

概要

日英統計翻訳の困難な点の1つとして、日本語と英語の語順が異なることが挙げられる。この問題を解決するために、日本語文の単語を並び替えて、英語文の語順に近づけてから、統計翻訳をする研究が盛んに行われている [1][2]。

岡崎ら [3] によると、日本語文の主語、目的語、動詞 (SOV) を主語、動詞、目的語 (SVO) の順に語順変更を行うだけでは翻訳精度は向上しなかった。一方、星野ら [4] によると、述語項構造に基づく語順変更を行ってから、句に基づく統計翻訳を行うことで、翻訳精度が向上した。しかし、これらの研究は翻訳される英語文の語順を推定して、日本語文の語順変更を行っている。そのため、目標とする英語文の最適な語順に並び替えられていない。

そこで、本研究では語順変更の手法の限界を調査するため、テスト文に参照文があると仮定し、日本語文を英語の参照文の語順に並び替える。その後、句に基づく統計翻訳を行い、語順変更の効果を調査した。

実験の結果、日本語文を英語文の語順に並び替えてから、句に基づく統計翻訳を行っても、人手評価で差は見られなかった。このことから、語順変更では翻訳精度の大きな向上が見られないと考える。

目次

第1章	はじめに	1
第2章	統計翻訳システム	2
2.1	統計翻訳の概要	2
2.2	日英統計翻訳における基本モデル	2
2.3	翻訳モデル	3
2.3.1	翻訳モデルの概要	3
2.3.2	IBM 翻訳モデル	4
2.3.3	Model1	5
2.3.4	Model2	6
2.3.5	Model3	7
2.3.6	Model4	8
2.3.7	Model5	9
2.3.8	GIZA++	9
2.4	言語モデル	9
2.4.1	言語モデルの概要	9
2.4.2	N -gram モデル	10
2.4.3	スムージング	11
2.5	reordering-model	12
2.6	デコーダー	12
2.7	評価システム	13
2.8	パラメータチューニング	13
第3章	日本語文の語順変更による翻訳	14
3.1	先行研究	14
3.1.1	動詞の語順変更 [3]	14
3.1.2	述語項構造に基づく語順変更 [4]	15

3.1.3	語順変更および句語順変更を用いた手法 [11]	16
3.2	提案手法	18
第4章	実験環境	21
4.1	実験データ	21
4.1.1	単文コーパス	21
4.1.2	重文複文コーパス	22
4.2	対訳句辞書	22
4.3	翻訳モデルの学習	25
4.4	言語モデルの学習	25
4.5	デコーダーに関するパラメータ	25
4.6	ベースライン	26
4.7	提案手法	26
4.8	評価方法	26
第5章	翻訳実験	28
5.1	実験内容	28
5.2	単文の実験	28
5.2.1	語順変更の精度	28
5.2.2	自動評価	30
5.2.3	人手評価	31
5.2.4	翻訳例	31
5.3	重文複文の実験	32
5.3.1	語順変更の精度	32
5.3.2	自動評価	34
5.3.3	人手評価	35
5.3.4	翻訳例	35
5.4	実験のまとめ	36
第6章	考察	38
6.1	先行研究との比較	38
6.2	自動評価と人手評価	38
6.2.1	実験内容	38

6.2.2	実験のまとめ	40
6.3	語順変更の有効性	40
6.3.1	実験内容	40
6.3.2	単文の実験	40
6.3.3	重文複文の実験	43
6.3.4	実験のまとめ	44
6.4	“reordering-model” の効果	44
6.4.1	実験内容	44
6.4.2	単文の実験	45
6.4.3	重文複文の実験	46
6.4.4	実験のまとめ	47
6.5	日本語文の語順をランダムに変えた実験	47
6.5.1	実験内容	47
6.5.2	検証手法	47
6.5.3	日本変更文の例	49
6.5.4	単文の実験	50
6.5.5	重文複文の実験	52
6.5.6	実験のまとめ	54
6.6	鳥バンクのパターンを使用した実験	54
6.6.1	実験内容	54
6.6.2	人手作成パターンの例	55
6.6.3	実験結果	55
6.6.4	実験のまとめ	57
6.7	“distortion-limit” の値を変えた実験	58
6.7.1	実験内容	58
6.7.2	単文の実験	58
6.7.3	重文複文の実験	59
6.7.4	実験のまとめ	59
6.8	語順変更の研究の今後	59
第7章 おわりに		60

目 次

2.1	日英統計翻訳の枠組	3
2.2	アライメントの例	4
2.3	デコーダーにおける翻訳候補の生成例	12
2.4	評価システム例	13
3.1	動詞の語順変更の例	14
3.2	述語項構造に基づく語順変更の例	15
3.3	動詞の語順変更の例	16
3.4	目的語の語順変更の例	16
3.5	助詞の語順変更の例	17
3.6	動詞+目的語+助詞の語順変更の例	17
3.7	変数化	19
3.8	語順変更	20
6.1	変数化	48
6.2	ランダムに並び替え	48

表 目 次

3.1	自動評価結果	14
3.2	対比較評価結果	15
3.3	自動評価結果	16
3.4	自動評価結果	18
3.5	対比較評価結果	18
4.1	実験データの文数と平均単語数	21
4.2	削除する対象となる単語	23
4.3	間違っただ対応を取ると判断する基準	23
4.4	対訳句数	23
4.5	鳥バンクの対訳句の精度	24
5.1	単文の語順変更の精度	28
5.2	単文の語順変更の成功例	29
5.3	単文の語順変更の失敗例	30
5.4	単文の実験の自動評価結果	31
5.5	単文の実験の対比較評価結果	31
5.6	提案手法の方が優れている翻訳例	31
5.7	ベースラインの方が優れている翻訳例	32
5.8	重文複文の語順変更の精度	32
5.9	重文複文の語順変更の成功例	33
5.10	重文複文の語順変更の失敗例	34
5.11	重文複文の実験の自動評価結果	35
5.12	重文複文の実験の対比較評価結果	35
5.13	提案手法の方が優れている翻訳例	35
5.14	ベースラインの方が優れている翻訳例	36

6.1	スコアが高いが不自然な翻訳文の例	39
6.2	スコアは低い正しい翻訳文の例	39
6.3	語順変更成功した 90 文の対比較評価結果	40
6.4	提案手法の方が優れている翻訳例	41
6.5	ベースラインの方が優れている翻訳例	42
6.6	語順変更失敗した 10 文の対比較評価結果	42
6.7	ベースラインの方が優れている翻訳例	42
6.8	語順変更成功した 87 文の対比較評価結果	43
6.9	提案手法の方が優れている翻訳例	43
6.10	ベースラインの方が優れている翻訳例	43
6.11	語順変更失敗した 13 文の対比較評価結果	44
6.12	提案手法の方が優れている翻訳例	44
6.13	単文の実験の自動評価結果	45
6.14	単文の実験の対比較評価結果	45
6.15	“reordering-model あり”の方が優れている翻訳例	45
6.16	“reordering-model なし”の方が優れている翻訳例	46
6.17	重文複文の実験の自動評価結果	46
6.18	重文複文の実験の対比較評価結果	46
6.19	“reordering-model あり”の方が優れている翻訳例	46
6.20	“reordering-model なし”の方が優れている翻訳例	47
6.21	単文の日本語変更文の例	49
6.22	重文複文の日本語変更文の例	50
6.23	単文の実験の自動評価結果	50
6.24	検証手法とベースラインの対比較評価結果	51
6.25	検証手法の方が優れている翻訳例	51
6.26	ベースラインの方が優れている翻訳例	52
6.27	重文複文の実験の自動評価結果	52
6.28	重文複文の実験の対比較評価結果	53
6.29	検証手法の方が優れている翻訳例	53
6.30	ベースラインの方が優れている翻訳例	54
6.31	人手作成パターン例	55
6.32	対比較評価結果	55

6.33 自動作成パターンの方が優れている例	56
6.34 人手作成パターンの方が優れている例	57
6.35 単文の自動評価結果	58
6.36 自動評価結果	59

第1章 はじめに

機械翻訳は、人手で対訳コーパスや文法のルールを作成するルールベース翻訳がはじめである。しかし、ルールベース翻訳は莫大な時間と労力がかかるという問題点がある。この問題点を改善するために、語に基づく統計翻訳が提案されたが、莫大なデータ量や計算機のパワーが必要であるため、あまり発展しなかった。しかし、句に基づく統計翻訳が提案されたときは、語に基づく統計翻訳の時代と比較して、計算機のパワーが大きく改良されていた。また、句に基づく統計翻訳は、語に基づく統計翻訳と比較して、翻訳精度が高いため、統計翻訳が急速に広がった。

一般的に、統計翻訳では、イタリア語から英語へ翻訳する場合、語彙や語順が似ているため、翻訳精度が高い傾向にある。しかし、日本語から英語へ翻訳する場合、語彙や語順が異なるため、翻訳精度が低い傾向にある。この問題を解決するために、日本語文の単語を並び替えて、英語の語順に近づけてから、統計翻訳を行う研究がされている。

岡崎ら [3] によると、日本語文の主語、目的語、動詞 (SOV) を主語、動詞、目的語 (SVO) の順に語順変更を行うだけでは翻訳精度は向上しなかった。一方、星野ら [4] によると、述語項構造に基づく語順変更を行ってから、句に基づく統計翻訳を行うことで、翻訳精度が向上した。しかし、これらの研究は翻訳される英語文の語順を推定して、日本語文の語順変更を行っている。そのため、目標とする英語文の最適な語順に並び替えられていない。日本語文を英語文の最適な語順に並び替えることで、翻訳精度が向上すると考えた。

そこで、本研究では語順変更の手法の限界を調査するため、テスト文に参照文があると仮定し、日本語文を英語の参照文の語順に並び替える。その後、句に基づく統計翻訳を行い、語順変更の効果を調査する。

第2章 統計翻訳システム

2.1 統計翻訳の概要

統計翻訳において、「単語に基づく統計翻訳」と「句に基づく統計翻訳」がある。初期の統計翻訳は、単語に基づく統計翻訳であった。しかし、近年提案された句に基づく統計翻訳は、語順の並び替えや文脈における訳語の選択や翻訳精度において、単語に基づく統計翻訳よりも優れている。よって、現在は句に基づく統計翻訳が主流となっている。そのため、本研究で扱う統計翻訳システムにおいても、句に基づく統計翻訳を用いる。本章では、句に基づく統計翻訳について説明する。

2.2 日英統計翻訳における基本モデル

日英統計翻訳は、翻訳元言語の日本語文 j が与えられた場合、全ての組合せの中から、確率値が最大となる翻訳先言語の英語文 \hat{e} を探索することにより翻訳を行う。以下にその基本モデルを示す。

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

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

ここで、 $P(j|e)$ は翻訳モデル、 $P(e)$ は言語モデルを表す。翻訳モデルは対訳学習文から学習し、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。

