比較)	59
9.18	提案手法が劣っていると判断した例 4(先行研究との比較)	60
9.19	提案手法が劣っていると判断した例 5(先行研究との比較)	61
9.20	ペナルティによる自動評価の変化	62
9.21	未知語の総数	63
11.1	フレーズテーブルの総数と追加した翻訳対の比較	68
11.2	重文複文における翻訳精度の評価	69
11.3	テスト文で使用された翻訳対の比較	69

第1章 はじめに

現在、グローバル世界で活動するに従い、言語を翻訳する仕事が急激に増加している。一方で、人手による翻訳は時間がかかり、コストが高く、言語によっては翻訳者の数も限られる。そのため、時間やコストを削減するために機械翻訳の技術の必要性が高まっている。従来の機械翻訳方式ではルールベース法が用いられてきた。ルールベース法は1つの言語間の翻訳規則を長い時間をかけて構築する必要がある。また、言語により翻訳規則が異なるため、多言語への拡張が難しい。そこで、現在、機械翻訳の分野において、対訳データから自動的に翻訳規則を生成し、翻訳を行う統計翻訳が注目されている。

統計翻訳は、獲得した翻訳規則を言語モデルと翻訳モデルで管理する。言語モデルは、単語の列や文字の列が起こる確率を与えるモデルである。翻訳モデルは、英語の単語の列から単語の列へ確率的に翻訳を行うためのモデルである。翻訳モデルには、主に単語に基づくモデルと句に基づくモデルがある。単語に基づくモデルは、単語の対応作成時に、対応がない単語にはNULLを対応させる。その結果、単語に基づくモデルにおける翻訳精度の低下を起こす。一方、句に基づくモデルは、単語に基づくモデルより、訳語の選択能力や局所的な語の並べ替え能力の高い、句に基づく翻訳モデル [2] が現在の主流になっている。

句に基づく翻訳モデルは、プログラムで自動作成されるフレーズテーブルと呼ばれる表で管理されている。句に基づくモデルは、句の対応をとるため、単語に基づくモデルで使用されていたNULLは使用しない。フレーズテーブルはプログラムにより自動作成されるため、カバー率は高いが、信頼性は低い。

対訳フレーズ対の信頼性が低いという問題に対し、鏡味らは人手で作成した対訳フレーズ対を統計翻訳に導入し、翻訳性能の向上を報告した [1]。この先行研究は人手で作成した対訳フレーズ対を導入するために、自動作成した対訳フレーズ対の翻訳確率を利用する。そのために、自動作成した対訳フレーズ対と人手で作成した対訳フレーズ対のマッチングを行う。そして完全に一致した場合のみ、人手で作成した対訳フレーズ対に対して翻訳確率を付与する。よって、導入可能な対訳フレーズ対の数が少ないという問題がある。

そこで本研究では、導入可能な対訳フレーズ対が少ないという問題に対処するために、

完全に一致した対訳フレーズ対だけでなく、部分的に一致した対訳フレーズ対を統計翻訳に導入し、翻訳性能の向上を試みる。

その結果、従来手法と比較すると、提案手法の翻訳精度は向上した。しかし、先行研究と比較すると、提案手法の翻訳精度は低いという結果であった。よって、提案手法の有効性は低いという結果であった。

2章において、統計翻訳システムの各部分の説明を行い、統計翻訳システムの概要を示す。3章において、本研究における、提案するシステムと先行研究のシステム、従来手法の説明を行い、それぞれのシステムの違いを示す。4章において、実験データと実験環境の説明を行い、パラメータなどの設定を示す。5章において、システムの出力を評価する手法の説明を行い、評価方法の違いを示す。6章、7章において、提案手法と従来手法、先行研究の比較を行い、その結果を示す。8章において、先行研究と提案手法を組み合わせたシステムと先行研究の比較を行い、その結果を示す。9章で考察を示し、最後に10章で結論を述べ、まとめる。

第2章 統計翻訳システム

本章では、統計翻訳システムを成す、翻訳モデルと言語モデルとデコーダの説明を行い、統計翻訳システム全体の概要を示す。日英統計翻訳は、日本語学習文と英語学習文から成る、対訳学習文から自動的に翻訳規則を生成し、その翻訳規則に基づき、入力文に対して最尤な出力文を探索するシステムである。

2.1 基本概念

日英統計翻訳は、日本語文 j が与えられたとき、全ての組合せの中から確率が最大となる英語文 \hat{e} を探索することにより翻訳を行う。

$$\hat{e} = \operatorname{argmax}_e P(e | j) \quad (2.1)$$

$$= \operatorname{argmax}_e \frac{P(j | e)P(e)}{P(j)} \quad (2.2)$$

$$\approx \operatorname{argmax}_e P(j | e)P(e) \quad (2.3)$$

$P(j | e)$ は翻訳モデル、 $P(e)$ は言語モデルである。2.1 式はベイズの法則¹より、2.2 式に展開される。2.2 式の分母は e によつては変わらないため、大小の比較は結局 2.3 式に近似できる。

これは、2 言語間のモデル $P(e | j)$ を正確に推定することが困難なためである。そのため、一般に、高い信頼性を持つことが知られている言語モデルと翻訳モデルを併用し、翻訳精度を高めている。日英統計翻訳の枠組を図 2.1 に示す。

図 2.1 で示すように、翻訳モデルは日本語学習文と英語学習文が集まった、対訳学習文から学習して作成する。また、言語モデルは、出力文の言語である英語学習文から学習して作成する。翻訳モデルと言語モデルを用いて、 \hat{e} を探索する翻訳システムが図中のデコーダである。

¹ベイズの法則: $P(B | A) = \frac{P(A|B)P(B)}{P(A)}$

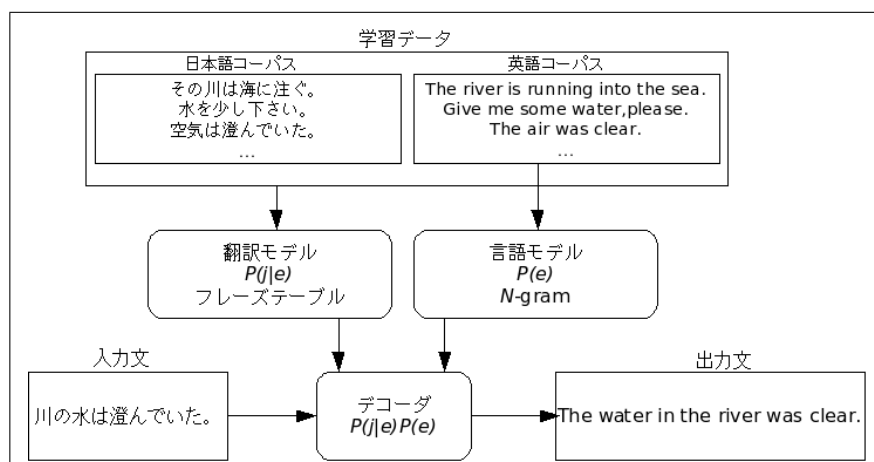


図 2.1: 日英統計翻訳の枠組

2.2 翻訳モデル

翻訳モデルは、英語の単語の列から単語の列へ確率的に翻訳を行うためのモデルである。翻訳モデルには、主に単語に基づくモデルと句に基づくモデルがある。現在は、訳語の選択能力や局所的な語の並べ替え能力の高い、句に基づく翻訳モデルが現在の主流になっている。句に基づく翻訳モデルは表 2.1 のようなフレーズテーブルと呼ばれる表で管理されている。

表 2.1: フレーズテーブルの例

あらし		The storm		0.025	0.122856	0.0555556	0.105324
きのう		Yesterday	,	0.2	0.235839	0.0131579	0.00783859
それぞれ		each		0.0714286	0.0183908	0.125	0.0418848
厳しい寒さ		intense cold		1	0.00103928	1	0.00334861
玄関の前		front of the entrance		1	0.015335	0.5	0.0018312

左から、日本語フレーズ、英語フレーズ、フレーズの日英翻訳確率 $P(j|e)$ 、単語の日英翻訳確率の積、フレーズの英日翻訳確率 $P(e|j)$ 、単語の英日翻訳確率の積である。なお、本研究では、日本語フレーズ、英語フレーズ、各種確率をまとめて、フレーズ対と呼ぶ。

2.2.1 IBM 翻訳モデル

統計翻訳の手法として、代表的なものに IBM の仏英翻訳モデル [4] がある。IBM 翻訳モデルは、Model1 から Model5 の 5 つのモデルから成り、順に複雑な計算となる。本研究では日本語から英語への翻訳を行うが、IBM 翻訳モデルではフランス語から英語への翻訳を想定しており、2.2.1 節では便宜上、原言語をフランス語 f 、目的言語を英語 e として、説明を行う。IBM モデルでは、フランス語文 f と英語文 e の翻訳モデル $P(f | e)$ を計算するために、アライメント a と呼ばれる概念を導入し、2.4 式を考える。

$$P(j | e) = \sum_a P(j, a | e) \quad (2.4)$$

アライメント a とは、ある仏単語 f と英単語 e の対応関係を示す。IBM 翻訳モデルでは英単語は $1:n$ の対応を持ち、仏単語は 1 つの英単語のみと対応すると仮定する。また、英語文頭に特殊文字 e_0 があるとし、仏単語の対応として適切な英単語がなかった場合はそこに対応づける。フランス語英語翻訳における、アライメントの例 2.2 を図に示す。

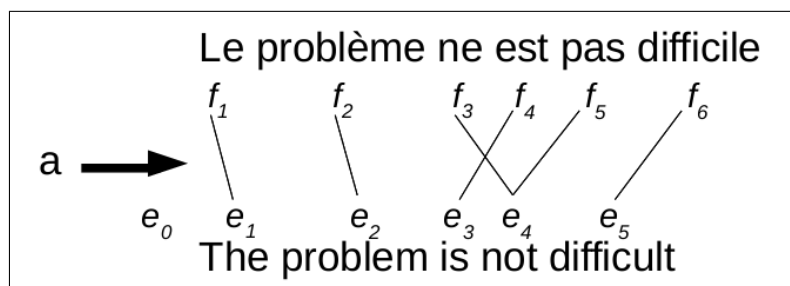


図 2.2: フランス語英語翻訳における、アライメントの例

翻訳モデルのパラメータは EM アルゴリズムを用いて推定する。一般に初期値としては、より簡単なモデルから得られた推定値を与えることで、より信頼性の高い値を得ることができる。次に各モデルについての詳しい説明を行う。

Model1

2.4 式は 2.5 式のように分解することができる。 m は仏語文の長さ、 a_1^{j-1} は仏語文の単語位置 1 から $j-1$ までのアライメントとし、 f_1^{j-1} は仏語文の位置 1 から $j-1$ までを表すとする。

$$P(f, a | e) = P(m | e) \prod_{j=1}^m P(a_j | a_1^{j-1}, f_1^{j-1}, m, e) P(f_j | a_1^{j-1}, j_1^{j-1}, m, e) \quad (2.5)$$

このままでは複雑すぎるため，計算が困難である．そこで，model1 では以下の仮定を置き，簡略化する．

- 仏語文の長さの確率は m ， e に依存しないとし，定数で置く．

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

- 単語の対応付けの確率は全て一定とし，英語文の長さ l にのみ依存する．

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

- 仏単語は前後の単語の出現には依存せず，自身に対応付けられる英単語のみに依存する．

$$P(f_j | a_1^{j-1}, j_1^{j-1}, m, e) = t(f_j | e_{a_j})$$

以上の仮定を用い，2.5 式は 2.6 式のように簡略化される．また，model1 において，確率の積の総和を総和の積に置き換えることが可能であり，2.8 式のように置くことができる．

$$P(f, a | e) = \frac{\epsilon}{(l+1)^m} \prod_{j=1}^m t(f_j | e_{a_j}) \quad (2.6)$$

$$P(f | e) = \frac{\epsilon}{(l+1)^m} \sum_{a_1}^l \cdots \sum_{a_m}^l \prod_{j=1}^m t(f_j | e_{a_j}) \quad (2.7)$$

$$= \frac{\epsilon}{(l+1)^m} \prod_{j=1}^m \sum_{i=0}^l t(f_j | e_{a_j}) \quad (2.8)$$

ラグランジェの未定計数法を用いて，制約 $\sum_f t(f | e) = 1$ のもとで $P(f | e)$ を最大化する問題を解くと，学習に用いた対訳文において， e と f が対応付けられる回数の期待値を 2.9 式のように求めることができる．

$$C(f | e; f, e) = \frac{t(f | e)}{f(f | e_0) + \cdots + f(f | e_l)} \sum_{j=1}^m \delta(f, f_j) \sum_{i=0}^m \delta e, e_i \quad (2.9)$$

$\delta(f, f_j)$ は仏文 f 中に仏単語 f が出現する回数, δ_{e, e_i} は英文 e 中に英単語 e が出現する回数を表す. このことより, 次のような EM アルゴリズムによって確率を推定することができる.

1. $t(f | e)$ に適用な初期値を設定する.
2. 対訳ペアの集合 $(f^{(s)}, e^{(s)}), 1 \leq s \leq S$ において, e と f が対応付けられる回数の期待値を 2.9 式によって求める.
3. e 中に 1 回以上出現した e に対し, $t(f | e)$ を 2.10 式で再推定する.

$$t(f | e) = \frac{\sum_{s=1}^S c(f | e; f^s, e^s)}{\sum_f \sum_{s=1}^S c(f | e; f^s, e^s)} \quad (2.10)$$

4. 2 と 3 を $t(f | e)$ が収束するまで繰り返す.

EM アルゴリズムで得られる解は最適解となる保証はない. しかし, Model1 において, 極大値は 1 つしかなく, 初期値に依らず最適解に収束する.