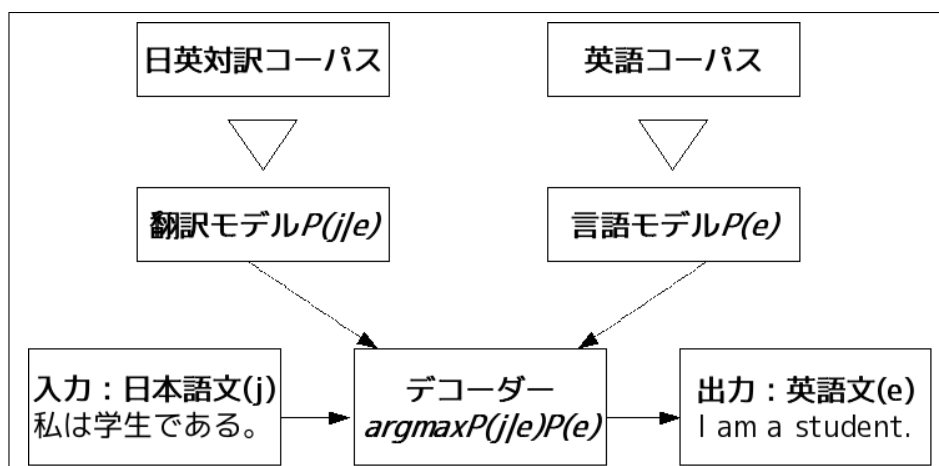


図 2.1: 日英統計翻訳の枠組

2.3 翻訳モデル

2.3.1 翻訳モデルの概要

翻訳モデルは、日本語の単語列から英語の単語列へと確率的に翻訳を行うためのモデルである。句に基づく翻訳モデルは、IBM 翻訳モデル [6] を用いた GIZA++ [7] から生成され、以下に示すフレーズテーブルという句の表により管理される。

フレーズテーブルの例

```

この ||| This ||| 0.666405 0.717667 0.427636 0.414483 2.718
この ||| This is ||| 0.0350318 0.717667 0.00184688 0.00875099 2.718
この ||| This is the ||| 0.0338983 0.717667 0.000335796 0.000587052 2.718
この 本 ||| This book ||| 0.623288 0.585559 0.535294 0.219275 2.718
この 本 ||| This book is ||| 0.235294 0.585559 0.0235294 0.00462955 2.718
...
```

左から、日本語フレーズ、英語フレーズ、そして、フレーズ対応の確率値5つである。確率値は、左から、英日方向の翻訳確率、英日方向の単語の翻訳確率の積 (IBM 翻訳モデル)、日英方向の翻訳確率、日英方向の単語の翻訳確率の積 (IBM 翻訳モデル)、そして、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。

フレーズテーブルを生成する際、フレーズ内の単語の数を制限するものとして、`max-phrase-length` (Maximum Phrase Length) が定義されている。例として、`max-phrase-length` が 10 の場合、日英方向または英日方向の単語列の長さが 11 以上のフレーズは生成され

ない. max-phrase-length でフレーズ内の単語を制限することで, 翻訳精度が向上する場合がある.

2.3.2 IBM 翻訳モデル

統計翻訳で代表的なモデルとして, Brown らが提案したフランス語英語翻訳モデルがある. この翻訳モデルは, 提案した Brown ら全員が IBM の社員であったため, 通称 IBM 翻訳モデル (IBM Model) といわれる. IBM 翻訳モデルは, Model1 から Model5 まで 5 つのモデルからなり, 順に複雑な計算となる. 第 2.3.2 章の説明では, 翻訳する言語をフランス語 f , 目的言語を英語 e と想定して行う.

IBM 翻訳モデルでは, 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 α を用いる. (2.3) に IBM 翻訳モデルの基本的な計算式を示す.

$$P(f|e) = \sum_a P(f, a|e) \quad (2.3)$$

アライメント α はフランス単語 f と英単語 e における対応関係を示す. IBM 翻訳モデルにおけるアライメントは, フランス語側の各単語 f において, それぞれ英語側に対応する単語 e が一つある. また, 英語側の各単語 e は, フランス語側に対応する単語 f が $0 \sim n$ 個ある. また, フランス単語 f において, 適切な英単語 e との対応関係がなかった場合, 英語文の先頭に e_0 という空単語があると仮定し, そのフランス語と空単語 e_0 を対応づける. フランス語英語翻訳におけるアライメントの例を図 2.2 に示す.

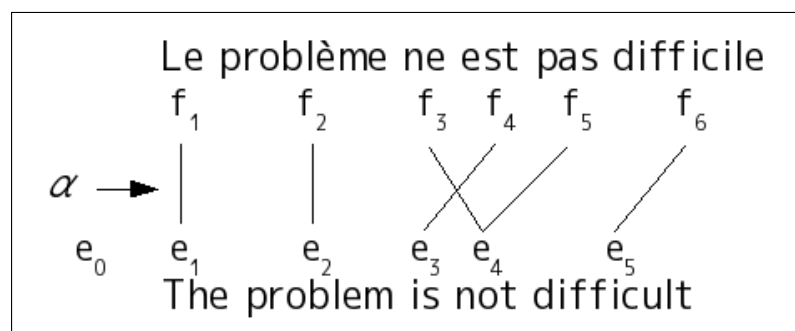


図 2.2: アライメントの例

α は $\alpha_1^n = \alpha_1, \alpha_2, \dots, \alpha_n$ である. また, α_1 は, f_1 の単語 Le と $e_{\alpha_1=1}$ の単語 The にお

る対応関係を表す.

2.3.3 Model1

(2.3) 式は次のように書き換えることができる. このとき, $f = f_1^m$ はフランス語文において 1 単語目から m 単語まで, $e = e_1^l$ は英語文において 1 単語目から l 単語目まで, $a = a_1^m$ はフランス語文において 1 単語目から m 単語目までのアライメントを表す.

$$P(f|e) = P(m|e) \prod_{j=1}^m P(a_j|a_1^{j-1}, f_1^{j-1}, m, e) P(f_j|a_1^j, f_1^{j-1}, m, e) \quad (2.4)$$

(2.4) 式の右辺は, パラメータが多く複雑なため, 計算が困難である. そこで, Model1 では (2.4) 式のパラメータを簡略化する.

- 文の長さの確率 $\epsilon (= \epsilon(m|l))$ は, m および e に独立する

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

- アライメントの確率は英語文の長さ l に依存する

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

- フランス語の翻訳確率 $t(f_j|e_{a_j})$ は, フランス単語 f_j に対応づけられる英単語 e_{a_j} に依存する

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

パラメータを簡略化した場合の $P(f, a|e), P(f|e)$ は以下になる.

$$P(f, a|e) = \frac{\epsilon}{(l+1)^m} \prod_{j=1}^m t(f_j|e_{a_j}) \quad (2.5)$$

$$P(f|e) = \frac{\epsilon}{(l+1)^m} \sum_{a_1=0}^l \cdots \sum_{a_m=0}^l \prod_{j=1}^m t(f_j|e_{a_j}) \quad (2.6)$$

$$= \frac{\epsilon}{(l+1)^m} \prod_{j=1}^m \sum_{i=0}^l t(f_j|e_{a_j}) \quad (2.7)$$

Model1 は, $t(f|e)$ の初期値が 0 以外の時, EM アルゴリズムを繰り返して得られる唯一の極大値より最適解を推定する. EM アルゴリズムは以下の手順で行われる.

1. 翻訳確率 $t(f|e)$ の初期値を設定する.
2. フランス語と英語の対訳 $(f^{(s)}, e^{(s)})$, $1 \leq s \leq S$ において, 英単語 e とフランス単語 f が対応する回数の期待値 $c(f|e; f^{(s)}, e^{(s)})$ を計算する. なお, 期待値 $c(f|e; f^{(s)}, e^{(s)})$ は以下の式になる.

$$c(f|e; f^{(s)}, e^{(s)}) = \frac{t(f|e)}{t(f|e_0) + \dots + t(f|e_l)} \sum_{j=1}^m \delta(f|f_j) \sum_{i=1}^l \delta(e|e_i) \quad (2.8)$$

フランス語文 f のうちフランス単語 f_j が出現する回数 $\sum_{j=1}^m \delta(f|f_j)$, 英語文 e のうち英単語 e_i が出現する回数 $\sum_{i=1}^l \delta(e|e_i)$ である

3. 英語文 e のうち最低一回出現する英単語 $e^{(s)}$ に対し,
 - 定数 λ_e を以下の式で計算する.

$$\lambda_e = \sum_f \sum_{s=1}^S c(f|e; f^{(s)}, e^{(s)}) \quad (2.9)$$

- 翻訳確率 $t(f|e)$ を以下の式で再計算する.

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

4. 翻訳確率 $t(f|e)$ が収束するまでステップ2と3を繰り返す.

2.3.4 Model2

Model1 において, $P(f_j|a_1^{j-1}, f_1^{j-1}, m, e)$ は英語文の長さ l のみに依存する. しかし, Model2 では, j 単語目のアライメント a_j , フランス語文の長さ m にも依存する. この場合, 以下のような関係となる.

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

これより, (2.6) 式は以下の式のようになる.

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

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

Model2 における期待値は $c(f|e; \mathbf{f}, \mathbf{e})$ と $c(i|j, m, l; \mathbf{f}, \mathbf{e})$ の二種類ある. これらは以下の式で求められる.

$$\begin{aligned} c(f|e; \mathbf{f}, \mathbf{e}) &= \frac{t(f|e)}{t(f|e_0) + \cdots + t(f|e_l)} \sum_{j=1}^m \delta(f|f_j) \sum_{i=1}^l \delta(e|e_i) \\ &= \frac{t(f|e) \alpha(i|j, m, l) \delta(f|f_j) \delta(e|e_i)}{t(f_j|e_0) \alpha(0|j, m, l) + \cdots + t(f_j|e_l) \alpha(l|j, m, l)} \end{aligned} \quad (2.14)$$

$$\begin{aligned} c(i|j, m, l; \mathbf{f}, \mathbf{e}) &= \sum_a P(a|e, f) \delta(i, a_j) \\ &= \frac{t(f|e) \alpha(i|j, m, l)}{t(f_j|e_0) \alpha(0|j, m, l) + \cdots + t(f_j|e_l) \alpha(l|j, m, l)} \end{aligned} \quad (2.15)$$

なお, Model2 は, EM アルゴリズムで計算した場合, 複数の極大値を持ち, 最適解を求める保証はない. しかし, Model1 は, Model2 において, $a(a_j|j, m, l) = (l+1)^{-1}$ となる特殊な状態であり, 最適解を求めることができる. このため, Model2 で最適解を求めるとき, Model1 を用いる.

2.3.5 Model3

Model3 は, Model1 や Model2 とは異なり, ある英単語がフランス単語に何個対応するか, 対応する英単語とフランス単語の距離を考慮する. なお, Model3 は以下 4 つのパラメータを用いる.

- $t(f|e)$
フランス単語 f が英単語 e に翻訳されるとき確率
- $n(\phi|e)$
英単語 e がフランス単語に ϕ 個対応するとき確率
- $d(j|i, l, m)$
フランス語文が m 単語, 英語文が l 単語あるとき, 単語位置 i である英単語が単語位置 j であるフランス単語に対応するときの歪み確率

- p_0, p_1

英単語に対応しないフランス単語が ϕ_0 個あるときの確率

なお p_0 は以下の式で表される.

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

よって, Model3 は以下の式で表される.

$$\begin{aligned} P(f|e) &= \sum_{a_1=0}^l \dots \sum_{a_m=0}^l P(f, a|e) \\ &= \sum_{a_1=0}^l \dots \sum_{a_m=0}^l \binom{m - \phi_0}{\phi_0} p_0^{m-2\phi_0} p_1^{\phi_0} \prod_{i=1}^l \phi_i! n(\phi_i | e_i) \\ &\quad \times \prod_{j=1}^m t(f_j | e_{a_j}) d(j | a_j, l, m) \end{aligned} \quad (2.17)$$

Model3 は, すべてのアライメントを計算する量が莫大であるため, EM アルゴリズムによる近似で求める. また, 英単語とフランス単語における位置は絶対位置で考慮する.

2.3.6 Model4

Model4 は, Model2 や Model3 と違い, 単語の位置を絶対位置ではなく相対位置で考慮する. また, 英語とフランス語における名詞と形容詞の位置など, 各単語ごとの位置を考慮する. なお, Model4 は, 歪み確率 $d(j|i, l, m)$ を二つの場合について考慮する.

- 繁殖数が 1 以上である英単語に対応するフランス単語のうち, 単語が最も文頭に近い場合

$$P(\Pi_{[i]1} = j | \pi_1^{[i]-1}, \gamma_0^l, e) = d_1(j - \odot_{i-1} | A(e_{[i-1]}), B(f_i))$$

- 上記以外の場合

$$P(\Pi_{[i]1} = j | \pi_1^{[i]-1}, \gamma_0^l, e) = d_{>1}(j - \pi_{[i]k-1} | B(f_i))$$

2.3.7 Model5

Model4は、フランス語における複数の単語を同じ位置に重ねたり、単語が生成されない位置が生成される欠点がある。その欠点を改善したのがModel5である。

2.3.8 GIZA++

GIZA++は、統計翻訳で用いることを前提にして、OchとNeyによって作成されたツールである。これは、IBM Model1~5を用いて翻訳する言語と目的言語における単語の対応関係 (Word Alignment) の確率値を計算する。GIZA++を用いた場合、以下のファイルが出力される。

1. T TABLE (Translation Table)

T TABLEは、Model1からModel3により作成された翻訳確率 $P(f|e)$ のデータである。 f は翻訳する言語で、 e は目的言語である。T TABLEは各行が、目的言語の単語ID($e_i d$)、翻訳する言語の単語ID($f_i d$)、翻訳する言語の単語から目的言語の単語へ翻訳する確率 ($P(f_i d|e_i d)$) で構成される。

2. N TABLE (Fertility Table)

N TABLEは、目的言語の単語における繁殖数を表したデータである。N TABLEは各行が、目的言語の単語ID($e_i d$)、繁殖数が0である確率 (p_0)、繁殖数が1である確率 (p_1)、 \dots 、繁殖数が n である確率 (p_n) で構成される。

2.4 言語モデル

2.4.1 言語モデルの概要

言語モデルは、単語列や文字列に対して、それらが起こる確率を与えるモデルである。日英統計翻訳においては、より英語らしい翻訳候補に対して高い確率を与えることにより、翻訳候補から英語として自然な文を選出する。言語モデルの例を下記に示す。

-2.136928 many books -0.09209893
 -1.276793 many people -0.142345
 -2.582711 many points -0.124561
 -2.138576 many questions -0.2320607
 ...

一番上の行は、左から、many の後に books が来る確率を常用対数 \log_{10} で表した値 $\log_{10}(P(\text{books}|\text{many}))$ 、2-gram で表された単語列 many books、バックオフスムージングにより、many の後に books が来る確率を常用対数 \log_{10} で表した値 $\log_{10}(P(\text{books}|\text{many}))$ である。バックオフスムージングについては、第 2.4.3 章にて述べる。

2.4.2 N-gram モデル

言語モデルは、N-gram モデルが代表的である。N-gram モデルは、“単語列 $w_1^n = w_1, w_2, \dots, w_n$ 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 $w_1 w_2 \dots w_n$ の確率を $P(w_1 w_2 \dots w_n)$ とすると、(2.18) の計算式を表現することが可能である。

$$P(w_1 w_2 \dots w_n) = \prod_{i=1}^n P(w_i | w_{i-N+1}, w_{i-N+2}, \dots, w_{i-1}) \quad (2.18)$$

例えば、「I am a student .」という文字列に対する 2-gram モデルを以下に示す。

$$P(e = \text{“I am a student .”}) \approx P(I)P(am|I)P(a|am)P(student|a)P(.|student)$$

3-gram モデルであれば $P(a|I am)$ 、4-gram モデルであれば $P(student|I am a)$ というように、 $(N - 1)$ 単語の次にくる単語が「a」, 「student」である確率を考える。

しかし、N-gram モデルは局所的な情報であり、語順の情報を持たない。例えば日英統計翻訳のような異なる語順間の翻訳は、日本語から英語へ語順が変化するため、同じ語順間の翻訳と比較して、翻訳精度が低下する傾向となる。

2.4.3 スムージング

N -gram の確率は、最尤推定法を用いて計算することが可能である。単語列 $w_{i-2}w_{i-1}w_i$ を w_i^{i-2} 、単語列 $w_{i-2}w_{i-1}w_i$ の確率を $N(w_i^{i-2})$ とすると、(2.19) から 3-gram の確率を計算することが可能である。

$$p(w_i|w_{i-2}) = \frac{N(w_i^{i-2})}{N(w_{i-2}^{i-1})} \quad (2.19)$$

しかし、入力文において学習データに存在しない単語列がある場合、その単語列の確率は 0 となり、翻訳精度が悪化する傾向となる。そこで、確率が 0 となるのを防ぐために、大きい確率はより小さく、小さい確率はより大きくし、全体の確率値を平滑化及び近似を行う。これはスムージングという方法であり、代表的には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、学習データには出現していない N -gram 単語列の確率を、 $N-1$ -gram 単語列の確率でスムージングを行う。 $N=3$ -gram の場合のバックオフスムージングによる確率は以下の式に求められる。

$$P(w_i|w_{i-1}^{i-2}) = \begin{cases} \lambda(w_{i-2}^i)p(w_i|w_{i-2}^i) & \text{if } N(w_{i-2}^i) \\ (1 - \lambda_0(w_{i-2}^{i-1}))\alpha P(w_i|w_{i-1}^{i-2}) & \text{else } N(w_{i-2}^{i-1}) \\ P(w_i|w_{i-1}^{i-1}) & \text{other} \end{cases} \quad (2.20)$$

λ はディスカウント係数と呼ばれ、学習データに存在しない N -gram に対し、学習データに存在する $N-1$ -gram から確率値を求める。 α は確率の総和を 1 にするための正規化係数である。 λ_0 及び α は以下に求められる。

$$\lambda_0(w_{i-2}^i) = \sum_{w_i} \lambda(w_{i-2}^i)p(w_i|w_{i-2}^i) \quad (2.21)$$

$$\alpha = (1 - \sum_{N(w_{i-2}^i) > 0} P(w_i|w_{i-1}^{i-2}))^{-1} \quad (2.22)$$

ディスカウント係数を求める方法は多く、ukndiscount(Unmodified Kneser-Ney discounting) や kndiscount(Chen and Goodman's modified Kneser-Ney discounting) がある。

ukndiscount と kndiscount の違いは、一つまたは二つの単語を考慮してスムージングを行う (undiscount) か、一つ、二つまたは三つの N -gram 単語列を考慮してスムージングを行う (kndiscount) かである。

2.5 reordering-model

reordering-model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。reorderingには、Lexicalized Block Orientationモデル[8]および、歪みペナルティ[9]が広く使用されている。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reorderingの評価を行う。

2.6 デコーダー

デコーダーは、翻訳モデルと言語モデルを用いて、 $\hat{e} = \operatorname{argmax}_e P(e|j)$ が最大となる翻訳候補を探索し、出力する。入力文として、「私は学生です。」を用いたときの翻訳例を図2.3に示す。

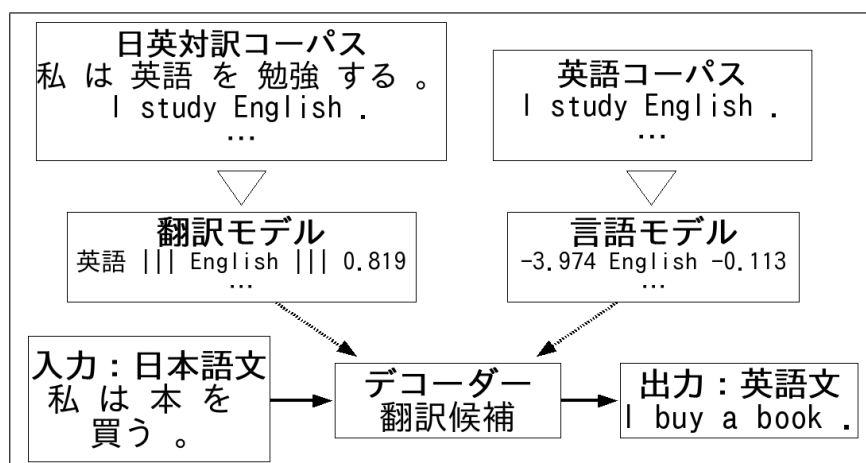


図 2.3: デコーダーにおける翻訳候補の生成例

日英統計翻訳において、 $\hat{e} = \operatorname{argmax}_e P(e|j)$ の確率が最大となる翻訳文を出力するためには、適切な日本語と英語の単語対応を選択する必要がある。しかし、適切なものを選択するためには、莫大な計算量と時間が必要となる。莫大な計算量と時間を削減する方法として、ビームサーチ法が存在する。

2.7 評価システム

統計翻訳の評価方法として、自動評価が用いられる。自動評価ではデコーダーにより、入力文を翻訳した翻訳文と正解翻訳例である参照文との一致率などを計算し、スコアを出力する。評価システム例を図 2.4 に示す。

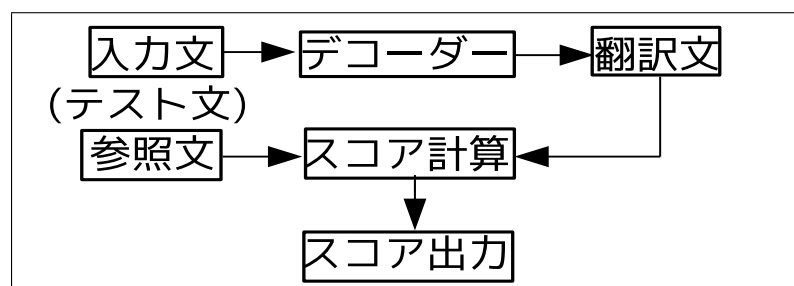


図 2.4: 評価システム例

2.8 パラメータチューニング

パラメータチューニングは、デコーダーで用いるパラメータの最適値を設定し、Minimum Error Rate Training(MERT)[10]という手法を用いることが一般的である。MERTは評価関数(通常 BLEU)を最大にする翻訳結果が選ばれるために、パラメータを調整する。パラメータチューニングを行う際、ディベロップメント文と呼ばれる試し翻訳を行うデータを用いる。

ディベロップメント文の各文から、上位 100 個程度の翻訳候補を出力し、その候補の中でパラメータの重みを変え、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。しかし、MERT によって生成されたパラメータは、ディベロップメント文に依存するため、そのパラメータを別のテストデータに使用しても翻訳精度が向上されない場合がある。

日英統計翻訳において、語順が単純な単文 100 文を用いて、チューニングを行った場合、30 分で終了する。しかし、特許文をはじめとする語順が複雑な文 100 文を用いて、チューニングを行った場合、チューニングが終了されるのに、一日以上時間がかかる場合がある。

第3章 日本語文の語順変更による翻訳

3.1 先行研究

現在，統計翻訳において，語順変更を行い，日本語文を英語の語順に近づけてから，統計翻訳をする研究が行われている．語順変更を用いた研究について，以下で説明する．

3.1.1 動詞の語順変更 [3]