Model2

Model1 では全ての単語の対応付けの確率が一定であるとしており, 現実的ではない. そこで, Model2 では j 番目の位置である仏単語 f_j と対応する英単語の位置 a_j は, j と英文の長さ l , 仏文の長さ m に依存するとし, 2.11 式のようにする.

$$a(a_j | j, m, l) = P(a_j | a_1^{j-1}, f_1^{j-1}, m, l) \quad (2.11)$$

この仮定を使い, 2.11 式は 2.12 式ようになる. また, Model1 と同じく重要な特徴として, 和と積を入れ換えることができるため, 計算効率を向上させることができる.

$$P(f | e) = \epsilon \sum_{a_1=0}^l \dots \sum_{a_m=0}^l \prod_{j=1}^m t(f_j | e_{a_j}) a(a_j | j, m, l) \quad (2.12)$$

$$= \epsilon \prod_{j=1}^m \sum_{i=0}^l t(f_j | e_{a_j}) a(a_j | j, m, l) \quad (2.13)$$

Model1 と同じく, ラグランジェの未定係数を使い, $\sum_f t(f | e) = 1, \sum_{i=0}^l a(i | j, m, l) = 1$ の制約のもとで 2.13 式を最大化する問題を解くことで, 2.14 式, 2.15 式の 2 つの式が得られる.

$$c(f | e; f, e) = \sum_{j=1}^m \sum_{i=0}^l \frac{t(f | e)a(i | j, m, l)\delta(f, f_j)\delta(e, e_i)}{t(f | e_0)a(0 | j, m, l) + \dots + t(f | e_l)a(l | j, m, l)} \quad (2.14)$$

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

$c(f | e; f, e)$ は対訳ペア中の英単語 e と仏単語 f が対応付けられる回数の期待値を表し, $c(i | j, m, l; f, e)$ は英単語の位置 i が仏単語の位置 j に対応付けられる回数の期待値を表す. Model2 は次の繰り返し計算によって推定される.

1. $t(f | e)$ に適当な初期値を設定する.
2. 対訳ペアの集合 $(f^{(s)}, e^{(s)}), 1 \leq s \leq S$ において, $c(f | e; f, e), c(i | j, m, l; f, e)$ を求める.
3. 2 からそれぞれの総和を求め, 正規化することで $t(f | e)$ を再推定する.
4. 2 と 3 を $t(f | e)$ が収束するまで繰り返す.

Model2 は複数の極大値を持ち, 必ずしも最適解を得られるとは限らない. しかし, Model1 は $a(i | j, m, l) = (l + 1)^{-1}$ となる Model2 の特別な場合であると考えられる. しかし, Model1 は最適解が保証されているため, 初期値として Model1 によって求められたパラメータを使い, Model2 の精度を高めることができる.

Model3

Model1, Model2 では, 英単語と仏単語の対応は 1 対 1 の場合のみを想定していた. しかし, 実際の翻訳では, 複数の単語が 1 つの単語に翻訳されることや, 翻訳先で省略されることがしばしば起こる. このような英単語が対応する仏単語の数を繁殖数と呼ぶ.

また, 対応付けられる英単語と仏単語は必ずしも近い位置にあるとは限らず, 離れた位置にあることもある. この現象を歪みと呼ぶ. Model3 ではこの 2 つの現象を想定し, 次の 3 つのパラメータにより表される.

- 翻訳確率 $P(f | e)$ 仏単語 f が英単語 e に翻訳される確率
- 繁殖数 $P(\phi | e)$ 英単語 e が ϕ 個の仏単語と対応する確率

- 歪み確率 $d(j | i, l, m)$ 英文の長さが l , 仏文の長さが m のとき, 単語位置 i の英単語が j の仏単語に翻訳される確率

これに加え, 英単語に翻訳されない仏単語の数を ϕ_0 とし, その確率 p_0 を 2.16 式で求める. 歪み確率は $1/\phi_0!$ とし, $p_0 + p_1 = 1$ で p_0 と p_1 は 0 ではないとする.

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

以上により Model3 は 2.17 式のようになる.

$$\begin{aligned} P(f | e) &= \sum_{a_1=0}^l \dots \sum_{a_m=0}^l P(f, a | e) \\ &= \sum_{a_1=0}^l \dots \sum_{a_m=0}^l \binom{m - \phi_0}{\phi_0} \phi_0^{m-2\phi_0} p_1^{\phi_0} \prod_{i=1}^l \phi_i! n(\phi_i | e_i) \times \prod_{j=1}^m t(f_i | e_{a_j}) d(j | a_j, m, l) \end{aligned} \quad (2.17)$$

Model3 は Model1 や Model2 と異なり, 効率的に計算を行うことができない. 全ての単語対応を考慮すると莫大な計算量が必要となるため, 期待値を近似によって求める.

対訳ペア中の全てのアライメントにおいて, 最大となるアライメントをビタビアライメントと呼び, $V(f | e)$ とする. また, あるアライメント a 中の, 1つの仏単語 f の対応先を別の英単語に移動させたことによって得られるアライメント a' の場合において, a' は a の近傍であると定義する. 近傍の集合を $N(a)$ で表す.

Model3 では, a の近傍である a' の確率を, a の確率から効率的に求めることができる. 位置 j にある仏単語の対応先を, 位置 i にある英単語から i' に変更したときのアライメントを a' とし, i と i' は 0 でなおとすると, 2.18 式が成り立つ.

$$P(a' | e, f) = P(a | e, f) \frac{(\phi_{i'} + 1) n(\phi_{i'} + 1 | e_{i'}) n(\phi_{i'} - 1 | e_{i'}) t(f_i | e_{i'}) d(j | i', m, l)}{\phi_i n(\phi_i | e_i) n(\phi_i | e_{i'}) t(f_i | e_i) d(j | i, m, l)} \quad (2.18)$$

$\phi_{i'}$ はアライメント a における位置 i' の単語の繁殖数であり, この単語のアライメント a' における繁殖数は $\phi_{i'} + 1$ である. 同様に, i もしくは i' が 0 であるとき, 交換による近傍の場合についても計算することができる.

このことより, Model2 において, 最大確率となるアライメント $V(f | e; 2)$ から, その近傍で確率最大となるアライメントを選択することで, Model3 のビタビアライメントを近似することができる. また, Model2 のビタビアライメントがある対応 $a_j = i$ を固定し

た計算も加えることで、妥当な範囲で広く計算することができ、より確実性の高い推定を行うことができる。

Model4

Model3 では単語の位置は絶対位置で考えていたが、Model4 では英単語と仏単語の位置を相対位置で表現する。また、Model3 では各単語同士の位置は考慮に入られていない。例えば、名詞と形容詞の出現位置を見たとき、英語における形容詞は名詞の前に出現するが、仏語では名詞の後ろに出現する性質がある。そのような場合を表現するため、英単語 e_i に対応する仏単語の位置を j 、直前の英単語 e_{i-1} に対応する仏単語の位置を j' とし、その相対位置 $j - j'$ は、直前の英単語と仏単語のクラスに依存するとする。

繁殖数を考慮し、2つの場合を考える。ある繁殖数 1 以上の英単語に対応する仏単語のうち、最も文頭に近い単語を *head* と呼ぶ。単語が *head* のとき、Model3 の歪み確率 $d(j | i, m, l)$ は 2.19 式のようになる。

$$d_1(k - k' | \text{class}(e_{|i-1|}), \text{class}(f_i)) \quad (2.19)$$

$e_{|i-1|}$ は直前の繁殖数が 0 でない英単語、 j' は $e_{|i-1|}$ に対応する仏単語の位置を表す。ただし、複数の仏単語が対応している場合、その単語の位置の平均をとった整数値とする。*class* は単語クラス 50 前後の種類を用いる。

次に単語が *head* でない場合、単語のクラスのみ依存するとし、2.20 式のような式となる。ただし、このときの j' は同じ英単語に対応付けられている、直前に出現した仏単語の位置とする。

$$d_{>1}(k - k' | \text{class}(f_i)) \quad (2.20)$$

Model4 も Model3 と同様の方法で推定を行うことができる。Model3 のビタビアライメントから近傍を取ることで、Model4 のビタビアライメントを求める。

Model5

Model4 では単語の位置に関して、直前の単語についてしか考慮していない。そのため、複数の単語を同じ位置に配置したり、単語が存在しない位置が生成されてしまう。こ

のような問題を回避するために、単語を空白部分に配置するように制約を加えたものが Model5 である。

2.2.2 GIZA++

IBM 翻訳モデルを推定するツールとして GIZA++[3] がある。GIZA++ は Franz によって作成され、Model1 から Model5 までの IBM モデルと HMM アライメントモデルを推定することができる。それぞれのモデルの推定に用いる EM アルゴリズムの繰り返し計算数は任意に設定することができる。デフォルトでは HMM、Model1、Model3、Model4 をそれぞれ 5 回ずつ試行する。

2.2.3 フレーズテーブルの作成法

手順1 最尤な単語 alignment の獲得

まず，GIZA++により IBM 翻訳モデルを推定することで最尤な単語 alignment を得る．これを英日，日英の両方向に対して行う．なお，IBM モデルは単語を基本単位とした翻訳モデルである．対訳文，“言語はコミュニケーションの道具である．”，“language is a means of communication.” を学習文とした時の例を図 2.3 に示す．

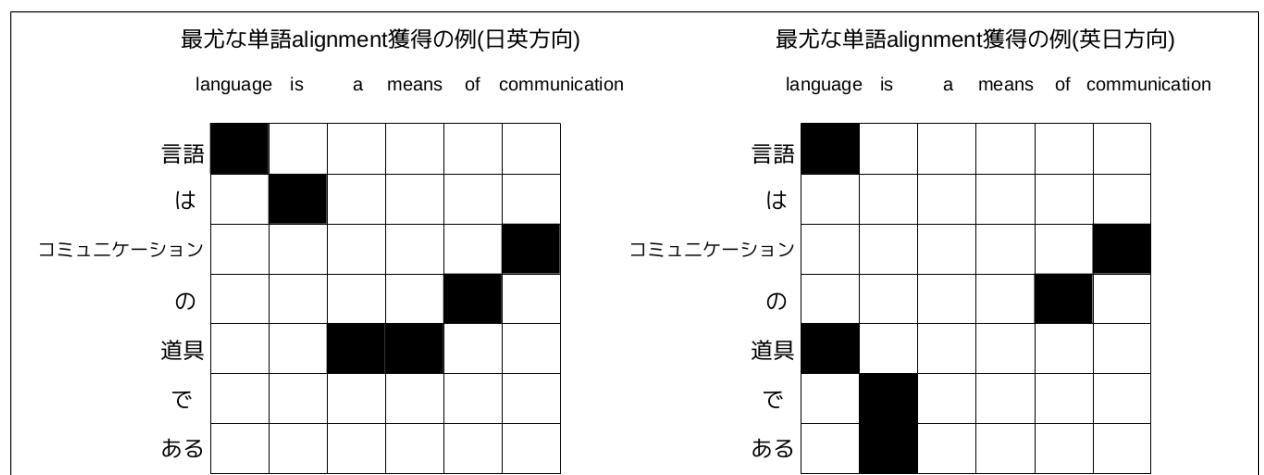


図 2.3: 日英，英日方向の最尤な単語 alignment の例

手順2 対称化された単語 alignment の計算

次に，両方向の alignment から，両方向に 1 対多の対応を認めた，単語 alignment を計算する．この単語 alignment は基本的に両方向の単語対応の積集合と和集合の中間をヒューリスティックスで求める．なお，積集合 (intersection) は，両方向ともに単語対応が存在する場合のみ対応を残し，和集合 (union) は，少なくとも片方向に単語対応が存在する場合，単語対応を残す．対称な単語対応を求めるヒューリスティックス (grow-diag-final など) は，まず積集合から始まり，和集合にしかない単語対応が妥当であるかを判断しながら，単語対応を除々に加える．対称化された単語 alignment の獲得の例を，図 2.4(intersection)，図 2.5(union)，図 2.6(grow-diag)，図 2.7(grow-diag-final) に示す．

図 2.4 が示すように，intersection は日英，英日の両方向ともに単語対応が存在する場合のみ単語対応を残す．

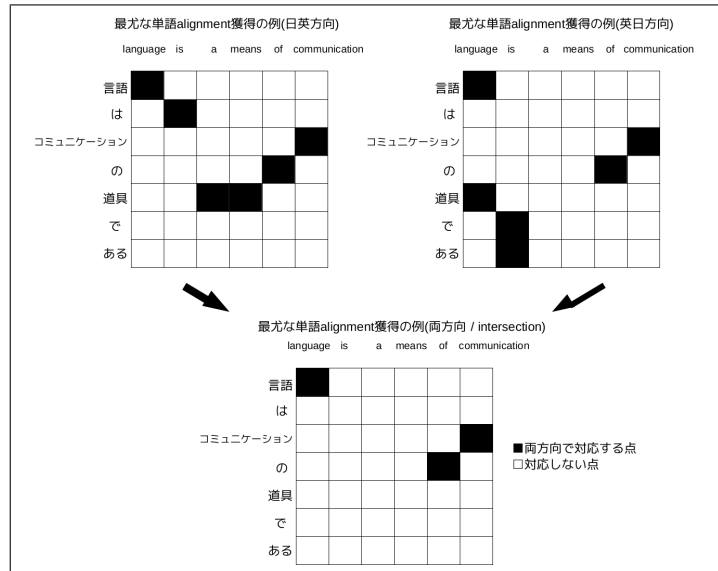


図 2.4: 最尤な単語 alignment の例 (intersection)

図 2.5 が示すように union は，少なくとも片方向に単語対応が存在する場合，単語対応を残す．

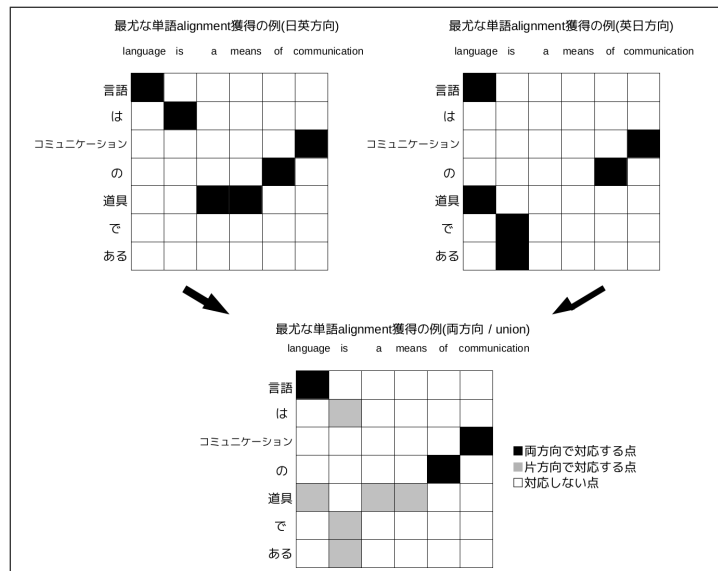


図 2.5: 最尤な単語 alignment の例 (union)