岡崎らによる動詞の語順変更の手法は日本語文を英語の語順に近づけるために，日本語文の主語，目的語，動詞 (SOV) を主語，動詞，目的語 (SVO) の順に並び替え，日英統計翻訳を行う．動詞の語順変更の例を図 3.1 に示す．

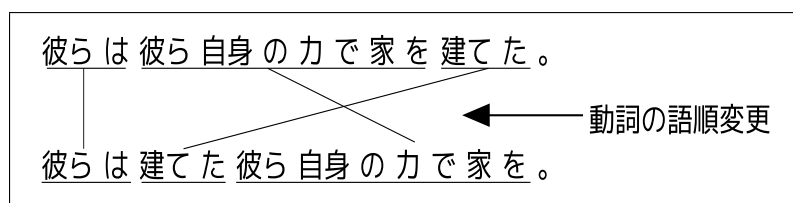


図 3.1: 動詞の語順変更の例

岡崎らは単文を用いて，実験を行った．岡崎らによる動詞の語順変更の手法での自動評価の結果を表 3.1，対比較評価の結果を表 3.2 に示す．

表 3.1: 自動評価結果

	BLEU	NIST	METEOR
動詞の語順変更	0.1239	4.0997	0.3619
通常の統計翻訳	0.1208	4.0818	0.3587

表 3.2: 対比較評価結果

動詞の語順変更○	通常の統計翻訳○	差なし
41	38	21

3.1.2 述語項構造に基づく語順変更 [4]

星野らによる述語項構造に基づく語順変更の手法を以下に示す。

1. 係り受け・述語項構造解析

日本語文に対して、係り受け・述語項構造解析を行う。

2. 文の語順変更

係り受け関係を用いて、深さ優先に基づき、並び替える。

3. 助詞の語順変更

目的言語での順序に基づき、句中の内容語と機能語を並び替える。

述語項構造に基づく語順変更 [4] の例を図 3.2 に示す。

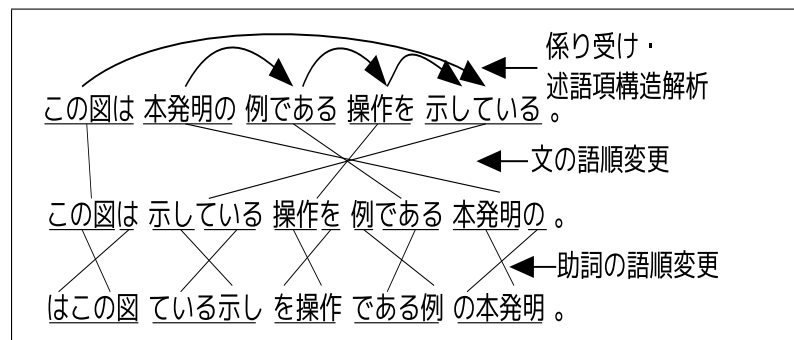


図 3.2: 述語項構造に基づく語順変更の例

星野らは特許文を用いて、実験を行った。星野らによる述語項構造に基づく語順変更の手法での自動評価の結果を表 3.3 に示す。

表 3.3: 自動評価結果

	BLEU	RIBES
述語項構造に基づく語順変更	0.3001	0.724
通常の統計翻訳	0.2919	0.685

3.1.3 語順変更および句語順変更を用いた手法 [11]

畑中らによる語順変更および句語順変更を用いた手法を以下に示す。

1. 動詞の語順変更

日本語文の主語，述語，動詞 (SOV) を主語，動詞，述語 (SVO) の順に並び替える。この語順変更は岡崎らの動詞の語順変更の手法と，ほぼ同等である。動詞の語順変更の例を表 3.3 に示す。

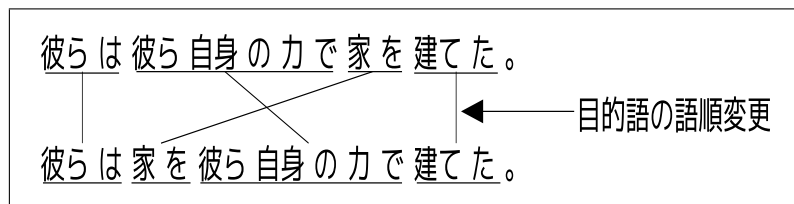


図 3.3: 動詞の語順変更の例

2. 目的語の語順変更

目的語の句を後ろから順に並び替える。なお，この語順変更は星野らの述語項構造に基づく事前並べ替えの手法と，ほぼ同等である。目的語の語順変更の例を表 3.4 に示す。

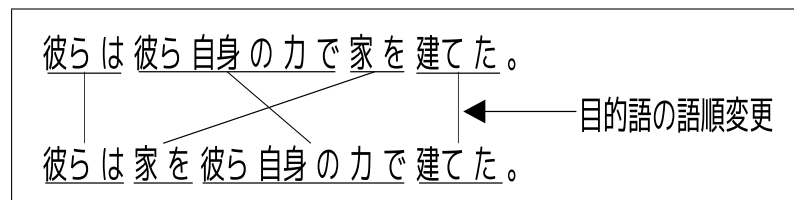


図 3.4: 目的語の語順変更の例

3. 助詞の語順変更

目的語内の文節の内容語と機能語を並び替える。助詞の語順変更の例を表 3.5 に示す。

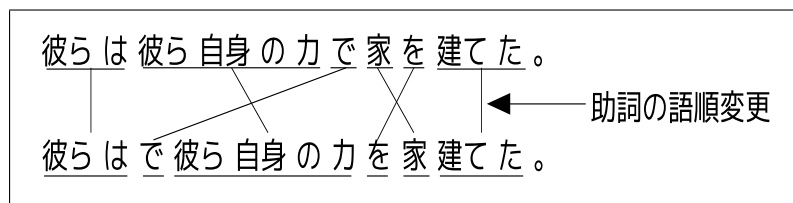


図 3.5: 助詞の語順変更の例

動詞の語順変更，目的語の語順変更と助詞の語順変更を組み合わせたときの例を図 3.6 に示す。

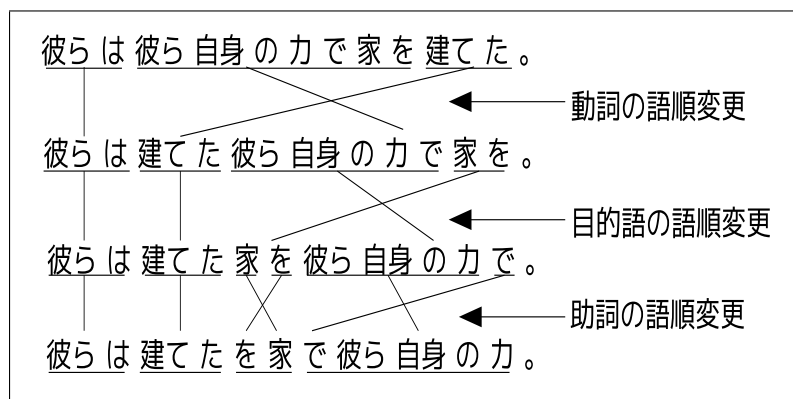


図 3.6: 動詞+目的語+助詞の語順変更の例

畑中らは単文を用いて，実験を行った。畑中らによる語順変更の手法での自動評価の結果を表 3.4，対比較評価の結果を表 3.5 に示す。

表 3.4: 自動評価結果

	BLEU	NIST	METEOR	RIBES
動詞	0.1361	4.7408	0.4550	0.7227
目的語	0.1343	4.7575	0.4548	0.7201
助詞	0.1305	4.7002	0.4522	0.7197
動詞+目的語	0.1352	4.8356	0.4644	0.7276
動詞+助詞	0.1326	4.7369	0.4572	0.7223
目的語+助詞	0.1269	4.6393	0.4474	0.7194
動詞+目的語+助詞	0.1365	4.8538	0.4659	0.7264
ベースライン	0.1343	4.7646	0.4519	0.7196

表 3.5: 対比較評価結果

	ベースライン○	提案手法○	差なし	同一出力
動詞	6	8	75	11
目的語	4	6	63	27
助詞	5	11	73	11
動詞+目的語	3	13	71	13
動詞+助詞	6	12	71	11
目的語+助詞	6	10	73	10
動詞+目的語+助詞	6	14	69	11

3.2 提案手法

岡崎 [3] や星野 [4] 研究は翻訳される英語文の語順を推定して、日本語文の語順変更を行っている。そのため、目標とする英語文の最適な語順に並び替えられていない。

そこで、本研究では語順変更の手法の限界を調査するため、テスト文に参照文があると仮定し、日本語文を英語の参照文の語順に並び替える。その後、句に基づく統計翻訳を行い、語順変更の効果を調査する。以下に手順を示す。

手順1 変数化

対訳句辞書と対訳文を用いて、学習文、テスト文、ディベロップメント文に変数化を行う。対訳句辞書と対訳文を比較し、共通する箇所を変数にする。なお、対訳句辞書には鳥バンク [12] 約 21 万句を使用する。変数化の手順を図 3.7 に示す。

日本語文
わたしは <u>3</u> 時間 も 森 の <u>小道</u> を <u>歩いて</u> <u>その</u> 宿 に <u>着いた</u> 。
X1 X2 X3 X4 X5 X6 X7 X8 X9
対訳句辞書(鳥バンク209,539句)
わたし / I 3 / three 時間 / hours 森 / the woodland 小道 / path 歩い / walking その / the 宿 / inn 着い / got
英語文
<u>I</u> got to <u>the</u> <u>inn</u> after <u>walking</u> <u>three</u> <u>hours</u> along <u>the</u> <u>woodland</u> <u>path</u> .
X1 X9 X7 X8 X6 X2 X3 X4 X5

図 3.7: 変数化

日本語文の文頭の“わたし”から順番に変数化を行う。英語文の文頭に“I”があり、対訳句辞書にも“わたし/I”があるので、変数化を行い X1 とする。次に、日本語文の“は”を変数化しようと試みる。英語文の文頭の“I”は X1 を変数にするときに使用したため、次の単語“got”から順番に対訳句辞書と比較する。つまり、“は/got”が対訳句辞書にあるかどうか調べる。対訳句辞書にない場合、英語文の“got”の次の単語“to”で対訳句辞書と比較する。つまり、“は/to”が対訳句辞書にあるかどうか調べる。この作業を繰り返し、英語の文末の単語(“path”)まで対訳句辞書と比較しても、共通する対訳句がないので字面として残す。次に、日本語文の“3”を変数化しようと試みる。英語文の“got”から順番に対訳句辞書と比較していくと、英語文の“three”のときに、対訳句辞書に“3/three”があるので、変数化を行い X2 とする。以降、同様に日本語文と英語文に対して変数化を行う。

手順2 語順変更

対訳文を用いて、学習文、テスト文、ディベロップメント文の日本語文を英語文の変数の順番に並び替え、出力を日本語変更文とする。なお、移動単位は次の変数までとする。語順変更の手順を図 3.8 に示す。

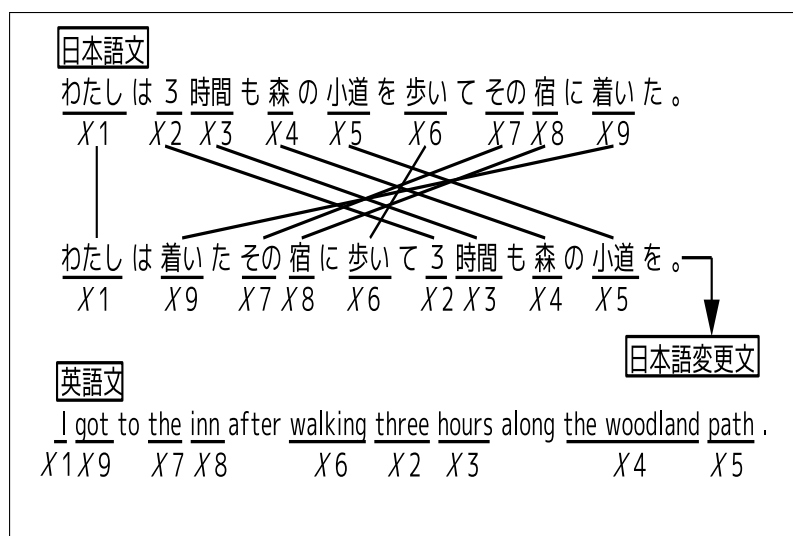


図 3.8: 語順変更

手順1で作成した英語文の変数の順番, “X1, X9, X7, X8, X6, X2, X3, X4, X5”の順番に日本語文を並び替える. まず, X1の“わたし”を移動させる. なお, 移動単位は次の変数までなので, 次の変数(X2)までの“わたしは”を移動させる. 次に, 英語文の変数の順番に従って, X9の“着い”を移動させる. なお, X9は変数の最後なので, 文末までの“着いた”を移動させる. 以降, 同様に日本語文を英語文の変数の順番に並び替える.

手順3 翻訳実験

学習文, テスト文, デイベロップメント文に語順変更を行ったデータを用いて, 句に基づく統計翻訳を行う.

第4章 実験環境

4.1 実験データ

本研究では，辞書の例文から抽出した対訳データの単文コーパス [13] と重文複文コーパス [13] を用いる．実験データの文数と平均単語数を表 4.1 に示す．

表 4.1: 実験データの文数と平均単語数

	単文コーパス	重文複文コーパス
学習文	99,989 文	97,765 文
テスト文	9,999 文	9,784 文
ディベロップメント文	1,000 文	976 文
平均単語数 (実験データの文数)	10.5 単語 (110,988 文)	14.0 単語 (108,525 文)

単文コーパスと重文複文コーパスは対訳コーパスであり，日本語コーパスと英語コーパスに分けられている．対訳データの日本語コーパスに対しては，Cabocha[14] を用いてわかち書きを行う．英語コーパスに対しては，tokenizer.perl を用いて，二つ以上の単語が結合している単語を分離するが，lowercase.perl を用いた英字の小文字化は行わない．

4.1.1 単文コーパス

日英対訳コーパスと日本語変更文コーパスの例を以下に示す．

日本語コーパス

組合はストライキに参加する。
私はいつも辞書を手近に置いている。
この本は学生の要求に応えるだろう。

—— 英語コーパス ——

The union will join in the strike .
I always keep a dictionary at hand .
This book will meet the needs of students .

—— 日本語変更文コーパス ——

組合は参加するストライキに。
私はいつも置いている辞書を手近に。
この本は応えるだろう要求に学生の。

4.1.2 重文複文コーパス

日英対訳コーパスと日本語変更文コーパスの例を以下に示す。

—— 日本語コーパス ——

広々とした牧場に牛が群れているのが見える。
その交友関係を求めることには下心があった。
現実には君の考えているよりも厳しいものだ。

—— 英語コーパス ——

We see herds of cows in the open meadow .
There was an ulterior motive in cultivating the friendship .
Reality is harsher than you think .

—— 日本語変更文コーパス ——

見える群れているのが牛が広々とした牧場に。
下心があった求めることにはその交友関係を。
現実には厳しいものだ君の考えているよりも。

4.2 対訳句辞書

本研究では対訳句として、鳥バンク [12] を用いる。鳥バンクのパターンと原文を比較した変数部から品詞が *AJ*(形容詞), *V*(動詞), *N*(名詞), *ADV*(副詞), *AJV*(形容動詞), *VP*(動詞句), *AJVP*(形容動詞句), *ADVVP*(形容詞句) の対訳句を抽出し、UNIQ をかけて同じ対訳句を削除し、対訳句辞書を作成する。なお、間違った対応を取る単語を手手で

削除した。削除する対象となる単語を表 4.2 に示す。

表 4.2: 削除する対象となる単語

彼, 彼女, それら, 私, 私たち, あなた
I, my, me, mine, myself, you, your, yours, yourself,
he, his, him, himself, she, her, herself, they, their, them, theirs, themselves,
a, an, the, it, that, this, these, those, be, is, are, was, were

表 4.2 の削除する対象となる単語から, 例として “彼”, “私”, “be”, “it” の間違っただ対応を取ると判断する基準を表 4.3 に示す。

表 4.3: 間違っただ対応を取ると判断する基準

日本語句	英語句	判断
彼	He	正しい対応 (残す)
彼	John	間違っただ対応 (削除)
彼	I	間違っただ対応 (削除)
私	I	正しい対応 (残す)
私	she	間違っただ対応 (削除)
私	one	間違っただ対応 (削除)
be	ある	正しい対応 (残す)
be	起き	間違っただ対応 (削除)
be	出席し	間違っただ対応 (削除)
it	それ	正しい対応 (残す)
it	お皿	間違っただ対応 (削除)
it	コーヒー	間違っただ対応 (削除)

表 4.3 の判断基準によりクリーニングを行った。クリーニング前とクリーニング後の対訳句数を表 4.4 に示す。

表 4.4: 対訳句数

クリーニング前の対訳句数	クリーニング後の対訳句数	削除した対訳句数
213,511 句	209,539 句	3,972 句

抽出した対訳句辞書の例を以下に示す。

— 対訳句辞書の例 —

若い		young (<i>AJ</i>)
思わ		expected (<i>V</i>)
彼女		She (<i>N</i>)
いつも		always (<i>ADV</i>)
とても 静かな		very quiet (<i>AJV</i>)
ぼくにあいさつし		greeted me (<i>VP</i>)
わたしの好きな		my favorite (<i>AJVP</i>)
長い間		for a long time (<i>ADVP</i>)

鳥バンクの対訳句の精度を調べるために、鳥バンクの対訳句からランダムに 100 句抽出して、正しい対応が取られているか調査した。鳥バンクの対訳句の精度を表 4.5 に示す。

表 4.5: 鳥バンクの対訳句の精度

対訳句の対応○	対訳句の対応×
92	8

また、正しい対応と判断した対訳句と間違った対応と判断した対訳句の例を以下に示す。

— 対訳句の対応○の例 —

誇り		proud
子ども		children
ひどく		desperately
ノックする		knock
達する		run up to
新メニュー		a new menu
驚きを期待する		anticipates a surprise
宿題を終え		finished his homework

対訳句の対応×の例

俺		father
国民		a people
手術		a surgeon
作れ		cultivate
はつきりし		vague
恥も外聞も忘れ		shamelessly
フットボール生活		football career
営巣用の洞を掘る		excavate nesting cavities

4.3 翻訳モデルの学習

翻訳モデルの学習は，“train-factored-phrase-model.perl”を用いて行う。このプログラムは，IBM model1～5に基づくGIZA++を使用する。本実験において，フレーズテーブルを作成する学習法は，Mosesにおける標準的学習法であるgrow-diag-finalを用いて，max-phrase-lengthは20と設定する。

4.4 言語モデルの学習

言語モデルの学習には，“SRILM (Stanford Research Institute Language Model)”[15]のngram-countを使用する。本研究で使用するスムージングはkndiscountである。また， $N=5$ -gramとする。

4.5 デコーダーに関するパラメータ

デコーダーは，moses[16]を使用する。mosesはパラメータを設定することができる。本研究では，次のようにパラメータを設定する。翻訳モデルの重みである“weight-t”の値は，“0.2 0.2 0.2 0.2 0.2”とする。また，日本語から英語への翻訳は，動詞の位置が大きく変化する。そこで，フレーズの並び替え確率の重み“distortion-weight”の値は“0.3”，フレーズテーブルの読み込み制限を“0(無制限)”，言語モデルの重みである“weight-l”の値を“0.1”，“ttable-limit”の値を“40”，ワードペナルティ“weight-w”の値を“-1”とする。

moses には, “reordering-model” というモデルが存在する. これは, 単語ごとの並び替えに確率をつけることで, 自然な英語文となる翻訳候補を出力する. 本研究では, “reordering-model” を用いて, 実験を行う.

4.6 ベースライン

語順変更を行わない実験をベースラインとする. ただし, フレーズの並び替え制限である “distortion-limit” の値は “-1” に設定する.

4.7 提案手法

対訳句辞書と対訳文を用いて, 学習文, テスト文, デイベロップメント文の日本語文を英語の参照文の語順に並び替える. ただし, “distortion-limit” の値は “6” に設定する.

4.8 評価方法

評価方法に自動評価と人手評価を用いる. 自動評価には BLEU[17], NIST[17], METEOR[18], RIBES[19], TER[20], WER[20] を用いる. BLEU は出力された翻訳文と正解翻訳文を 4-gram で, 単語列が正しいか評価する. NIST は, 出力された翻訳文と正解翻訳文を 5-gram で, 単語訳が正しいか評価する. METEOR は意味などの単語属性が正しいか評価する. RIBES は文全体の大局的な並びが正しいか評価する. TER は翻訳編集率であり, 人間が機械翻訳結果の後編集を行った際のコストに着目した評価尺度である. WER は単語誤り率であり, 音声認識の評価などで広く用いられる. BLEU と METEOR と RIBES における評価値の範囲は 0 から 1 であり, 1 が最大となる. TER, WER における評価値は 0 が最大となる. NIST における評価値の範囲は 0 から無限大であり, 評価値が高いほど良い評価となる.

人手評価は翻訳文からランダムに 100 文取り出し，ベースラインとの対比較実験を行う。なお，評価基準は以下のようにする。

- 提案手法○：提案手法の方が良い
- ベースライン○：ベースラインの方が良い
- 差なし：翻訳精度に明確な差がない
- 同一出力：完全に同一の出力

第5章 翻訳実験

5.1 実験内容

対訳句辞書と対訳文を用いて，学習文，テスト文，ディベロップメント文の日本語文を英語の参照文の語順に並び替える．また，実験は“reordering-model”を使用し，句に基づく統計翻訳を行う．なお，翻訳は日英統計翻訳で行う．

5.2 単文の実験

5.2.1 語順変更の精度

本研究で用いた語順変更のプログラムの精度を調べるため，単文のテスト文から，ランダムに100文抽出して，正しく語順変更されているか調査した．単文の語順変更の精度を表5.1.に示す．

表 5.1: 単文の語順変更の精度

語順変更○	語順変更×
90	10

誤り率は10%であった．語順変更に成功した例を表5.2，失敗した例を表5.3に示す．なお，下線部は間違いと判断した箇所を表す．

表 5.2: 単文の語順変更の成功例

例 1	
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 .
日本語変更文	見ましたその試合はテレビで。
日本語パターン	X4 ました X1 X2 は X3 で。
英語パターン	I X4 X1 X2 on X3 .
例 2	
入力文	あの大学に入るには高い学力が必要だ。
参照文	Entering that university requires high scholastic ability .
日本語変更文	入るにはあの大学に必要な高い学力が。
日本語パターン	X3 には X1 X2 に X6 だ X4 X5 が。
英語パターン	X3 X1 X2 X6 X4 scholastic X5 .
例 3	
入力文	その新しい奨学金は若い技術者に刺激となるだろう。
参照文	The new scholarship will provide a stimulus for young engineers .
日本語変更文	その新しい奨学金はなるだろう刺激と若い技術者に。
日本語パターン	X1 X2 X3 は X7 だろう X6 と X4 X5 に。
英語パターン	X1 X2 X3 will X7 X6 for X4 X5 .