図 2.6 が示すように grow-diag は、intersection の拡張である。grow-diag は、intersection の対応点に加えて、片方向で対応点があり、かつ、intersection の隣り合う点を両方向の対応点にする。

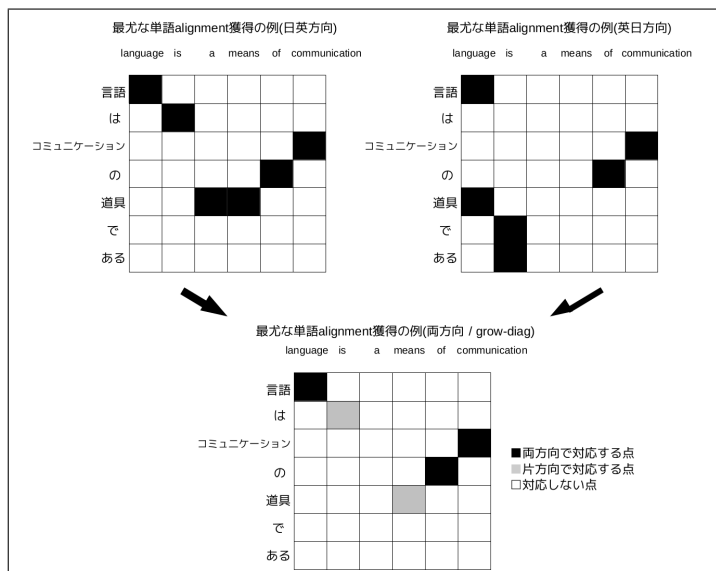


図 2.6: 最尤な単語 alignment の例 (grow-diag)

図 2.7 が示すように grow-daig-final は、grow-diag の拡張である。grow-diag-final は片方向で対応点があり、かつ、grow-diag において、少なくとも片方の単語の対応しない点を grow-diag に追加する。

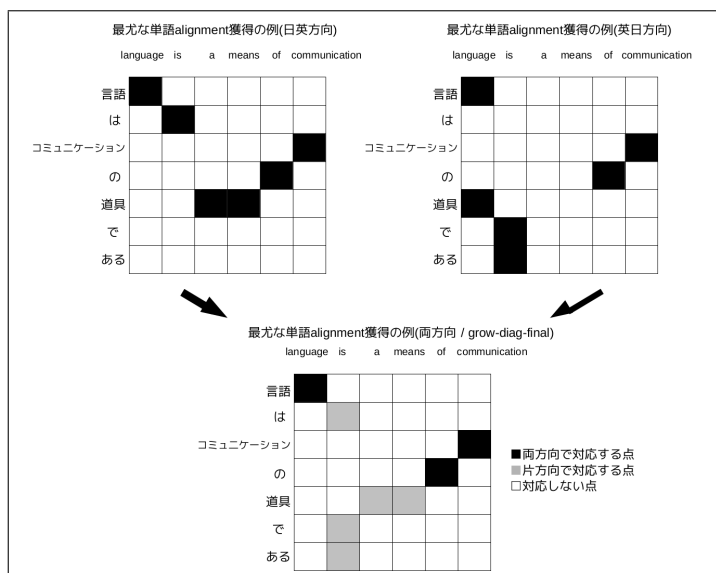


図 2.7: 最尤な単語 alignment の例 (grow-diag-final)

手順3 フレーズテーブルの抽出

対称化された単語 alignment のうち矛盾しないすべてのフレーズ対応を得る．そのフレーズに対して翻訳確率を計算し，値を付与する．作成されたフレーズテーブルを表 2.2 に示す．

表 2.2: 作成されたフレーズテーブルの例 (grow-diag-final)

言語	language	1 1 1 1
コミュニケ - ション	communication	1 1 1 1
コミュニケーションの	of communication	1 1 1 1
コミュニケーションの 道具	a means of communication	1 1 1 1
道具	a means	1 1 1 1
の	of	1 1 1 1
の 道具	a means of	1 1 1 1
言語はコミュニケーションの道具である	language is a means of communication	1 0.0987654 1 9.28785e-05

2.2.4 フレーズ対応の抽出

grow-diag-final による，両方向の単語対応から，フレーズ対応を抽出する．例を図 2.8 に示す．

矛盾しないフレーズの対応						矛盾したフレーズの対応					
	is	a	means	of	Communi- cation		is	a	means	of	Communi- cation
は	■					は	■				
コミュニケーション					■	コミュニケーション					■
の					■	の				■	
道具		■	■			道具		■	■		

図 2.8: フレーズ対応の抽出例

抽出されるフレーズ対応の条件は，フレーズ対応が“矛盾しない”ことである．“矛盾しない”とは，フレーズ対応に含まれている単語が互いに対応しており，かつ，他の語

に対応しないということを表す．例えば，フレーズ対応 “コミュニケーション の ||| of communication” において，日本語の単語 “コミュニケーション” と “の” と，英語の単語 “of” と “communication” は，それぞれフレーズ対応に含まれる単語にしか対応していない．そのため，“コミュニケーション の ||| of communication” は “矛盾しない” と判断し，抽出できる．

一方，フレーズ対応 “コミュニケーション の 道具 ||| means of communication” は，日本語の単語 “道具” が，フレーズ対応に含まれない英単語 “a” と対応している．そのため，“コミュニケーション の 道具 ||| means of communication” は “矛盾した” と判断し，抽出しない．ただし，フレーズ対応を “コミュニケーション の 道具 ||| a means of communication” ととった場合は，フレーズ対応に含まれている単語が互いに対応しており，かつ，他の語に対応しないため，“矛盾しない” ことに注意したい．

図 2.8 から抽出した “矛盾しない” フレーズ対応を表 2.3 に示す．

表 2.3: 図 2.8 から抽出した “矛盾しない” フレーズ対応の例

は is
コミュニケーション communication
の of
コミュニケーション の of communication
コミュニケーション の 道具 a means of communication
道具 a means

2.2.5 抽出したフレーズ対応の確率付け

抽出したフレーズ対応に対して，確率付けを行う．日本語フレーズ J_{phrase} と英語フレーズ E_{phrase} からなるフレーズ対応の確率は以下の式で計算される．

$$P(J_{phrase} | E_{phrase}) = \frac{\text{対訳学習文中で } J_{phrase} \text{ と } E_{phrase} \text{ が同時に出現した数}}{\text{対訳学習文中で } E_{phrase} \text{ が出現した数}}$$

$$P(E_{phrase} | J_{phrase}) = \frac{\text{対訳学習文中で } J_{phrase} \text{ と } E_{phrase} \text{ が同時に出現した数}}{\text{対訳学習文中で } J_{phrase} \text{ が出現した数}}$$

2.3 言語モデル

言語モデルは単語の列，文字の列が起こる確率を与えるモデルである．日英翻訳では，より英語らしい文に対して，高い確率を与えることで，翻訳モデルで翻訳された訳文候

補の中から英語として自然な文を選出する。

代表的な言語モデルに N -gram がある。 N -gram は “単語の列 $w_1^i = w_1, w_2, \dots, w_i$ の i 番目の単語 w_i の生起確率 $P(w_i)$ は直前の $(i - 1)$ 単語に依存する”，という仮説に基づくモデルである。計算式を 2.21 式に示す。

$$P(w_1^i) = \prod_{i=1}^{k=1} P(w_i | w_{i-1}) \quad (2.21)$$

例えば、「I eat dinner.」という文字の列に対する 2-gram モデルは以下のようになる。

$$P(e = I \text{ eat dinner.}) \approx P(He) \times P(eat | He) \times P(dinner | eat) \times P(. | dinner)$$

3-gram であれば、「I eat」という 2 単語の次にくる単語が「dinner」である確率を考える。なお、学習文に表れない N 単語連鎖を確率 0 として、 N -gram モデルを作成すると、性能が悪化する。よって、 N -gram モデルはスムージング手法によって 0 でない確率を割り当てて作成する。代表的なスムージング手法として、バックオフスムージングがある。バックオフスムージングは、高次の N -gram が存在しない場合、低次の N -gram で代用する。この代用する低次の確率を改良した手法が、Kneser-Ney スムージングである。言語モデルにおける N -gram の作成には一般に Kneser-Ney スムージングが用いられている。3gram の例を 2.22 式に示し、言語モデルの例を表 2.4 に示す。

$$P(w_i | w^{i-1i-2}) = \begin{cases} \alpha \times p(w_i | w_{i-2}^{i-1}) & 3\text{gram が存在する} \\ \beta \times p(w_n | w_{n-1}) & 3\text{gram は存在せず } 2\text{gram が存在する} \\ p(w_n | w_{n-1}) & \text{それ以外} \end{cases} \quad (2.22)$$

ここで、 α をディスカウント係数、 β をバックオフ係数と呼ぶ。

表 2.4: 言語モデルの例

-1.557193	Because of
-1.92458	absence from
-0.8959993	account for

例えば、1 番目の行は $\log_{10}(P(of | Becaouse)) = -1.557193$ を意味している。

言語モデルの学習に用いた英文の例を 2.5 に示し，各スムージングのパラメータを用いたときの，2-gram までの範囲で作成した言語モデルの例を表 2.6 に示す．

表 2.5: 言語モデル作成に用いた英文の例

Faith can movemountains .
Please move over more to the right .
His way of thinking goes too far .
She has learned a good deal of literature .
A computer employs the two digits of the binary system .
As you know , the Hong Kong Textile Show falls on April 24 .

表 2.6: スムージングパラメータ別の言語モデルの例

パラメータ ukndiscount を用いた場合の例
-3.202558 the boy
-2.963157 the child
-2.298372 at all
パラメータ kndiscount を用いた場合の例
-3.179705 the boy
-2.978517 the child
-2.283957 at all
パラメータ ndiscount を用いた場合の例
-3.068275 the boy
-3.068275 the child
-1.493827 at all
デフォルトのパラメータ (Good-Tuning-discounting) を用いた場合の例
-3.385125 the boy
-3.385125 the child
-2.405329 at all

2.3.1 リオーダーリングモデル

リオーダーリングモデルとは，翻訳の前後での語順変化に対して，確率を与えるためのモデルである．リオーダーリングには，Lexicalized Block Orientation モデル [5] および，歪みペナルティ [6] が広く用いられている．どちらも目的言語側において， i 番目および $i + 1$ 番目，すなわち目的言語側が隣接しているフレーズ対のペアに対する局所的なリオーダーリングの評価を行う．次にそれぞれのモデルについて説明する．

Lexicalized Block Orientation モデル

Lexicalized Block Orientation (IBO) モデルでは，目的言語側における， i 番目および $i + 1$ 番目のフレーズ \bar{e}_i, \bar{e}_{i+1} に対応するフレーズ対のペアに対して，その原言語側フレーズ \bar{f}_{c_i} と $\bar{f}_{c_{i+1}}$ との位置関係を次のような 3 クラスに分類する．

$$class(c_i, c_{i+1}) = \begin{cases} monotone(c_{i+1} = c_i + 1 \text{ のとき}) \\ swap(c_{i+1} = c_i - 1 \text{ のとき}) \\ discontinuous(\text{それ以外}) \end{cases}$$

目的言語側で隣接するフレーズに対して，*monotone* は原言語側において隣接かつ目的言語と同順であること，*swap* は原言語側において隣接かつ目的言語側と逆順であること，*discontinuous* は原言語側において両者が離れていることを表す [5] ．

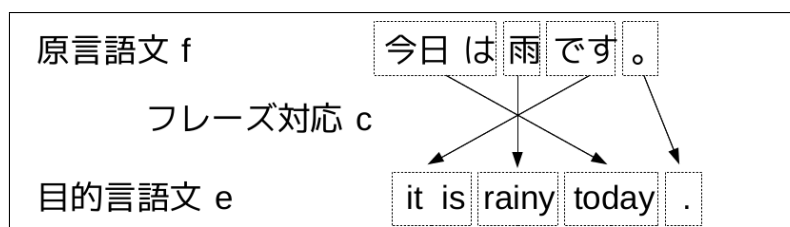


図 2.9: reordering における swap の例

例えば，図 2.10 において，“today”-“今日は” のフレーズ対と “雨”-“rainy” のフレーズ対とは *swap* の関係をなしている．

LBO モデルでは，上記の 3 クラスを用いた各フレーズ対間のリオーダーリング確率の積として，リオーダーリングの確率を 2.23 式のように近似する [5] ．

$$P(c_1^I | e) \approx \prod_{i=1}^I P(class(c_i, c_{i+1}) | \bar{e}_i, \bar{e}_{i+1}) \quad (2.23)$$

対数線形モデルの枠組みにおいては翻訳精度向上のために、 $P(class(c_i, c_{i+1}) | \bar{e}_i, \bar{e}_{i+1})$ の代わりとして、より多くの情報をもつ $P(class(c_i, c_{i+1}) | \bar{f}_i, \bar{f}_{i+1}, \bar{e}_i, \bar{e}_{i+1})$ を素性に使用する場合が多い。

歪みペナルティ

歪みペナルティは、目的言語側が隣接しているフレーズ対の原言語側フレーズ \bar{f}_{c_i} と $\bar{f}_{c_{i+1}}$ との間に含まれる単語数 (単語距離) に比例してリオーダーリングに対するペナルティを与える。すなわち、この素性は翻訳結果の *monotone* 性に対してバイアスをかける。文頭を 0 としたときの \bar{f}_{c_i} の終端単語位置を end_i 、 $\bar{f}_{c_{i+1}}$ の開始単語位置を $start_{i+1}$ とすれば、歪みペナルティ $d(end_i, start_{i+1})$ は 2.24 式のように表せる [6]。

$$d(end_i, start_{i+1}) = |end_i - start_{i+1} + 1| \quad (2.24)$$

なお、歪みペナルティは対数線形モデルの枠組みでのみ考慮可能な素性である。

2.4 デコーダ

デコーダは翻訳モデルと言語モデルの確率が最大となる出力文を探索し、出力する。入力文として「言語はコミュニケーションの道具である」を与えた場合の探索の例を図 2.10 に示す。