表 5.3: 単文の語順変更の失敗例

例 1	
入力文	彼は頭が天井に着くほど背が高い。
参照文	He is so tall that his head touches the ceiling .
日本語変更文	彼は高い頭が着くほど背が天井に。
日本語パターン	X1 は X5 X2 が X4 ほど <u>背が</u> X3 に。
英語パターン	X1 is so X5 that his X2 X4 the X3 .
例 2	
入力文	あの人は日本語がちょっとできます。
参照文	That person can speak a little Japanese .
日本語変更文	あのできますちょっと人は日本語が。
日本語パターン	X1 X4 ます X3 <u>X2</u> は日本語が。
英語パターン	X1 <u>person</u> can X4 X3 <u>X2</u> .
例 3	
入力文	わが国もこの問題で引き続き協力したい。
参照文	Japan , too , must continue to cooperate in <u>solving Mexico's debt problem</u> .
日本語変更文	わが国もこの引き続き協力したい問題で。
日本語パターン	X1 も <u>この</u> X3 X4 したい X2 で。
英語パターン	X1 , too , must X3 to X4 in <u>solving Mexico's debt</u> X2 .

語順変更失敗の原因を以下に示す。

- 例 1：入力文の“背”，に対応する単語が参照文になく，字面として残ったため，語順変更失敗した。
- 例 2：変数 X2 が“人/person”ではなく，“人/Japanese”で変数になってしまったので，語順変更失敗した。
- 例 3：入力文の“この”に対応するのが，参照文では“solving Mexico's debt”になり，対訳句辞書になかった。そのため，“この”が変数にならず，字面として残ったため，語順変更失敗した。

5.2.2 自動評価

自動評価の結果を表 5.4 に示す。

表 5.4: 単文の実験の自動評価結果

	BLEU	NIST	METEOR	RIBES	TER	WER
提案手法	0.1430	4.9098	0.4617	0.7345	0.6820	0.6970
ベースライン	0.1340	4.7326	0.4513	0.7024	0.7178	0.7456

5.2.3 人手評価

人手評価の結果を表 5.5 に示す.

表 5.5: 単文の実験の対比較評価結果

提案手法○	ベースライン○	差なし	同一出力
7	5	86	2

5.2.4 翻訳例

提案手法の方が優れている翻訳例を表 5.6, ベースラインの方が優れている翻訳例を表 5.7 に示す. なお, 下線部は間違いと判断した箇所を表す.

表 5.6: 提案手法の方が優れている翻訳例

例 1	
入力文	彼はわがチームの主将としてふさわしい。
語順変更文	彼は主将としてふさわしいわがチームの。
提案手法○	He is fit as the captain of our team .
ベースライン	He <u>which</u> as the captain of our team .
参照文	He is working to be captain of our team .
例 2	
入力文	ひどい精神 錯乱 となっている。
語順変更文	ひどい精神 錯乱 となっている。
提案手法○	His mind was badly .
ベースライン	He is <u>錯乱</u> badly .
参照文	He is seriously deranged .
例 3	
入力文	地元の経済においてその会社の存在は重要である。
語順変更文	その会社の存在は重要である地元の経済において。
提案手法○	The company is very important in the local economy .
ベースライン	It is important in the presence of the company <u>local economy</u> .
参照文	The company bulks large in the local economy .

表 5.7: ベースラインの方が優れている翻訳例

例 1	
入力文	この風邪薬はよく効く。
語順変更文	この風邪薬はよく効く。
提案手法	This medicine is effective <u>cold</u> .
ベースライン○	This cold medicine works well .
参照文	This cold medicine is effective .
例 2	
入力文	自動車が動かない。
語順変更文	自動車が動かない。
提案手法	The car <u>stationary</u> .
ベースライン○	The car is not working .
参照文	The car will not start .
例 3	
入力文	そのひどい知らせにすっかり打ちひしがれた。
語順変更文	すっかり打ちひしがれたそのひどい知らせに。
提案手法	I was completely <u>overwhelmed the bad news</u> .
ベースライン○	She was absolutely crushed by the bad news .
参照文	We were totally crushed by the terrible news .

5.3 重文複文の実験

5.3.1 語順変更の精度

本研究で用いた語順変更のプログラムの精度を調べるため、重文複文のテスト文から、ランダムに 100 文抽出して、正しく語順変更されているか調査した。重文複文の語順変更の精度を表 5.8 に示す。

表 5.8: 重文複文の語順変更の精度

語順変更○	語順変更×
87	13

誤り率は 13% であった。語順変更に成功した例を表 5.9, 失敗した例を表 5.10 に示す。なお、下線部は間違いと判断した箇所を表す。

表 5.9: 重文複文の語順変更の成功例

例 1	
入力文	その建物は私たちが到着したとき盛んに燃えていた。
参照文	The building was blazing furiously when we arrived .
日本語変更文	その建物は燃えていた盛んにとき私たちが到着した。
日本語パターン	X1 X2 は X7 ていた X6 X5 X3 が X4 した。
英語パターン	X1 X2 was X7 X6 X5 X3 X4 .
例 2	
入力文	わたしは 3 時間も森の小道を歩いてその宿に着いた。
参照文	I got to the inn after walking three hours along the woodland path .
日本語変更文	わたしは着いたその宿に歩いて 3 時間も森の小道を。
日本語パターン	X1 は X9 た X7 X8 に X6 て X2 X3 も X4 の X5 を。
英語パターン	X1 X9 to X7 X8 after X6 X2 X3 along X4 X5 .
例 3	
入力文	机の上に本を広げたままどこかへ行ってしまった。
参照文	He is left his books open on the desk and gone off somewhere .
日本語変更文	本を広げたまま上に机の行ってしまったどこかへ。
日本語パターン	X3 を X4 たまま X2 に X1 の X6 てしまった X5 かへ。
英語パターン	He is left his X3 X4 X2 the X1 and X6 off X5 .

表 5.10: 重文複文の語順変更の失敗例

例 1	
入力文	彼は体の釣り合いを失ってその場に倒れた。
参照文	He lost his balance and fell down on the spot .
日本語変更文	彼は <u>体の</u> 失って釣り合いを倒れたその場に。
日本語パターン	X1 は <u>体の</u> X3 て X2 を X6 た X4 X5 に。
英語パターン	X1 X3 his X2 and fell X6 on X4 X5 .
例 2	
入力文	あいつが <u>そんな</u> 事をしたら私は黙っていない。
参照文	If he does <u>such</u> a thing , I can not let <u>it</u> go unchallenged .
日本語変更文	あいつがしたら事を私は黙っていない <u>そんな</u> 。
日本語パターン	X1 が X4 たら X3 を X5 は黙っていない <u>X2</u> 。
英語パターン	If X1 X4 <u>such</u> X3 , X5 can not let <u>X2</u> go unchallenged .
例 3	
入力文	蛇口からポトポトしたたる水音が彼の神経に <u>さわり</u> はじめた。
参照文	The water dripping from the faucet began to <u>jangle</u> on his nerves .
日本語変更文	蛇口からポトポトしたたる水音がはじめた彼の神経に <u>さわり</u> 。
日本語パターン	X1 からポトポトしたたる水音が X4 た X2 の X3 に <u>さわり</u> 。
英語パターン	The water dripping from the X1 X4 to <u>jangle</u> on X2 X3 .

語順変更失敗の原因を以下に示す。

- 例 1: 入力文の“体”に対応する単語が参照文になく、字面として残ったため、語順変更失敗した。
- 例 2: 変数 X2 が“そんな/such”ではなく、“そんな/it”で変数になってしまったので、語順変更失敗した。
- 例 3: 入力文の“さわり”に対応するのが参照文では“jangle”になり、対訳句辞書になかった。そのため、“さわり”が変数にならず、字面として残ったため、語順変更失敗した。

5.3.2 自動評価

自動評価の結果を表 5.11 に示す。

表 5.11: 重文複文の実験の自動評価結果

	BLEU	NIST	METEOR	RIBES	TER	WER
提案手法	0.1295	4.4601	0.4255	0.7181	0.7412	0.7565
ベースライン	0.1175	4.4218	0.4125	0.6749	0.7693	0.8027

5.3.3 人手評価

人手評価の結果を表 5.12 に示す.

表 5.12: 重文複文の実験の対比較評価結果

提案手法○	ベースライン○	差なし	同一出力
2	5	89	4

5.3.4 翻訳例

提案手法の方が優れている翻訳例を表 5.13, ベースラインの方が優れている翻訳例を表 5.14 に示す. なお, 下線部は間違いと判断した箇所を表す.

表 5.13: 提案手法の方が優れている翻訳例

例 1	
入力文	これらはわれわれの無視できない事実だ。
語順変更文	これらは事実だわれわれの無視できない。
提案手法○	These are the fact that we can not ignore .
ベースライン	<u>These are not our ignored .</u>
参照文	These are facts to which we cannot close our eyes .
例 2	
入力文	彼は体の釣り合いを失ってその場に倒れた。
語順変更文	彼は体の失って釣り合いを倒れたその場に。
提案手法○	He lost his balance and fell down on the spot .
ベースライン	He lost his balance and fell to the body .
参照文	He lost his balance and fell down on the spot .

表 5.14: ベースラインの方が優れている翻訳例

例 1	
入力文	彼女は私の遠い親戚だ。
語順変更文	彼女は遠い親戚だ私の。
提案手法	She <u>from my relatives</u> .
ベースライン○	She is my distant relative .
参照文	She is a distant relative of mine .
例 2	
入力文	その物音を聞いて彼らはおびえた。
語順変更文	その物音を聞いて彼らはおびえた。
提案手法	The noise <u>of their</u> frightened at the news .
ベースライン○	They were frightened at the noise .
参照文	The noise made them timorous .
例 3	
入力文	私なんぞがチームの代表だなんて分不相応だ。
語順変更文	私なんぞが代表だなんて分不相応だチームの。
提案手法	It is absurd that I am the representative of the team <u>不相応 minutes</u> .
ベースライン○	It is laughable that I am the representative of the team .
参照文	It is ridiculous that I am the representative of the team .