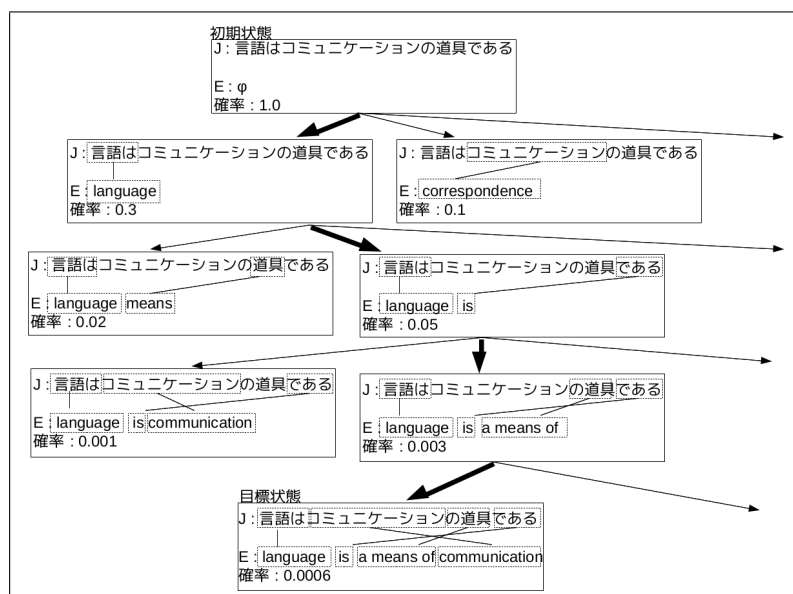


図 2.10: 入力文「言語はコミュニケーションの道具である」の探索例

デコーダは全く翻訳されていない空の初期仮説から探索を始める．そして，ある仮説において1つのフレーズを選択し，翻訳モデルのフレーズ対を用いて，新しい仮説を作り，そのステップを繰り返す．このとき，翻訳候補の確率を計算し，その翻訳候補のスコアとする．1つのフレーズを選択では，原言語の文で未選択の単語列から語の部分文の文末側に加えていく．仮説は優先順位つきキューで管理し，原言語の文の全ての単語を選択したら探索は終了する．

しかし，文が長くなると，用いるフレーズ対の組み合わせは膨大な数になる．そのため，全ての翻訳候補を探索し，最適な出力文を決定することは困難である．そこで，現在，計算量を減らす手法として，ビームサーチ法とマルチスタック法を組み合わせた手法が一般的に用いられている．

代表的なデコーダに `moses`[7] がある．`moses` はいくつかのパラメータを設定することができる．

2.4.1 ビームサーチ法

ビームサーチ法は，探索の計算量を減らすために用いる．ビームサーチ法は，翻訳候補の探索木において，翻訳確率の低い翻訳候補を枝刈りし，探索範囲を限定する．枝刈りは“`histogram pruning`”と“`threshold pruning`”によって行う．“`histogram pruning`”は確率の高い翻訳候補のみを一定数残す枝刈り法である．“`threshold pruning`”は一定の確率以上の翻訳候補のみを残す枝刈り法である．この2つの枝刈り法を用いて，探索範囲を限定する．しかし，図 2.10 からわかるように，翻訳が進むほど，翻訳候補の確率は小さくなる．そのため，翻訳が進んだ翻訳候補と翻訳が進んでいない翻訳候補を比較したとき，翻訳が進んだ翻訳候補ほど枝刈りの対象となる可能性が高い．

2.4.2 マルチスタック法

ビームサーチ法の問題を解決するために，ビームサーチ法とマルチスタック法を組み合わせる．マルチスタック法は翻訳候補を翻訳した単語の数毎に分ける．そして，分けた翻訳候補の中で，ビームサーチ法を適用する．図 2.11 にマルチスタック・ビームサーチ法を適用した例を示す．

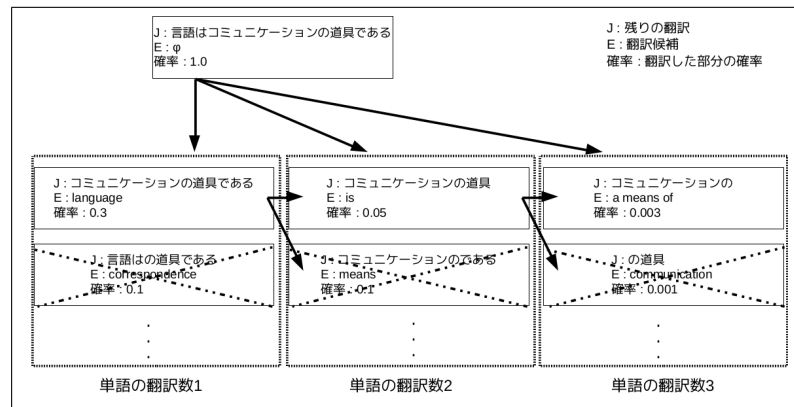


図 2.11: マルチスタック・ビームサーチ法の適用例 (入力文「言語はコミュニケーションの道具である」)

図 2.11 において，“histogram pruning”の残す翻訳候補を 1 とした場合，× のついた翻訳候補からは，探索が行われない。

2.4.3 moses のパラメータ

moses で設定できるパラメータの例を以下に示す。

- weight-l: 言語モデルの重み
- weight-t: 翻訳モデルの重み
- weight-d: リオーダーリングモデルの重み
- weight-w: 目的言語の長さに関するペナルティー
- distorsion-limit: フレーズの並び替えの範囲制限

なお，パラメータチューニングによって，パラメータの最適値を求めることができる。

2.4.4 パラメータチューニング

デコーダは言語モデルや翻訳モデルに対して重みを与えることができる。例えば，言語モデルに対して高い重みを与え，翻訳モデルに低い重みを与えた場合，デコーダは言語モデルの確率 $P(e)$ を重視した翻訳を行う。しかし，各モデルに与える重みを，決定することは難しい。そこで，それぞれのパラメータに対して，パラメータチューニングを

行う。パラメータチューニングには、Minimum Error Rate Training(MERT)[15]を用いることが一般的である。MERTは目的の評価関数を最大とするような翻訳結果が選ばれるように、パラメータを調整する。その際、development データと呼ばれる、試し翻訳を行うデータを与え、各文について上位100個程度の翻訳候補を出力し、その候補の中で重みを変え、より良い翻訳候補が上位にくるようにパラメータを調整する。MERTは後述する自動評価法 BLEU が最大となる翻訳結果が選ばれる重み $\hat{\lambda}$ を計算する。n 個のパラメータの最適化は 2.25 式で表せる。

$$\hat{\lambda}_1^n = \operatorname{argmax}_{\lambda_1^n} BLEU(\operatorname{smt}\lambda_1^n, e_{ref}) \quad (2.25)$$

ここで、 $\operatorname{smt}(\lambda)$ はパラメータ λ が与えられたときの、デコーダの出力文である。また、 $BLEU()$ は BLEU のスコアであり、デコーダの出力文と、入力文に対してあらかじめ用意された正解文 e_{ref} から計算される。重みの最適化は、具体的には以下の手順で行われる。

手順1 λ に初期値を与える。

手順2 λ を用いてデコーディングを行い、確率の高い上位 N 文を出力する。

手順3 上位 N 文の中で BLEU スコアが高い文が上位にくるよう λ を最適化する。

手順4 重みが収束するまで、手順2、手順3を繰り返す。

第3章 それぞれの手法の説明

本章では，提案手法と先行研究の手法，従来手法，追加実験のシステムの説明を行う．提案手法では，2章で述べた日英統計翻訳システムに“英辞郎 [8]”の対訳フレーズを導入する．具体的には，moses を用いた日英統計翻訳システムの翻訳モデルのフレーズテーブルに英辞郎を追加する．そのための手法を述べる．先行研究は，提案手法と同様に，moses を用いた日英統計翻訳システムに対訳フレーズを導入する．先行研究では“日英重文複文文型パターン辞書 [9]”を，日英統計翻訳システムの翻訳モデルに追加する．しかし，本研究では，提案手法と先行研究の手法のみの違いによる翻訳精度を調べるために，日英重文複文文型パターン辞書の代わりに，英辞郎を用いる．従来手法は，moses を用いた日英統計翻訳である．追加実験では，提案手法で追加した翻訳対と先行研究で追加した翻訳対を日英統計翻訳のフレーズテーブルに追加する．

3.1 提案手法

本研究では統計翻訳に翻訳対を導入するために，翻訳対に対して翻訳確率の付与を行う．翻訳確率を付与するために，2.2節で述べたフレーズ対の翻訳確率を利用する．翻訳対がフレーズ対と完全に一致すれば，フレーズ対に付与されている翻訳確率を翻訳対に付与する．部分的に一致すれば，対訳フレーズ対に付与されている翻訳確率にペナルティをかけた値を翻訳対に付与する．

3.1.1 英辞郎

英辞郎は，EDP(Electronic Dictionary Project) がアップデートし続けている英和・和英辞書である．そのため，英辞郎には通常の英語辞書にない新しい語彙や複雑な言い回しも含まれる．本研究では人手で作成した対訳フレーズ対として英辞郎の対訳フレーズ対を用いる．また，英辞郎の日本語フレーズ，英語フレーズをまとめて，翻訳対と呼ぶ．本研究で使用した翻訳対の総数は1,587,241文対である．表3.1に英辞郎の例を示し，表3.2にフレーズテーブルの形式に変換を行った，翻訳対の例を示す．

表 3.1: 英辞郎の例

hoe one's own row	: 自分で仕事をする、自分一人でやっていく
have an operation	: 手術を受ける / 【用例】 You need to have an operation. : 《旅行・病気・診断》手術が必要です。 医者の診断
medical team	: 医療チーム、医療班
child-rearing tasks	: 育児 {いくじ} の仕事 {しごと}
feel uneasy	: 胸騒ぎがする、気がとがめる、気が気でない、気が落ち着かない、気になる、気がかりである、気兼ねする、心もとなく思う、心細い、心配する、不安に思う、不安の念を抱く
indicate one's personal habits	: 個人 {こじん} の習慣 {しゅうかん} を示す

表 3.2: 翻訳対の例

自分で仕事をする	hoe one 's own row
自分で仕事をする	hoe my own row
自分で仕事をする	hoe your own row
手術を受ける	have an operation
手術を受ける	come under the knife
手術を受ける	go through an operation
医療 チーム	medical team
医療 テクノロジー アセスメント	medical technology assessment
医療 データ 処理	medical data processing
育児 の	child-care
育児 の	childcare
育児 の 仕事	child-rearing tasks
気になる	anxious
気になる	become curious
気になる	feel uneasy
個人の 習慣 を 示す	indicate one 's personal habits
個人の 習慣 を 示す	indicate my personal habits
個人の 習慣 を 示す	indicate your personal habits
個人の 習慣 を 示す	indicate his personal habits
個人の 習慣 を 示す	indicate her personal habits

3.1.2 部分的に一致した翻訳対の翻訳確率

本研究ではより多くの翻訳対を統計翻訳に導入するために，翻訳対とフレーズ対が部分的に一致した際の翻訳確率を計算する．より類似したフレーズ対の翻訳確率を利用するために，翻訳対とフレーズ対の DP マッチングを行う．DP マッチングにより，翻訳対と最も一致率の高いフレーズ対を抽出する．部分的に一致した翻訳対の例を表 3.3 に示す．

表 3.3: 部分的に一致した翻訳対の例

翻訳対	自分で仕事をする hoe my own row
フレーズ対	仕事をする do the work 0.5 0.002 1 0.005
翻訳対	手術を受ける have an operation
フレーズ対	手術を受けた He had an operation 1 4.4e-05 0.33 2.28e-05
翻訳対	医療チーム medical team
フレーズ対	わがチーム Our team 0.46 0.0525101 0.92 0.366894
翻訳対	育児の仕事 child-rearing tasks
フレーズ対	この仕事 The job 0.05 0.0129424 0.00980392 0.023981
翻訳対	気になる feel uneasy
フレーズ対	になる It will 0.142857 1.81559e-05 0.0196078 0.000642405
翻訳対	個人の習慣を示す indicate her personal habits
フレーズ対	の習慣を the habit of 0.5 0.0160856 1 0.00794435

3.1.3 部分的に一致した翻訳対の確率のペナルティ

翻訳対とフレーズ対が部分的に一致した際，フレーズ対に付与されている翻訳確率を翻訳対に付与することは不適切だと考える．よって，本研究ではフレーズ対のそれぞれの翻訳確率に対して，ペナルティをかける．ペナルティは日本語フレーズと英語フレーズにおいて，それぞれ一致した単語数をそれぞれの総単語数で割った値とする．ペナルティの値を翻訳確率にかけることにより，フレーズ対と部分的に一致した翻訳対に翻訳確率を与える．ペナルティの式を 3.1 式に示す．また，翻訳確率を付与した翻訳対の例を表 3.4 に示す．

$$\text{ペナルティ} = \frac{\text{一致した日単語数}}{\text{全日単語数}} * \frac{\text{一致した英単語数}}{\text{全英単語数}} \quad (3.1)$$

表 3.4: 翻訳確率を付与した翻訳対の例

手術	を受ける		have an operation						
4.444444e-01	1.964551e-05	1.481480e-01	1.017133e-05	2.718			1	3	
医療	チーム		medical team						
1.150000e-01	1.312752e-02	2.300000e-01	9.172350e-02	2.718			50	25	

3.1.4 翻訳確率の付与手順

翻訳対に対して、翻訳確率を付与する手順の説明を以下に示し、手順の図を図 3.1 を以下に示す。

- 手順 1 対訳学習文から grow-diag-final を用いてフレーズテーブルを作成
- 手順 2 翻訳対の形式をフレーズテーブルの形式に変換
- 手順 3 DP マッチングを用いてフレーズテーブルから翻訳確率を抽出し、ペナルティをかけ、翻訳対に付与
- 手順 4 翻訳確率を付与した翻訳対をフレーズテーブルに追加

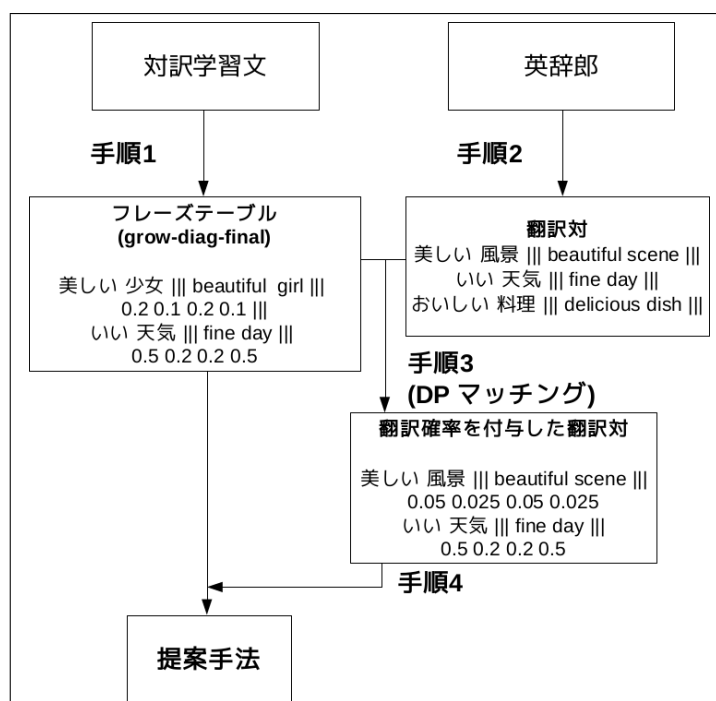


図 3.1: 提案手法の翻訳確率の付与手順

3.2 先行研究

先行研究では翻訳対に対して翻訳確率を付与するために，intersection で作成したフレーズテーブルのフレーズ対と英辞郎で作成した翻訳対のマッチングを行う．フレーズ対と翻訳対が完全に一致した場合にのみ翻訳対にフレーズテーブルで翻訳確率を付与する．そして，grow-diag-final で作成したフレーズテーブルに，翻訳確率を付与した翻訳対を追加する．先行研究の手順の説明を以下に示し，手順の図を図 3.2 に示す．

- 手順 1 対訳学習文から grow-diag-final を用いてフレーズテーブルを作成
- 手順 2 対訳学習文から intersection を用いてフレーズテーブルを作成
- 手順 3 翻訳対の形式をフレーズテーブルの形式に変換
- 手順 4 フレーズテーブルのフレーズ対と完全に一致した翻訳対に翻訳確率付与
- 手順 5 翻訳確率を付与した翻訳対をフレーズテーブルに追加

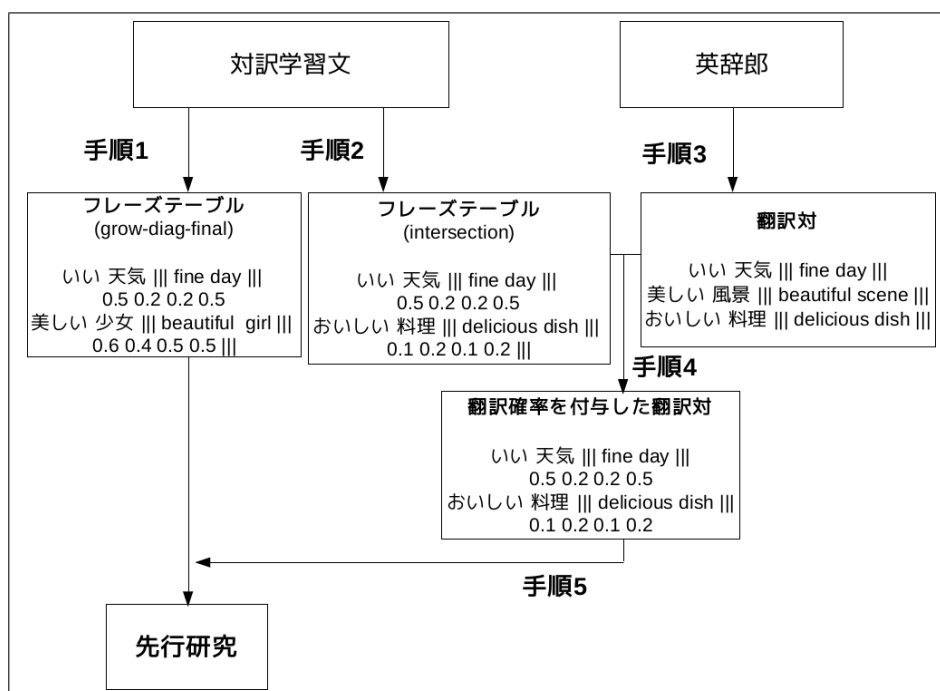


図 3.2: 先行研究の翻訳確率の付与手順

3.3 従来手法

従来手法では，対訳学習文から grow-diag-final で作成したフレーズテーブルを用いて翻訳を行う．従来手法の手順の説明を以下に示し，手順の図を図 3.3 に示す．

手順 1 対訳学習文から grow-diag-final を用いてフレーズテーブルを作成

手順 2 手順 1 で作成したフレーズテーブルを用いて翻訳

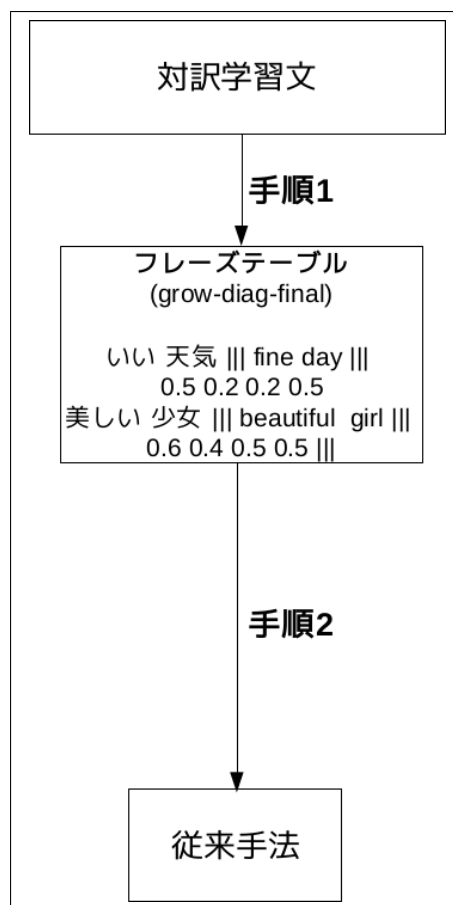


図 3.3: 従来手法の手順

3.4 先行研究と提案手法を組み合わせた手法

追加実験は、先行研究で追加した翻訳対と提案手法で追加した翻訳対を grow-diag-final で作成したフレーズテーブルに追加した翻訳実験を行う。追加実験の手順の説明を以下に示し、手順の図を図 3.4 に示す。

- 手順 1 対訳学習文から grow-diag-final を用いてフレーズテーブルを作成
- 手順 2 対訳学習文から intersection を用いてフレーズテーブルを作成
- 手順 3 翻訳対の形式をフレーズテーブルの形式に変換
- 手順 4 DP マッチングを用いてフレーズテーブルから翻訳確率を抽出し、ペナルティをかけ、翻訳対に付与
- 手順 5 フレーズテーブルのフレーズ対と完全に一致した翻訳対に翻訳確率付与
- 手順 6 翻訳確率を付与した翻訳対をフレーズテーブルに追加

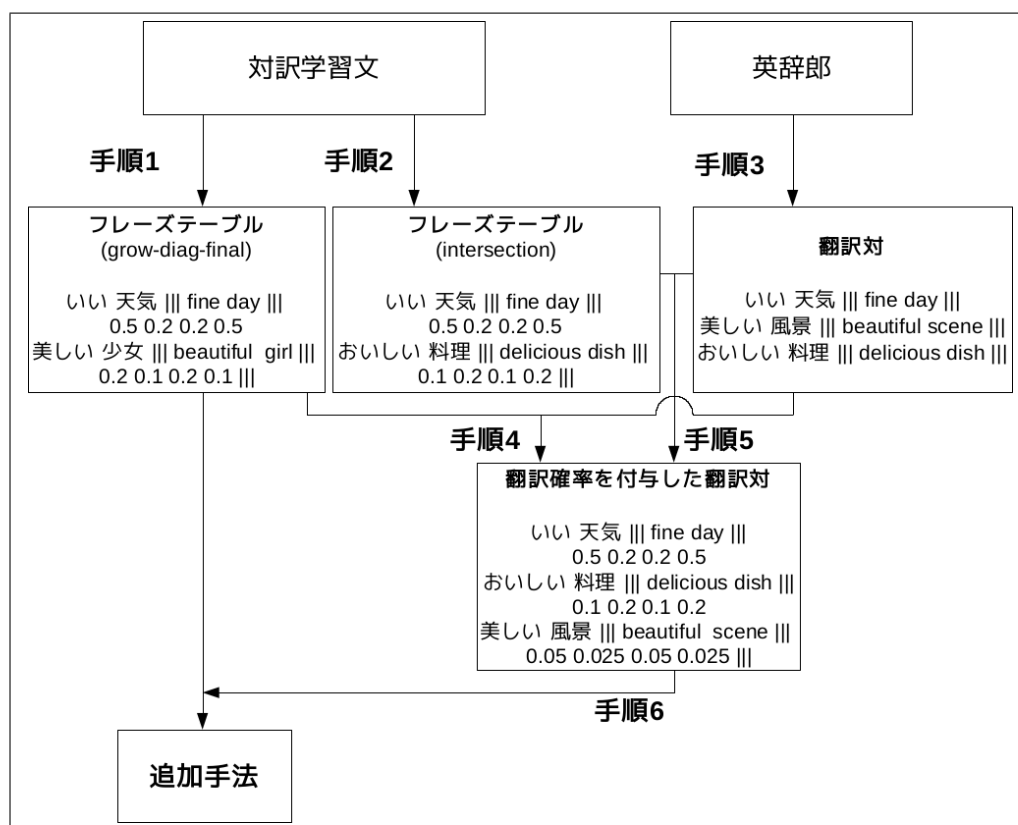


図 3.4: 追加実験の翻訳確率の付与手順

第4章 実験データ・実験環境

本章では、3章で述べた手法で実験を行うための、実験データ、実験環境について述べる。実験データとしては、日英統計翻訳の学習に用いる学習文と翻訳を行うためのテスト文、3章で述べた手法で追加した翻訳対を示す。実験環境としては、翻訳モデルのフレーズテーブル作成のために、使用したヒューリスティックと言語モデル作成のために、使用したスムージング、デコーダのパラメータにおける重みの値について示す。

4.1 実験データ

一般に、日英統計翻訳では前処理として、日本語文を形態素解析を用いて単語に区切る。本研究では、形態素解析器として MeCab[11] を用いる。また、英語文に対して句読点の前後にスペースを入れる。この処理を対訳学習文、テスト文それぞれに行う。

4.1.1 対訳学習文

単文の実験では、対訳学習文として、辞書から抽出した単文の対訳文 100,000 文対を用いる。表 4.1 に単文の対訳学習文の例を示す。

表 4.1: 単文の対訳学習文の例

日本語文	私は映画を見に行く。
英語文	I go to see a movie .
日本語文	鉄道株が株式市場で暴落した。
英語文	Rail stocks took a dive on the stock market .
日本語文	自作のトマトを売っている。
英語文	We sell tomatoes of our own growing .
日本語文	彼は大声で泣いた。
英語文	He cried loudly .
日本語文	聴衆は1000人を越えた。
英語文	There were over one thousand people in the audience .

4.1.2 テスト文

単文の日英翻訳のテスト文として，辞書から抽出した単文 1,000 文対を用いる．

4.1.3 確率を付与した翻訳対

提案手法では翻訳対の総数，1,587,241 文対のうち，単文では 69,728 文対に確率を付与した．

4.2 実験環境

4.2.1 フレーズテーブル

本研究における，フレーズテーブルの作成には，`train-model.perl` を用いる．なお，`grow-diag-final` で作成したフレーズテーブルの総数は 306,895 文対であり，`intersection` で作成したフレーズテーブルの総数は 12,893,466 文対である．

4.2.2 N -gram モデルの学習

言語モデルには， N -gram モデルを用いる． N -gram モデルの学習には “SRILM” [10] を用いる．本研究では，5-gram モデルを用いる．なお，スムージングには `kndiscount` を用いる．

4.2.3 デコーダのパラメータ

本研究の翻訳実験では，MERT による翻訳パラメータの最適化 [15] は行わない．クロスエントロピーを用いるため，“`weight-t`” は “0.5 0.0 0.5 0.0 0.0” とする．翻訳時にフレーズの位置変化に柔軟に対応するため，“`distortion limit`” は “-1” とする．また，“`distortion weight`” は “0.2” とする．

第5章 評価方法

本章では3章で説明した手法を評価するための方法について説明する．本研究では自動評価と人手評価を行う．自動評価は機械翻訳コンテストなどで広く使用されている．自動評価は手軽に機械翻訳の出力を評価できるが，問題があることが知られている．よって，人手による出力文の評価も必要であると考え．そこで，本研究では人手による対比較評価を行う．対比較評価は，2つの機械翻訳の出力を比較し，どちらが入力文の出力として適しているかを判断する評価である．

5.1 自動評価

自動評価には，BLEU[12] と NIST[13] と METEOR[14] の3種類を用いる．

5.1.1 BLEU

BLEU(Bilingual Evaluation Understudy) は予め用意された参照文と比較して，語順が正しい場合に高いスコアを出す．BLEU は5.1式で計算される．

$$BLEU_{score} = BP_{BLEU} \times \exp\left(\frac{1}{N} \sum_{n=1}^N \log P_n\right) \quad (5.1)$$

$$BP_{BLEU} = \begin{cases} 1(c < r) \\ e^{1-\frac{r}{c}}(c \geq r) \end{cases} \quad (5.2)$$

$$P_n = \frac{\sum_{\text{全出力文}} \sum_{\text{出力文}} \text{参照文と一致する } N\text{-gram の数}}{\sum_{\text{全出力文}} \sum_{\text{出力文}} \text{全 } N\text{-gram 数}} \quad (5.3)$$

ここで， BP_{BLEU} は出力文の単語数を c ，参照文の単語数を r としており，出力文の文字数が参照文より長い場合は1であり，BLEU スコアに対して影響を及ぼさない．一方，出力文の文字数が参照文より短い場合は，その短さに応じて BP_{BLEU} が小さくなり，結果としてBLEU スコアが小さくなる． P_n は参照文と一致する N -gram 数を数える際に参照文の要素を重複して数えないように処理を行う．例えば，出力文が “the the the the”，