5.4 実験のまとめ

単文、重文複文の実験から以下のことが示せる。

1. 単文で語順変更失敗した10文中6文、重文複文で13文中9文は表5.3の例3のように、対訳文の単語対応に問題があったため、語順変更失敗した。これらの文は対訳文に問題があるので、改善することができない。また、単文で語順変更失敗した残りの4文、重文複文で語順変更失敗した残りの4文は表5.3の例1、2のように、対応する対訳句がない場合か間違った対応を取る対訳句が原因であった。これに関しては、対訳句を追加し、対訳句をクリーニングすることで改善できると考える。
2. 語順変更の精度で改善できるのは単文、重文複文で4文ずつであり、語順変更のプログラムはかなりの精度と言える。よって、語順変更の精度の向上による翻訳精度の向上は期待できない。

3. 自動評価では全ての評価において、提案手法の方がスコアが高い。しかし、人手評価では差は見られなかった。
4. 自動評価と人手評価の実験結果は一致しない。

第6章 考察

6.1 先行研究との比較

統計翻訳において、同じ語順間の翻訳では翻訳精度が高い傾向にある。しかし、異なる語順間の翻訳では翻訳精度が低い傾向にある。この問題を解決するために、日本語文の単語を並び替えて、英語の語順に近づけてから、統計翻訳を行う研究がされている。

岡崎らの研究では、自動評価と人手評価において、翻訳精度の向上は見られなかった。一方、星野らの研究では、自動評価でスコアの向上が見られた。しかし、人手評価では翻訳精度の向上は見られなかった。そこで、本研究では日本語文を英語の参照文の語順に並び替えてから、句に基づく統計翻訳を行った。実験の結果、自動評価ではスコアの向上が見られたが、人手評価では翻訳精度の向上が見られなかった。

これらの実験結果から、異なる語順間の翻訳で翻訳精度が低い原因は、語順が異なることではないことが分かり、原因が他にあると考える。今後、異なる語順間の翻訳において、翻訳精度を向上させる方法を調査する。

6.2 自動評価と人手評価

6.2.1 実験内容

本研究では、自動評価において、提案手法はベースラインよりスコアが高かったが、人手評価で差がないという結果になった。そこで、自動評価と人手評価で結果が異なる原因を調査した。実験は単文コーパスの提案手法の翻訳文を使用し、自動評価のスコア (METEOR, TER) と翻訳文を見て、自動評価ではスコアが高いが不自然な翻訳文が出力されている文を表 6.1、自動評価ではスコアが低い正しい翻訳文が出力されている文を表 6.2 に示す。なお、単文の提案手法での METEOR の平均値は 0.4617、TER の平均値は 0.6820 である。

表 6.1: スコアが高いが不自然な翻訳文の例

例 1	
スコア	METEOR : 0.8217 , TER : 0.3333
入力文	当地の気候は彼の性に合わなかった。
翻訳文	The climate here that he did not agree with me .
参照文	The climate here did not agree with him .
例 2	
スコア	METEOR : 0.9269 , TER : 0.3636
入力文	爆撃で町全体が焼け野原と化した。
翻訳文	The whole town was turned into a was reduced to ashes in the bombing .
参照文	The whole town was reduced to ashes in the bombing .
例 3	
スコア	METEOR : 0.7496 , TER : 0.8333
入力文	その事故は20人の命を奪った。
翻訳文	The accident took two of his life .
参照文	The accident claimed twenty lives .

表 6.2: スコアは低い正しい翻訳文の例

例 1	
スコア	METEOR : 0.09837 , TER : 1.500
入力文	そんなことはとても信じられない。
翻訳文	I can't believe such a thing .
参照文	That beggars belief .
例 2	
スコア	METEOR : 0.2307 , TER : 1.000
入力文	それは無理な注文というものだ。
翻訳文	It is an unreasonable request .
参照文	That's asking too much .
例 3	
スコア	METEOR : 0.2500 , TER : 0.6000
入力文	急速に衰弱している。
翻訳文	He quickly wasted .
参照文	He is failing fast .

6.2.2 実験のまとめ

自動評価と人手評価で結果が異なる原因として、自動評価のスコアの計算方法に問題があると考えます。自動評価のスコアの計算方法は、翻訳文と参照文を比較してスコアを計算している。そのため、翻訳文が正しい場合であっても、違う文法で書かれた文、参照文が間違っている文や意識されている文では、スコアが低くなることもある。また、自動評価のスコアは翻訳文を局所的に見て、スコアを計算している。語順変更により、英語の語順に並び替えてから、翻訳を行うことで、局所的には正しく翻訳されている。しかし、大局的に見れば、構文的に不自然な文が多く存在するため、人手評価では差がなかったと考える。また、表 6.1 の例 3 では METEOR は高いスコアを出しているが、TER は低いスコアを出している。このような文はよく見られる。これは表 6.2 の例 3 でも同じことが言え、自動評価の曖昧性が明らかである。

6.3 語順変更の有効性

6.3.1 実験内容

第 5.2.1 章と第 5.3.1 章の実験の語順変更成功した文を翻訳すると、ベースラインと比べて、翻訳精度が向上するのか調査した。実験は句に基づく統計翻訳を行い、提案手法とベースラインの対比較評価を行う。なお、“reordering-model”を用いる。また、語順変更失敗した文についても調査した。

6.3.2 単文の実験

語順変更成功した 90 文を翻訳し、ベースラインと比べて、翻訳精度が向上するのか調べた。語順変更成功した 90 文の対比較評価の結果を表 6.3 に示す。

表 6.3: 語順変更成功した 90 文の対比較評価結果

提案手法○	ベースライン○	差なし
7	4	79

また、提案手法の方が優れている翻訳例を表 6.4、ベースラインの方が優れている翻訳例を表 6.5 に示す。なお、下線部は間違いと判断した箇所を表す。

表 6.4: 提案手法の方が優れている翻訳例

例 1	
入力文	その 責めは私にある。
日本語変更文	私にある その 責めは。
提案手法○	I have the blame .
ベースライン	<u>The company persecuted is to me .</u>
参照文	I am responsible for it .
例 2	
入力文	軍隊にはいろいろの階級がある。
日本語変更文	いろいろの階級がある 軍隊には。
提案手法○	There is a lot of the rank of the army .
ベースライン	There are <u>all kinds of working</u> in the army .
参照文	There are various ranks in the military .
例 3	
入力文	東京から名古屋までぶっ通しに自動車を走らせた。
日本語変更文	自動車を走らせた東京から名古屋までぶっ通しに。
提案手法○	He drove his car from Tokyo to Nagoya Nonstop .
ベースライン	He drove his car Nonstop <u>driving was hard</u> from Tokyo to Nagoya .
参照文	He made a nonstop motorcar drive from Tokyo to Nagoya .

表 6.5: ベースラインの方が優れている翻訳例

例 1	
入力文	この風邪薬はよく効く。
日本語変更文	この風邪薬はよく効く。
提案手法	This medicine is effective <u>cold</u> .
ベースライン○	This cold medicine works well .
参照文	This cold medicine is effective .
例 2	
入力文	自動車が動かない。
日本語変更文	自動車が動かない。
提案手法	The car <u>stationary</u> .
ベースライン○	The car is not working .
参照文	The car will not start .
例 3	
入力文	そのひどい知らせにすっかり打ちひしがれた。
日本語変更文	すっかり打ちひしがれたそのひどい知らせに。
提案手法	I <u>was</u> completely <u>overwhelmed</u> the bad news .
ベースライン○	She was absolutely crushed by the bad news .
参照文	We were totally crushed by the terrible news .

次に、語順変更で失敗した 10 文を翻訳し、ベースラインと比べて、翻訳精度が低下するのかわかった。語順変更で失敗した 10 文の対比較評価の結果を表 6.6 に示す。

表 6.6: 語順変更で失敗した 10 文の対比較評価結果

提案手法○	ベースライン○	差なし
0	1	9

また、ベースラインの方が優れている翻訳例を表 6.7 に示す。なお、下線部は間違いと判断した箇所を表す。

表 6.7: ベースラインの方が優れている翻訳例

入力文	ぼくは純粋な動機からそう言っているのだ。
日本語変更文	ぼくは純粋な言っているのだ動機からそう。
提案手法	I am <u>those is</u> from the motives .
ベースライン○	I say that is from the neat motives .
参照文	I am speaking from a disinterested motive .

6.3.3 重文複文の実験

語順変更で成功した87文を翻訳し、ベースラインと比べて、翻訳精度が向上するのかわかった。語順変更で成功した87文の対比較評価の結果を表6.8に示す。

表 6.8: 語順変更で成功した87文の対比較評価結果

提案手法○	ベースライン○	差なし
1	5	81

また、提案手法の方が優れている翻訳例を表6.9、ベースラインの方が優れている翻訳例を表6.10に示す。なお、下線部は間違いと判断した箇所を表す。

表 6.9: 提案手法の方が優れている翻訳例

入力文	これらはわれわれの無視できない事実だ。
日本語変更文	これらは事実だわれわれの無視できない。
提案手法○	These are the fact that we can not ignore .
ベースライン	These are <u>not our ignored</u> .
参照文	These are facts to which we cannot close our eyes .

表 6.10: ベースラインの方が優れている翻訳例

例 1	
入力文	その物音を聞いて彼らはおびえた。
日本語変更文	その物音を聞いて彼らはおびえた。
提案手法	<u>The noise of their</u> frightened at the news .
ベースライン○	They were frightened at the noise .
参照文	The noise made them timorous .
例 2	
入力文	彼女は私の遠い親戚だ。
日本語変更文	彼女は遠い親戚だ私の。
提案手法	She <u>from my relatives</u> .
ベースライン○	She is my distant relative .
参照文	She is a distant relative of mine .
例 3	
入力文	彼女の帽子は実にへんてこなものだ。
日本語変更文	彼女の帽子は実にへんてこなものだ。
提案手法	Her hat is a very <u>へんてこ</u> .
ベースライン○	Her hat is weird .
参照文	Her hat is a horror .

次に、語順変更で失敗した 13 文を翻訳し、ベースラインと比べて、翻訳精度が低下するのかわかった。語順変更で失敗した 13 文の対比較評価の結果を表 6.11 に示す。

表 6.11: 語順変更で失敗した 13 文の対比較評価結果

提案手法○	ベースライン○	差なし
1	0	12

また、提案手法の方が優れている翻訳例を表 6.12 に示す。なお、下線部は間違いと判断した箇所を表す。

表 6.12: 提案手法の方が優れている翻訳例

入力文	彼は体の釣り合いを失ってその場に倒れた。
日本語変更文	彼は体の失ってその釣り合いを倒れた場に。
提案手法○	He lost his balance and fell down on the spot .
ベースライン	He lost his balance and <u>fell to the body</u> .
参照文	He lost his balance and fell down on the spot .

6.3.4 実験のまとめ

単文、重文複文ともに、語順変更で成功しても、翻訳精度の向上に影響しないことが分かった。また、語順変更で失敗した文に関しても、翻訳精度の低下に影響しないことが分かった。この実験から、語順変更は翻訳精度の向上に影響しないことが示された。

6.4 “reordering-model” の効果

6.4.1 実験内容

一般的に、統計翻訳では語順の情報として、“reordering-model” を使用している。しかし、本研究では語順変更の有効性が見られなかった。そのため、日英統計翻訳における“reordering-model” の効果を調査する。“reordering-model” を使用した実験 (“reordering-model あり”) と、使用しない実験 (“reordering-model なし”) のベースラインで比較実験

を行った。なお，“distortion-limit”の値は“-1”に設定した。

6.4.2 単文の実験

単文の実験の自動評価の結果を表 6.13, 対比較評価の結果を表 6.14 に示す。

表 6.13: 単文の実験の自動評価結果

	BLEU	NIST	METEOR	RIBES	TER	WER
“reordering-model あり”	0.1340	4.7326	0.4513	0.7024	0.7178	0.7456
“reordering-model なし”	0.1336	4.8281	0.4511	0.7044	0.7080	0.7346

表 6.14: 単文の実験の対比較評価結果

“reordering-model あり” ○	“reordering-model なし” ○	差なし	同一出力
4	1	73	22

また，“reordering-model あり”の方が優れている翻訳例を表 6.15, “reordering-model なし”の方が優れている翻訳例を表 6.16 に示す。なお、下線部は間違いと判断した箇所を表す。

表 6.15: “reordering-model あり”の方が優れている翻訳例

例 1	
入力文	この風邪薬はよく効く。
“reordering-model あり” ○	This cold medicine works well .
“reordering-model なし”	This medicine works well <u>cold</u> .
参照文	This cold medicine is effective .
例 2	
入力文	ぼくは純粋な動機からそう言っているのだ。
“reordering-model あり” ○	I say that is from the neat motives .
“reordering-model なし”	I <u>always neat say that from motive</u> .
参照文	I am speaking from a disinterested motive .
例 3	
入力文	空気は黒い煙でもうもうとしていた。
“reordering-model あり” ○	The air was filled with black smoke .
“reordering-model なし”	The air was filled with <u>of</u> black smoke .
参照文	The air was thick with black smoke .