参照文が “This is the best buy” である場合に修正されない場合，出力文の “the” と一致するものとして参照文の “the” が重複して数えられ， $P_1=5/5$ となる．しかし，修正されると，重複して数えないため， $P_1=1/5$ となる．また，参照文に対して出力文の方が長い場合， P_n の値が小さくなる．即ち，BLEU スコアが小さくなるため，出力文が参照文より長くなる場合のペナルティが式に含まれていることがわかる．

5.1.2 NIST

NIST(National Institute of Standards and Technology) も BLEU と同様に予め用意された参照文と比較し，語順が正しい場合に高いスコアを出す．BLEU と異なる点として，NIST は情報量による重み付けを行う点と 5-gram で評価する点である．

$$NIST_{score} = BP_{NIST} \times \sum_{n=1}^N \frac{\sum_{\text{全出力文}} \sum_{\text{参照文と一致する } w_1 \dots w_n} \text{Info}(w_1 \dots w_n)}{\sum_{\text{全出力文}} \sum_{\text{全 } N\text{-gram 数}} \text{Info}(w_1 \dots w_n)} \quad (5.4)$$

$$BP_{NIST} = \exp\{\beta \log^2 \left[\min\left(\frac{c}{r}, 1\right)\right]\} \quad (5.5)$$

$$\text{Info}(w_i \dots w_n) = \log_2 \frac{\text{評価コーパス中の } w_1 \dots w_{n-1} \text{ 数}}{\text{評価コーパス中の } w_1 \dots w_n \text{ 数}} \quad (5.6)$$

ここで， BP_{NIST} は出力文が参照文より短い場合のペナルティ係数であり，評価する出力文の単語数を c ，全ての参照文を対象として参照文の平均単語数を r としている．これにより，参照文に比べて非常に短い出力文には強いペナルティが課せられ，評価がより低くなり，参照文に長さが近い出力文に課せられるペナルティは軽減され，評価に影響を与えなくなる． $\text{Info}(w_i \dots w_n)$ は情報量の重みを示す．これにより，より出現頻度の低い N -gram，即ち，より情報量多い N -gram がより重み付けされる．2-gram の場合で簡単に説明する．例えば，評価コーパス中に “the” が 10 回出現し，9 回が “the cat” というフレーズであり，1 回が “the dog” というフレーズであったとする．この場合には，明らかに 1 回しか出現しない “the dog” の方が他にない情報を与えてくれる．即ち，情報量が多い．逆に，9 回出現する “the cat” の 1 つ 1 つが持つ情報量は少ない．この場合に 5.6 式より，“the dog” がより重み付けされる．具体的には，“the dog” は 5.6 式による計算値が $\log_2(10/1)$ となり，“the cat” の計算値は $\log_2(10/9)$ となる．このように，NIST では，出力文の内容の正しさに比重を置いた精度判断がされる．

5.1.3 METEOR

METEOR(Metric for Evaluation of Translation with Explicit ORdering) は予め用意された参照文と比較して、単語属性が正しい場合に高いスコアを出す。METEORは5.3式で計算される。

$$METEOR_{score} = F_{mean} \times (1 - P_{en}) \quad (5.7)$$

$$F_{mean} = \frac{P \times R}{\alpha \times P + (1 - \alpha) \times R} \quad (5.8)$$

$$P_{en} = \gamma \times \left(\frac{c}{m}\right)^\beta \quad (5.9)$$

METEORはまず再現率 R と適合率 P に基づく F 値を求め、次に単語の非連続性に対するペナルティとして関数 P_{en} を与える。ペナルティ関数 P_{en} において、 m は出力文と参照文の一致率を表す。そして、 c は一致した単語を対象に、参照文と語順が同じものを1つのまとまりとして統合した場合の、まとまりの数を表す。そのため、出力文と参照文が同じ文であるとき、 $c=1$ となる。また、一致率の計算において、WordNetによる類義語を用いて、似た意味を持つ単語は同一であると判断される。 α, β, γ の値はパラメータである。本研究では、 $\alpha=0.9, \beta=3.0, \gamma=0.5$ の値を用いる。

両評価とも0から1の間で評価され、出力文と参照文が同じ文であるとき1となり、最も良い評価である。

5.2 人手評価

本研究では、人手評価として対比較評価も行う。評価基準を以下に示す。なお、比較対象とは、従来手法もしくは先行研究である。

- 提案手法 提案手法の結果が比較対象の結果より優れている場合
- 提案手法 × 提案手法の結果が比較対象の結果より劣っている場合
- 差なし どちらの結果も同等に理解できるまたは、同等に理解できない場合
- 同一出力 提案手法の結果と先行研究の結果が同一の文である場合

第6章 従来手法と提案手法の比較

本章では3章で述べた従来手法と提案手法の比較について述べる．比較はフレーズテーブルの量，自動評価，人手評価である．

6.1 フレーズテーブルの総数の比較

表 6.1 に従来手法と提案手法における，フレーズテーブルの総数と追加した翻訳対の数を示す．

表 6.1: フレーズテーブルの総数と追加した翻訳対の比較

	フレーズテーブルの総数	追加した翻訳対の数
従来手法	306,895	0
提案手法	376,653	69,758

6.2 翻訳精度の自動評価

テスト文に単文を用いて翻訳実験を行う．表 6.2 に翻訳実験結果に示す．

表 6.2: 単文における翻訳精度の評価

単文	BLEU	NIST	METEOR
従来手法	0.1189	4.1477	0.3730
提案手法	0.1215	4.1668	0.3735

表 6.2 の結果から提案手法は従来手法より自動評価がよいことから，提案手法は有効である．

6.3 翻訳精度の人手評価

人手による対比較評価では，提案手法の結果と比較対象の結果から，それぞれ 100 文ランダムに抽出し，どちらの文が入力文として適切であるかを判断する．評価基準は 6.2 節で述べた基準である．表 6.3 に従来手法と提案手法における人手評価の結果を示す．

表 6.3: 従来手法と提案手法の対比較評価結果

	提案手法	提案手法 ×	差なし	同一出力
単文	5	0	5	90

表 6.3 の結果から提案手法は従来手法より人手評価がよいことから，提案手法は有効である．

6.3.1 翻訳結果の例

従来手法と提案手法の翻訳結果の例を示す．表 6.4 に提案手法が優れていると判断した例を示す．

表 6.4: 提案手法が優れていると判断した例

入力文	この仕事は細心の注意を要する。
参照文	This work needs close attention .
従来手法	This task requires a 細心 attention .
提案手法	This task requires a religious attention to detail .
入力文	難局を収拾する道はそれしかない。
参照文	It is the only way to save the difficult situation .
従来手法	難局 収拾 only the way it .
提案手法	The only way that save a difficult situation .
入力文	きょうはバスケットボールの練習がある。
参照文	Today we have basketball practice .
従来手法	There is a basketball for the today .
提案手法	There is a basketball practice today .
入力文	この国では情報は自由に流れる。
参照文	In this country data flows freely .
従来手法	Information flows in this country is at .
提案手法	Information flows in this country free .
入力文	アメリカやイギリスでは、第一次世界大戦のときに、燃料の節約を目的に導入された。
参照文	The United States and Britain introduced the system during World War I to conserve fuel .
従来手法	In the United States and Britain , at the end of World War was introduced to save and fuel .
提案手法	In the United States and Britain , at the end of World War I was introduced to save and fuel .

第7章 先行研究と提案手法の比較

本章では3章で述べた先行研究と提案手法の比較について述べる．比較はフレーズテーブルの量，自動評価，人手評価である．

7.1 フレーズテーブルの総数の比較

表 7.1 に先行研究と提案手法における，フレーズテーブルの総数と追加した翻訳対の数を示す．表 7.1 から，提案手法は従来手法と比較して，約 3.5 倍の翻訳対を追加した．

表 7.1: フレーズテーブルの総数と追加した翻訳対の比較

	フレーズテーブルの総数	追加した翻訳対の数
従来手法	324,972	18,077
提案手法	376,653	69,758

7.2 翻訳精度の自動評価

テスト文に単文を用いて翻訳実験を行う．表 7.2 に単文の翻訳実験結果を示す．

表 7.2: 単文における翻訳精度の評価

単文	BLEU	NIST	METEOR
先行研究	0.1267	4.3047	0.3850
提案手法	0.1215	4.1668	0.3735

表 7.2 の結果から提案手法は先行研究より自動評価が悪いことから，先行研究と比較すると有効ではない．

7.3 先行研究と提案手法の人手評価

表 7.3 に先行研究と提案手法における人手評価の結果を示す．なお，評価基準は 6.2 節で述べた基準である．

表 7.3: 先行研究と提案手法の対比較評価結果

	提案手法	提案手法 ×	差なし	同一出力
単文	3	5	8	84

表 7.3 の結果から提案手法は先行研究より人手評価が悪いことから，先行研究と比較すると有効ではない．

7.3.1 翻訳結果の例

従来手法と提案手法の翻訳結果の例を示す．表 7.4 に提案手法が優れていると判断した例を示し，表 7.5 に提案手法が劣っていると判断した例を示す．

表 7.4: 提案手法が優れていると判断した例

入力文	きょうはバスケットボールの練習がある。
参照文	Today we have basketball practice .
先行研究	There is a basketball for the today .
提案手法	There is a basketball practice today .
入力文	この国では情報は自由に流れる。
参照文	In this country data flows freely .
先行研究	Information flows in this country is at .
提案手法	Information flows in this country free .
入力文	アメリカやイギリスでは、第一次世界大戦のときに、燃料の節約を目的に導入された。
参照文	The United States and Britain introduced the system during World War I to conserve fuel .
先行研究	In the United States and Britain , at the end of World War was introduced to save and fuel .
提案手法	In the United States and Britain , at the end of World War I was introduced to save and fuel .

表 7.5: 提案手法が劣っていると判断した例

入力文	ポルシェがそのクラスで優勝した。
参照文	The Porsche won its class .
先行研究	The Porsche won the championship in the class
提案手法	The ポルシェ won the championship in the class .
入力文	この仕事は細心の注意を要する。
参照文	This work needs close attention .
先行研究	This task requires a meticulous attention .
提案手法	This task requires a religious attention to detail .
入力文	それは実質的には変わらない。
参照文	It is the same in substance .
先行研究	It is essentially the same .
提案手法	It is essentially a has not changed .
入力文	アメリカの大学は学校ごとにマスコットがある。
参照文	In America , there is a school mascot for every university .
先行研究	There is an American university mascot each school .
提案手法	There is an American university マスコット each school .
入力文	何千人ものデモ隊が街頭で抗議デモをした。
参照文	Thousands of demonstrators protested in the streets .
先行研究	The demonstrators thousands of protest demonstrated on the street .
提案手法	The demonstrators demonstrated the street lights of thousands of protest .

第8章 先行研究と追加実験の比較

本章では3章で述べた先行研究と追加実験の比較について述べる．比較はフレーズテーブルの量，自動評価，人手評価である．

8.1 フレーズテーブルの総数の比較

表 8.1 に先行研究と提案手法における，フレーズテーブルの総数と追加した翻訳対の数を示す．

表 8.1: フレーズテーブルの総数と追加した翻訳対の比較

	フレーズテーブルの総数	追加した翻訳対の数
従来手法	324,972	18,077
追加実験	394,730	87,835

表 8.1 から，追加実験は従来手法と比較して，約 4.8 倍の翻訳対を追加している．

8.2 翻訳精度の自動評価

テスト文に単文を用いて翻訳実験を行う．表 8.2 に単文の翻訳実験結果を示す．

表 8.2: 単文における翻訳精度の評価

単文	BLEU	NIST	METEOR
先行研究	0.1267	4.3047	0.3850
追加実験	0.1261	4.2965	0.3822

表 8.2 の結果から追加実験は先行研究より自動評価がよい．しかし，先行研究と比べると悪い．

8.3 先行研究と追加実験の人手評価の結果

表 8.3 に先行研究と追加実験における人手評価の結果を示す．なお，評価基準は 6.2 節で述べた基準である．

表 8.3: 先行研究と追加実験の対比較評価結果

	追加実験	追加実験 ×	差なし	同一出力
単文	4	0	5	91

表 8.3 の結果から追加実験は先行研究より人手評価がよいことから，追加実験は人手評価では有効である．

8.3.1 翻訳結果の例

先行研究と追加実験の翻訳結果の例を示す．表 8.4 に提案手法が優れていると判断した例を示す．

表 8.4: 単文において提案手法が優れていると判断した例

入力文	難局を收拾する道はそれしかない。
参照文	It is the only way to save the difficult situation .
先行研究	The situation 收拾 only the way it .
追加実験	The only way that save a difficult situation .
入力文	きょうはバスケットボールの練習がある。
参照文	Today we have basketball practice .
先行研究	There is a basketball for the today .
追加実験	There is a basketball practice today .
入力文	この国では情報は自由に流れる。
参照文	In this country data flows freely .
先行研究	Information flows in this country is at .
追加実験	Information flows in this country free .
入力文	アメリカやイギリスでは、第一次世界大戦のときに、燃料の節約を目的に導入された。
参照文	The United States and Britain introduced the system during World War I to conserve fuel .
先行研究	In the United States and Britain , at the end of World War was introduced to save and fuel .
追加実験	In the United States and Britain , at the end of World War I was introduced to save and fuel .

第9章 考察

この章では6章と7章,8章における比較の考察を行う。考察は提案手法において, 翻訳精度が向上もしくは低下した理由を, 翻訳に利用されたフレーズ対の観点から, 考察を行う。また, 先行研究と提案手法を比較し, 提案手法が劣っていた理由を, 追加した翻訳対の観点および, 部分マッチングで利用したペナルティの観点から考察する。

9.1 翻訳精度が向上した文の解析

この節では, 従来手法と比較して, 翻訳精度が向上した5文と先行研究と比較して, 翻訳精度が向上した3文とについて, 翻訳精度が向上した理由を解析した。また, 追加実験についても, 翻訳精度が向上した4文について同様の解析を行った。解析結果を表9.1に示す。

表 9.1: 翻訳精度向上の理由

翻訳精度向上の理由	提案手法 > 従来手法	提案手法 > 先行研究	追加実験 > 先行研究
適切なフレーズ対の増加	4	3	3
未知語の減少	1	0	1

9.1.1 適切なフレーズ対の増加

適切なフレーズ対が増加したため，提案手法の翻訳精度が向上した例を表 9.2 から表 9.11 に示す．

表 9.2: 提案手法が優れていると判断した例 1(従来手法との比較)

入力文	この 仕事 は 細心 の 注意 を 要する 。
参照文	This work needs close attention .
従来手法	This task requires a 細心 attention .
提案手法	This task requires a religious attention to detail .

使用されたフレーズ

先行研究	提案手法
この 仕事 This Task	この 仕事 This Task
要する requires	を 要する requires
は a	は a
細心 細心	細心 の 注意 religious attention to detail
の 注意 を	。 .
。 .	

この例では，従来手法において，“細心”が未知語として出力されている．一方で，提案手法では，“細心”は未知語として出力されていない．この原因は，従来手法のフレーズテーブルには“細心”に対するフレーズ対が存在しないことであった．一方で，提案手法では，DP マッチングにより，“細心の注意”，“religious attention to detail”という翻訳対をフレーズテーブルに追加できたため，未知語として出力されなかった．よって，この例では提案手法が有効であった．

表 9.3: 提案手法が優れていると判断した例 2(従来手法との比較)

入力文	きょうはバスケットボールの練習がある。
参照文	Today we have basketball practice .
従来手法	There is a basketball for the today .
提案手法	There is a basketball practice today .

使用されたフレーズ

先行研究	提案手法
がある There is は a バスケットボール basketball の練習 for the きょう today 。	がある There is a バスケットボールの練習 basketball practice きょうは today 。

この例では、従来手法において、“の練習”、“for the”という不適切なフレーズが使用されている。しかし、従来手法のフレーズテーブルには“の練習”、“practice”というフレーズ対が存在し、確率はこのフレーズ対の方が高い。しかし、言語モデルによるスコア付けやフレーズに対するペナルティにより、“の練習”、“for the”の方が高くなるため、従来手法では誤った英単語を出力している。一方で、提案手法では、DP マッチングにより、“バスケットボールの練習”、“basketball practice”という翻訳対をフレーズテーブルに追加でき、その確率が高かったため、言語モデルによるスコア付けでも高いスコアを保つことができ、誤った英単語を翻訳として出力しなかったと考えられる。よって、この例では提案手法が有効であった。

表 9.4: 提案手法が優れていると判断した例 3(従来手法との比較)

入力文	この国では情報は自由に流れる。
参照文	In this country data flows freely .
従来手法	Information flows in this country is at .
提案手法	Information flows in this country free .

使用されたフレーズ

先行研究	提案手法
情報は Infomation	情報は Infomation
流れる flows	流れる flows
この国では in this country	この国では in this country
自由に is at	自由に free
。 .	。 .

この例では、従来手法において、“自由に”、“is at”という不適切なフレーズが使用されている。しかし、従来手法のフレーズテーブルには“自由に”、“freely”というフレーズ対が存在し、確率はこのフレーズ対の方が高い。しかし、言語モデルによるスコア付けの際、“自由に”、“is at”の方が高くなるため、従来手法では誤った英単語を出力している。一方で、提案手法では、DP マッチングにより、“自由に”、“free”という翻訳対をフレーズテーブルに追加でき、その確率が高かったため、言語モデルによるスコア付けでも高いスコアを保つことができ、誤った英単語を翻訳として出力しなかったと考えられる。よって、この例では提案手法が有効であった。

表 9.5: 提案手法が優れていると判断した例 4(従来手法との比較)

入力文	アメリカやイギリスでは、第一次世界大戦のときに、燃料の節約を目的に導入された。
参照文	The United States and Britain introduced the system during World War I to conserve fuel .
従来手法	In the United States and Britain , at the end of World War was introduced to save and fuel .
提案手法	In the United States and Britain , at the end of World War I was introduced to save and fuel .

使用されたフレーズ

先行研究	提案手法
では In the アメリカやイギリス United States and Britain 、 , のときに at を the 目的 end の of 第一次世界大戦 World War に導入された was introduced to 節約 save 、燃料 and fuel 。 .	では In the アメリカやイギリス United States and Britain 、 , のときに at を the 目的 end の of 第一次世界大戦 World War I に導入された was introduced to 節約 save 、燃料 and fuel 。 .

この例では、従来手法において、“第一次世界大戦”、“World War”という適切なフレーズが使用されている。一方で、提案手法では、DP マッチングにより、“第一次世界大戦”、“World War I”というより適切な翻訳対をフレーズテーブルに追加できた、しかし、その確率は従来手法のフレーズ対に比べ低かった。しかし、言語モデルによるスコア付けにおいて、高いスコアを保つことができたため、提案手法の翻訳対を用いた翻訳が出力されたと考えられる。よって、この例では提案手法が有効であった。

表 9.6: 提案手法が優れていると判断した例 5(先行研究との比較)

入力文	きょうはバスケットボールの練習がある。
参照文	Today we have basketball practice .
先行研究	There is a basketball for the today .
提案手法	There is a basketball practice today .

使用されたフレーズ

先行研究	提案手法
がある There is は a バスケットボール basketball の練習 for the きょう today 。	がある There is a バスケットボールの練習 basketball practice きょうは today 。

この例では、先行研究において、“の練習”、“for the”という不適切なフレーズが使用されている。しかし、先行研究のフレーズテーブルには“の練習”、“plactice”というフレーズ対が存在し、確率はこのフレーズ対の方が高い。しかし、言語モデルによるスコア付けやフレーズに対するペナルティーにより、“の練習”、“for the”の方が高くなるため、先行研究では誤った英単語を出力している。一方で、提案手法では、DP マッチングにより、“バスケットボールの練習”、“basketball practice”という翻訳対をフレーズテーブルに追加でき、その確率が高かったため、言語モデルによるスコア付けでも高いスコアを保つことができ、誤った英単語を翻訳として出力しなかったと考えられる。よって、この例では提案手法が有効であった。

表 9.7: 提案手法が優れていると判断した例 6(先行研究との比較)

入力文	この国では情報は自由に流れる。
参照文	In this country data flows freely .
先行研究	Information flows in this country is at .
提案手法	Information flows in this country free .

使用されたフレーズ

先行研究	提案手法
情報は Information	情報は Information
流れる flows	流れる flows
この国では in this country	この国では in this country
自由に is at	自由に free
。 .	。

この例では、先行研究において、“自由に”、“is at”という不適切なフレーズが使用されている。しかし、先行研究のフレーズテーブルには“自由に”、“freely”というフレーズ対が存在し、確率はこのフレーズ対の方が高い。しかし、言語モデルによるスコア付けの際、“自由に”、“is at”の方が高くなるため、先行研究では誤った英単語を出力している。一方で、提案手法では、DP マッチングにより、“自由に”、“free”という翻訳対をフレーズテーブルに追加でき、その確率が高かったため、言語モデルによるスコア付けでも高いスコアを保つことができ、誤った英単語を翻訳として出力しなかったと考えられる。よって、この例では提案手法が有効であった。

表 9.8: 提案手法が優れていると判断した例 7(先行研究との比較)

入力文	アメリカ や イギリス では、 第 一 次 世界 大 戦 の と き に、 燃 料 の 節 約 を 目 的 に 導 入 さ れ た 。
参照文	The United States and Britain introduced the system during World War I to conserve fuel .
従来手法	In the United States and Britain , at the end of World War was introduced to save and fuel .
提案手法	In the United States and Britain , at the end of World War I was introduced to save and fuel .

使用されたフレーズ

先行研究	提案手法
では In the アメリカ や イギリス United States and Britain 、 , の と き に at を the 目的 end の of 第 一 次 世界 大 戦 World War に 導 入 さ れ た was introduced to 節 約 save 、 燃 料 and fuel 。 .	では In the アメリカ や イギリス United States and Britain 、 , の と き に at を the 目的 end の of 第 一 次 世界 大 戦 World War I に 導 入 さ れ た was introduced to 節 約 save 、 燃 料 and fuel 。 .

この例では、先行研究において、“第一次世界大戦”、“World War”という適切なフレーズが使用されている。一方で、提案手法では、DP マッチングにより、“第一次世界大戦”、“World War I”というより適切な翻訳対をフレーズテーブルに追加できた、しかし、その確率は先行研究のフレーズ対に比べ低かった。しかし、言語モデルによるスコア付けにおいて、高いスコアを保つことができたため、提案手法の翻訳対を用いた翻訳が出力されたと考えられる。よって、この例では提案手法が有効であった。

表 9.9: 追加実験が優れていると判断した例 8(追加実験と先行研究の比較)

入力文	きょうはバスケットボールの練習がある。
参照文	Today we have basketball practice .
先行研究	There is a basketball for the today .
追加実験	There is a basketball practice today .

使用されたフレーズ

先行研究	提案手法
がある There is は a バスケットボール basketball の練習 for the きょう today 。	がある There is a バスケットボールの練習 basketball practice きょうは today 。

この例では、先行研究において、“の練習”、“for the”という不適切なフレーズが使用されている。しかし、先行研究のフレーズテーブルには“の練習”、“plactice”というフレーズ対が存在し、確率はこのフレーズ対の方が高い。しかし、言語モデルによるスコア付けやフレーズに対するペナルティーにより、“の練習”、“for the”の方が高くなるため、先行研究では誤った英単語を出力している。一方で、追加実験では、DP マッチングにより、“バスケットボールの練習”、“basketball practice”という翻訳対をフレーズテーブルに追加でき、その確率が高かったため、言語モデルによるスコア付けでも高いスコアを保つことができ、誤った英単語を翻訳として出力しなかったと考えられる。よって、この例では追加実験が有効であった。

表 9.10: 追加実験が優れていると判断した例 9(追加実験と先行研究の比較)

入力文	この国では情報は自由に流れる。
参照文	In this country data flows freely .
先行研究	Information flows in this country is at .
追加実験	Information flows in this country free .

使用されたフレーズ

先行研究	提案手法
情報は Infomation	情報は Infomation
流れる flows	流れる flows
この国では in this country	この国では in this country
自由に is at	自由に free
。 .	。 .

この例では、先行研究において、“自由に”、“is at”という不適切なフレーズが使用されている。しかし、先行研究のフレーズテーブルには“自由に”、“freely”というフレーズ対が存在し、確率はこのフレーズ対の方が高い。しかし、言語モデルによるスコア付けの際、“自由に”、“is at”の方が高くなるため、先行研究では誤った英単語を出力している。一方で、追加実験では、DP マッチングにより、“自由に”、“free”という翻訳対をフレーズテーブルに追加でき、その確率が高かったため、言語モデルによるスコア付けでも高いスコアを保つことができ、誤った英単語を翻訳として出力しなかったと考えられる。よって、この例では追加実験が有効であった。

表 9.11: 追加実験が優れていると判断した例 10(追加実験と先行研究の比較)

入力文	アメリカ や イギリス では、 第 一 次 世界 大 戦 の と き に、 燃 料 の 節 約 を 目 的 に 導 入 さ れ た 。
参照文	The United States and Britain introduced the system during World War I to conserve fuel .
先行研究	In the United States and Britain , at the end of World War was introduced to save and fuel .
追加実験	In the United States and Britain , at the end of World War I was introduced to save and fuel .

使用されたフレーズ

先行研究	提案手法
では In the アメリカ や イギリス United States and Britain 、 , の と き に at を the 目的 end の of 第 一 次 世界 大 戦 World War に 導 入 さ れ た was introduced to 節 約 save 、 燃 料 and fuel 。 .	では In the アメリカ や イギリス United States and Britain 、 , の と き に at を the 目的 end の of 第 一 次 世界 大 戦 World War I に 導 入 さ れ た was introduced to 節 約 save 、 燃 料 and fuel 。 .

この例では、先行研究において、“第一次世界大戦”、“World War”という適切なフレーズが使用されている。一方で、追加実験では、DP マッチングにより、“第一次世界大戦”、“World War I”というより適切な翻訳対をフレーズテーブルに追加できた、しかし、その確率は先行研究のフレーズ対に比べ低かった。しかし、言語モデルによるスコア付けにおいて、高いスコアを保つことができたため、追加実験の翻訳対を用いた翻訳が出力されたと考えられる。よって、この例では追加実験が有効であった。

9.1.2 未知語の減少

未知語が減少したため，提案手法の翻訳精度が向上した例を表 9.12 と表 9.13 に示す．

表 9.12: 提案手法が優れていると判断した例 11(従来手法との比較)

入力文	難局 を 収拾 する 道 は それ しか ない 。
参照文	It is the only way to save the difficult situation .
従来手法	難局 収拾 only the way it .
提案手法	The only way that save a difficult situation .

使用されたフレーズ

先行研究	提案手法
難局 難局	は The
収拾 収拾	しか ない only
しか ない only	道 way
を the	それ that
する 道 way	難局 を 収拾 する save a difficult situation
は それ it	。 .
。 .	

この例では，従来手法において，“収拾”が未知語として出力されている．一方で，提案手法では，“収拾”は未知語として出力されていない．この原因は，従来手法のフレーズテーブルには“収拾”に対するフレーズ対が存在しないことであった．一方で，提案手法では，DP マッチングにより，“難局を収拾する”，“save a difficult situation”という翻訳対をフレーズテーブルに追加できたため，未知語として出力されなかった．よって，この例では提案手法が有効であった．

表 9.13: 追加実験が優れていると判断した例 12(追加実験と先行研究の比較)

入力文	難局 を 収拾 する 道 は それ しか ない 。
参照文	It is the only way to save the difficult situation .
先行研究	The situation 収拾 only the way it .
追加実験	The only way that save a difficult situation .

使用されたフレーズ	
先行研究	提案手法
は The	は The
難局 situation	しか ない only
収拾 収拾	道 way
しか ない only	それ that
を the	難局 を 収拾 する save a difficult situation
する 道 way	。 .
それ it	
。 .	

この例では，先行研究において，“収拾”が未知語として出力されている．一方で，追加実験では，“収拾”は未知語として出力されていない．この原因は，先行研究のフレーズテーブルには“収拾”に対するフレーズ対が存在しないことであった．一方で，追加実験では，DP マッチングにより，“難局を収拾する”，“save a difficult situation”という翻訳対をフレーズテーブルに追加できたため，未知語として出力されなかった．よって，この例では追加実験が有効であった．

9.2 翻訳精度が低下した文の解析

この節では，先行研究と比較して，翻訳精度が低下した3文について，翻訳精度が低下した理由を解析した．解析結果を表 9.14 に示す．

表 9.14: 翻訳精度低下の理由

翻訳精度低下の理由	提案手法 < 先行研究
不適切なフレーズ対の選択	2
未知語の影響	2
マッチングの対象の問題	1

9.2.1 不適切なフレーズ対の選択

不適切なフレーズ対が増加したため，提案手法の翻訳精度が低下した例を表 9.15 と表 9.16 に示す．

表 9.15: 提案手法が劣っていると判断した例 1(先行研究との比較)

入力文	それは実質的には変わらない。
参照文	It is the same in substance .
先行研究	It is essentially the same .
提案手法	It is essentially a has not changed .

使用されたフレーズ

先行研究	提案手法
それは It	それは It
的 is	的 is
実質 essentially	実質 essentially
には the	には a
変わらない same	変わらない has not changed .
。 .	

この例では，提案手法において，“変わらない”，“has not changed” のフレーズ対が不適切だといえる．しかし，このフレーズ対は提案手法によってフレーズテーブルに追加したものではない．このフレーズ対が選択された理由としては，提案手法により導入した翻訳対がフレーズテーブルに追加されることで，デコーディングの際に，本来必要なパスが枝刈りにより，カットされることで，不適切なフレーズ対が選択された翻訳が出力されたと考える．

表 9.16: 提案手法が劣っていると判断した例 2(先行研究との比較)

入力文	何千人ものデモ隊が街頭で抗議デモをした。
参照文	Thousands of demonstrators protested in the streets .
先行研究	The demonstrators thousands of protest demonstrated on the street .
提案手法	The demonstrators demonstrated the street lights of thousands of protest .

使用されたフレーズ

先行研究	提案手法
デモ隊が The demonstrators	デモ隊が The demonstrators
何千人も thousands	デモをした demonstrated
の of	で the
抗議 protest	街頭 street lights
デモをした demonstrated	何千人もの of thousands of
街頭で on the street	抗議 protest
。 .	。 .

この例では、提案手法において、“街頭”、“street lights”と“何千人もの”、“of thousands of”のフレーズ対が不適切だといえる。しかし、このフレーズ対は提案手法によってフレーズテーブルに追加したものではない。このフレーズ対が選択された理由としては、提案手法により導入した翻訳対がフレーズテーブルに追加されることで、デコーディングの際に、本来必要なパスが枝刈りにより、カットされることで、不適切なフレーズ対が選択された翻訳が出力されたと考える。

9.2.2 未知語の影響

未知語と対応したフレーズ対が選ばれているため，提案手法の翻訳精度が向上した例を表 9.17 と表 9.18 に示す．

表 9.17: 提案手法が劣っていると判断した例 3(先行研究との比較)

入力文	ポルシェ が その クラス で 優勝 した 。
参照文	The Porsche won its class .
先行研究	The Porsche won the championship in the class
提案手法	The ポルシェ won the championship in the class .

使用されたフレーズ

先行研究	提案手法
が The	が The
ポルシェ Porsche	ポルシェ ポルシェ
優勝 した won the championship	優勝 した won the championshi
で in	で in
その クラス the class	その クラス the class
。 .	。 .

この例では，先行研究において，“ポルシェ”が“Porsche”として出力されている．一方で，提案手法では，“ポルシェ”は未知語として出力されている．この原因は，提案手法のフレーズテーブルには“ポルシェ”に対するフレーズ対が存在しないことであった．一方で，提案手法では，intersection で作成したフレーズテーブルとのマッチングにより，“ポルシェ”，“Porsche”という翻訳対をフレーズテーブルに追加できたため，未知語として出力されなかった．よって，この例では提案手法が有効ではない．

表 9.18: 提案手法が劣っていると判断した例 4(先行研究との比較)

入力文	アメリカの大学は学校ごとにマスコットがある。
参照文	In America , there is a school mascot for every university .
先行研究	There is an American university mascot each school .
提案手法	There is an American university マスコット each school .

使用されたフレーズ

先行研究	提案手法
がある There is	がある There is
は an	は an
アメリカの大学 American university	アメリカの大学 American university
マスコット mascot	マスコット マスコット
ごとに each	ごとに each
学校 school	学校 school
。 .	。 .

この例では、先行研究において、“マスコット”が“mascot”として出力されている。一方で、提案手法では、“マスコット”は未知語として出力されている。この原因は、提案手法のフレーズテーブルには“マスコット”に対するフレーズ対が存在しないことであった。一方で、提案手法では、intersection で作成したフレーズテーブルとのマッチングにより、“マスコット”、“mascot”という翻訳対をフレーズテーブルに追加できたため、未知語として出力されなかった。よって、この例では提案手法が有効ではない。

9.2.3 マッチングの対象の問題

先行研究における，マッチング対象のフレーズテーブルにはフレーズ対があるが，提案手法における，マッチング対象のフレーズテーブルにはフレーズ対がないため，翻訳対に翻訳確率を付与することができないため，翻訳対を追加することができず，提案手法の翻訳精度が低下した例を表 9.19 に示す．

表 9.19: 提案手法が劣っていると判断した例 5(先行研究との比較)

入力文	この 仕事 は 細心 の 注意 を 要する 。
参照文	This work needs close attention .
先行研究	This task requires a meticulous attention .
提案手法	This task requires a religious attention to detail .

使用されたフレーズ	
先行研究	提案手法
この 仕事 This Task	この 仕事 This Task
を 要する requires	を 要する requires
は a	は a
細心 の meticulous	細心 の 注意 religious attention to detail
注意 を attention	。 .
。 .	

この例では，先行研究，提案手法ともに意味が通る文である．しかし，入力文により適した文は先行研究だと判断した．提案手法において，“細心の”，“meticulous”の翻訳対はフレーズテーブルに存在しない．その理由としては，DP マッチングの対象が grow-diag-final で作成したフレーズテーブルのため，作成されるフレーズ対の数が intersection で作成したフレーズテーブルと比較して，非常に小さいためだと考える．よって，intersection で作成したフレーズテーブルで DP マッチングを行えば，この翻訳対に対して確率を付与できるが，DP マッチングが有限時間内に終わらないため，現実的ではない．

9.3 翻訳候補文に使用された翻訳対の違い

提案手法では、69,728 文対の翻訳対に対して翻訳確率の付与を行い、先行研究では、18,077 文対の翻訳対に対して翻訳確率の付与を行った。先行研究と比較すると自動評価、人手評価共に評価が低いという結果になった。その原因は、統計翻訳において、翻訳候補として使用される翻訳対の数の違いだと考える。翻訳実験において、使用される翻訳対の数を調べたところ、先行研究では追加した翻訳対 18,077 文対中、翻訳候補文に使用された数は 2,481 文対であった。一方、提案手法は追加した翻訳対 69,728 文対中、翻訳候補文使用された数は 1,067 文対であり、先行研究の半数以下の翻訳対しか使用されていない。

この原因として、ペナルティにより、本来使用されるべき翻訳対を含んだ翻訳候補のパスがデコーディングの際に枝刈りされ、悪影響を及ぼしたと考えられる。

9.4 ペナルティの影響

表 9.20 に提案手法における、ペナルティの値を変化させたときの自動評価の結果を示す。

表 9.20: ペナルティによる自動評価の変化

ペナルティ	BLEU	NIST	METEOR
元のスコア	0.1215	4.1668	0.3735
× 1/4	0.1222	4.1773	0.3749
× 1/2	0.1222	4.1799	0.3749
× 2	0.1218	4.1671	0.3732
× 4	0.1229	4.1821	0.3736

表 9.20 において、ペナルティを変化させると、BLEU スコアで 0.004%、NIST スコアで 0.128%、METEOR スコアで 0.017%の変動がある。よって、本研究において与えるペナルティが適切でなかったと考える。今後は単語によるペナルティだけではなく、フレーズや頻度も考慮に入れた適切なペナルティを考えていきたい。

9.5 未知語の影響

表 9.21 にそれぞれの手法における未知語の総数と BLEU スコアを示す。

表 9.21: 未知語の総数

	未知語	BLEU	NIST	METEOR
従来手法	631	0.1189	4.1477	0.3730
先行研究	523	0.1267	4.3027	0.3850
提案手法	620	0.1215	4.1668	0.3735
追加実験	519	0.1261	4.2965	0.3822

表 9.21 から、未知語が減少するほど、METEOR スコアは向上している。先行研究と提案手法の未知語を比較すると、約 100 語の違いがある。よって、先行研究と提案手法の BLEU スコアの差は未知語が大きな原因だと考える。一方、先行研究と追加実験の未知語を比較すると、追加実験は僅かに未知語が少ない。しかし、BLEU スコアは僅かに先行研究の方がよい。その原因は、不適切な翻訳確率が付与された翻訳対がデコーディングの際、本来必要なパスの枝刈りを行い、悪影響を及ぼしたことだと考えている。

第10章 おわりに

本研究では、部分的に一致したフレーズを統計翻訳に導入することにより翻訳精度の向上を試みた。その結果、提案手法と従来手法を比較すると、自動評価、人手評価共に評価が高い。しかし、提案手法と先行研究を比較すると、自動評価、人手評価共に評価が低い。よって、提案手法はあまり効果がないことがわかった。その原因の1つは、翻訳候補に使用されるフレーズが少ないことだと考えている。また、提案手法において、ペナルティを変化させるとBLEUスコアで0.004%、NISTスコアで0.128%、METEORスコアで0.017%変動した。よって、本研究で与えるペナルティが適切ではなかったと考えている。今後は単語によるペナルティだけでなく、フレーズや頻度も考慮に入れた適切なペナルティを考えていきたい。

謝辞

最後に、3年間に渡り御指導いただきました鳥取大学大学院工学研究科情報エレクトロニクス専攻計算機工学講座Cの村上仁一准教授、徳久雅人講師、2年間に渡り御指導いただきました鳥取大学大学院工学研究科情報エレクトロニクス専攻計算機工学講座Cの村田真樹教授に厚く御礼申し上げます。また、参考にさせて頂いた論文の著者の方々に対して深く感謝申し上げます。

参考文献

- [1] 鏡味良太, 村上仁一, 徳久雅人, 池原悟, “統計翻訳における人手で作成された大規模フレーズテーブルの効果”, 言語処理学会第 15 回年次大会 pp.224-227, (2009).
- [2] Richard Zens, Franz Josef Och, and Hermann Ney, “Phrase-Based Statistical Machine Translation”, Springer-Verlag Berlin Heidelberg, pp.18-32, (2002).
- [3] <http://www.statmt.org/moses/giza/GIZA++.html>
- [4] Peter Brwon, Stephen Della Pietra, Vincent Della Pietra, Robert mercer, “The matematics of statistical machine translation, Parameter estimation”, Computational Linguistics, 263-311, (1993).
- [5] Tillmann Christoph and Zhang Tong, “A Localized Prediction Model for Statistical Machine Translation”, Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, pp.557-564, (2005).
- [6] Philipp Koehn, Franz Josef Och, Daniel Marcu, “Statistical Phrase-Based Translation”, Association for Computational Linguistic, pp.48-54, (2003).
- [7] Philipp Koehn, Marcello Federico, Brooke Cowan, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, Evan Herbst, “Moses: Open Source Toolkit for Statistical Machine Translation”, Association for Computational Linguistic, pp.177-180, (2007).
- [8] <http://www.alc.co.jp/>
- [9] 池原悟, 阿部さつき, 徳久雅人, 村上仁一, “非線形な表現構造に着目した日英文型パターン化”, 情報処理学会研究報告自然言語処理, pp.49-56, (2004).
- [10] The SRI Language Model Toolkit
<http://www.speech.sri.com/projects/srilm>

- [11] <http://mecab.sourceforge.net/>
- [12] Kishore Papineni, Salim Rukos, Todd Ward, and Wei-Jing Zhu, “BLEU: a Method for Automatic Evaluation of Machine Translation” , Association for Computational Linguistics, pp.311-318, (2002).
- [13] George Doddington, “Automatic evaluation of machine translation quality using n-gram co-occurrence statistics”, Proceedings of the second international conference on Human Language Technology Research, pp.138-145, (2002).
- [14] Satanjeev Banerjee and Alon Lavi, “METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments”, Association for Computational Linguistics, pp.65-72, (2005).
- [15] Franz Josef Och, “Minimum Error Rate Training in Statistical Machine Translation”, Association for Computational Linguistics, pp.160-167, (2003).

第11章 付録

付録では，重文複文での実験結果を載せる．

11.1 実験データ

11.1.1 学習データ

学習データには，辞書から抽出した，重文複文 100,000 文対を使用する．

11.1.2 テストデータ

テストデータには，辞書から抽出した，重文複文 1,000 文を使用する．

11.2 フレーズテーブルの総数の比較

表 11.1 に従来手法と先行研究と提案手法における，フレーズテーブルの総数と追加した翻訳対の数を示す．

表 11.1: フレーズテーブルの総数と追加した翻訳対の比較

	フレーズテーブルの総数	追加した翻訳対の数
従来手法	239,950	0
先行研究	257,345	17,395
提案手法	323,714	83,764

11.3 翻訳精度の自動評価

テスト文に重文複文を用いて翻訳実験を行う．表 11.2 に翻訳実験結果に示す．なお，重文複文に関しては，人手による対比較評価は行っていない．(時間がなかったため)

表 11.2: 重文複文における翻訳精度の評価

重文複文	BLEU	NIST	METEOR
従来手法	0.0857	3.5368	0.3270
先行研究	0.0881	3.6233	0.3344
提案手法	0.0848	3.5215	0.3259

表 11.2 の結果から提案手法は最も自動評価が低い。よって、重文複文において、提案手法は有効ではない。

11.4 使用された翻訳対の違い

表 11.3 に従来手法と先行研究と提案手法における、フレーズテーブルの総数と追加した翻訳対の数を示す。

表 11.3: テスト文で使用された翻訳対の比較

	追加した翻訳対の数	テスト文で使用された翻訳対の数
先行研究	17,395	2899(16.67%)
提案手法	83,764	1596(1.91%)

表 11.3 の結果から、提案手法は僅か 1.91% の翻訳対しか使用されておらず、かつ、テスト文で使用された翻訳対が悪影響を及ぼしたため、翻訳精度が低下したと考える。