表 6.16: “reordering-model なし”の方が優れている翻訳例

入力文	私は弟と三つ年が離れている。
“reordering-model あり”	I <u>have</u> three years apart with my brother .
“reordering-model なし” ○	My brother and I are three years apart .
参照文	My younger brother and I are three years apart in age .

6.4.3 重文複文の実験

重文複文の実験の自動評価の結果を表 6.17, 対比較評価の結果を表 6.18 に示す.

表 6.17: 重文複文の実験の自動評価結果

	BLEU	NIST	METEOR	RIBES	TER	WER
“reordering-model あり”	0.1175	4.4218	0.4125	0.6749	0.7693	0.8027
“reordering-model なし”	0.1140	4.3863	0.4092	0.6694	0.7778	0.8124

表 6.18: 重文複文の実験の対比較評価結果

“reordering-model あり” ○	“reordering-model なし” ○	差なし	同一出力
1	2	80	17

また, “reordering-model あり”の方が優れている翻訳例を表 6.19, “reordering-model なし”の方が優れている翻訳例を表 6.20 に示す. なお, 下線部は間違いと判断した箇所を表す.

表 6.19: “reordering-model あり”の方が優れている翻訳例

入力文	その物音を聞いて彼らはおびえた。
“reordering-model あり” ○	They were frightened at the noise .
“reordering-model なし”	They <u>lived</u> to hear the noise .
参照文	The noise made them timorous .

表 6.20: “reordering-model なし”の方が優れている翻訳例

例 1	
入力文	彼は体の釣り合いを失ってその場に倒れた。
“reordering-model あり”	He lost his balance and fell <u>to the body</u> .
“reordering-model なし” ○	He lost his balance and fell down on the spot .
参照文	He lost his balance and fell down on the spot .
例 2	
入力文	あいつがそんな事をしたら私は黙っていない。
“reordering-model あり”	<u>His resignation , I do not to do so</u> .
“reordering-model なし” ○	He is such a thing , but I do not keep silent .
参照文	If he does such a thing , I can not let it go unchallenged .

6.4.4 実験のまとめ

単文、重文複文ともに、自動評価で差が見られなかった。また、人手評価では“差なし”と“同一出力”が大半を占めていることから、“reordering-model”の効果が見られないことが分かる。

6.5 日本語文の語順をランダムに変えた実験

6.5.1 実験内容

第 6.3 章、第 6.4 章より、語順変更が翻訳精度の向上にあまり影響していないことが分かった。そこで、語順変更による翻訳精度の向上を検証するために、日本語文の語順を英語の語順に並び替えるのではなく、日本語文の語順をランダムに並べ替える。その後、句に基づく統計翻訳を行い、語順変更の効果を調査した。

6.5.2 検証手法

検証手法の手順を以下に示す。なお、手順 1 と手順 3 は第 3.2 章の提案手法と同様に行う。

手順1 変数化

対訳句辞書と対訳文を用いて、学習文、テスト文、ディベロップメント文に変数化を行う。対訳句辞書と対訳文を比べ、共通する箇所を変数にする。変数化の手順を図6.1に示す。

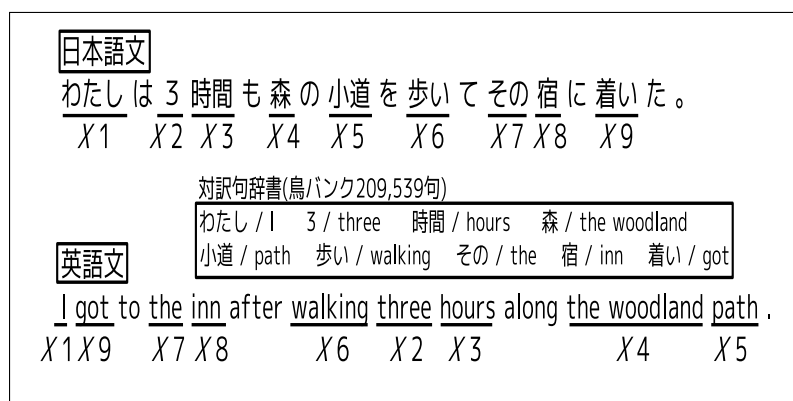


図 6.1: 変数化

手順2 ランダムに並び替え

対訳文を用いて、学習文、テスト文、ディベロップメント文の日本語文をランダムに並び替え、出力を日本語変更文とする。なお、移動単位は次の変数までとする。語順をランダムに並び替える手順を図6.2に示す。

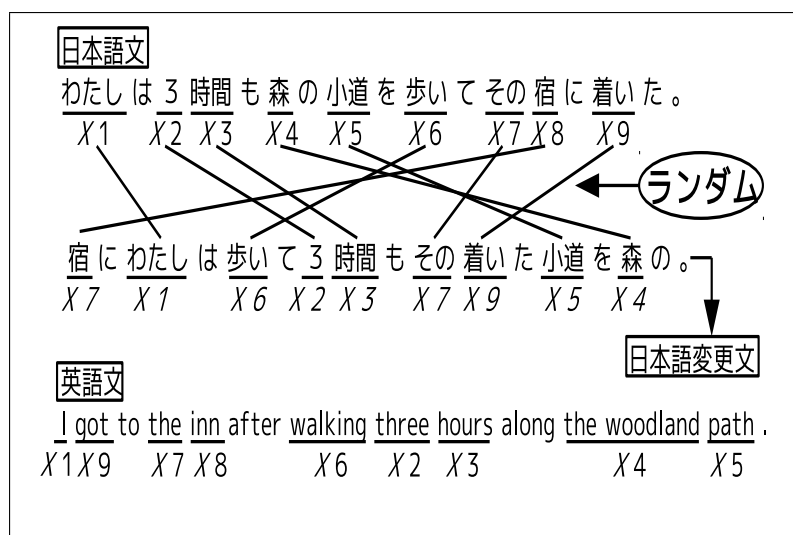


図 6.2: ランダムに並び替え

手順3 翻訳実験

学習文，テスト文，ディベロップメント文にランダムに並び替えたデータを用いて，句に基づく統計翻訳を行う．なお，“reordering-model”は用いない．

6.5.3 日本変更文の例

語順をランダムに並び替えた単文の日本語変更文の例を表 6.21，重文複文の日本語変更文の例を表 6.22 に示す．

表 6.21: 単文の日本語変更文の例

例 1	
入力文	新車の性能の試験をした。
参照文	The performance test of new cars was held .
日本語変更文	した性能の試験を新車の。
日本語パターン	X4 た X2 の X3 を X1 の。
英語パターン	The X2 X3 of X1 was X4 .
例 2	
入力文	この部屋は我々が去年泊まった部屋そのものです。
参照文	This is the identical room we stayed in last year .
日本語変更文	我々が部屋は去年泊まったこのそのものです。
日本語パターン	X3 が X2 は X4 X5 た X1 X6 です。
英語パターン	X1 is the X6 X2 X3 X5 in X4 year .
例 3	
入力文	彼はやっとの思いで、その少女に自分の気持ちを打ち明けた。
参照文	He finally got up the courage to reveal his feelings to the girl .
日本語変更文	思いで、打ち明けた少女に自分の気持ちをやっとのその彼は。
日本語パターン	X3 で、 X6 た X5 に自分の気持ちを X2 の X4 X1 は。
英語パターン	X1 X2 got up the courage to X6 his X3 to X4 X5 .

表 6.22: 重文複文の日本語変更文の例

例 1	
入力文	私の 今日 ある の は あなたの おかげ です 。
参照文	I owe what I am today to you .
日本語変更文	私 の ある の は 今日 あなたの おかげ です 。
日本語パターン	X1 の X3 の は X2 X4 の おかげ です 。
英語パターン	X1 X3 what I am X2 to X4 .
例 2	
入力文	子ども を 健全 に 育てる ため の 環境 作り が 望まれる 。
参照文	Providing good surroundings to raise children soundly is desirable .
日本語変更文	育てる ため の 作り が 環境 子ども を 望まれる 健全 に 。
日本語パターン	X3 ため の X5 が X4 X1 を X6 れる X2 。
英語パターン	X5 good X4 to X3 X1 X2 is X6 .
例 3	
入力文	自動車 を 動かす に は まず エンジン を かけ なければ なり ませ ん 。
参照文	In order to move a car , first you have to start the engine .
日本語変更文	動かす に は エンジン を まず かけ なければ なり ませ ん 自動車 を 。
日本語パターン	X2 に は X4 を X3 X5 なければ なり ませ ん X1 を 。
英語パターン	In order to X2 X1 , X3 you X5 to start the X4 .

6.5.4 単文の実験

自動評価の結果を表 6.23 に示す。

表 6.23: 単文の実験の自動評価結果

	BLEU	NIST	METEOR	RIBES	TER	WER
検証手法	0.1271	4.7223	0.4468	0.7010	0.7197	0.7467
ベースライン	0.1341	4.8670	0.4560	0.7040	0.7047	0.7327

検証手法とベースラインの対比較評価の結果を表 6.24 に示す。また、検証手法の方が優れている翻訳例を表 6.25，ベースラインの方が優れている翻訳例を表 6.26 に示す。なお，下線部は間違いと判断した箇所を表す。

表 6.24: 検証手法とベースラインの対比較評価結果

検証手法○	ベースライン○	差なし	同一出力
5	3	79	13

表 6.25: 検証手法の方が優れている翻訳例

例 1	
入力文	この風邪薬はよく効く。
日本語変更文(検証)	効く薬はよくこの風邪。
検証手法○	This medicine works very well for a cold .
ベースライン	This medicine works well cold .
参照文	This cold medicine is effective .
例 2	
入力文	いつでも喜んでお手伝いします。
日本語変更文(検証)	いつでも喜んでお手伝いします。
検証手法○	I will be glad to help you any time .
ベースライン	I help you any time .
参照文	I'm always ready to help .
例 3	
入力文	空気は黒い煙でもうもうとしていた。
日本語変更文(検証)	黒い空気は煙でもうもうとしていた。
検証手法○	The air was filled with black smoke .
ベースライン	The air was filled with of black smoke .
参照文	The air was thick with black smoke .

表 6.26: ベースラインの方が優れている翻訳例

例 1	
入力文	我が社には優秀な技術者が居ります。
日本語変更文(検証)	我が社には優秀な技術者が居ります。
検証手法	The 居り engineers a brilliant in our company .
ベースライン○	There is a brilliant engineers in our company .
参照文	Our company has some talented technicians .
例 2	
入力文	その試合はテレビで見ました。
日本語変更文(検証)	その試合は見ましたテレビで。
検証手法	The game I saw on television .
ベースライン○	I saw the match on TV .
参照文	I watched the game on television .
例 3	
入力文	私は弟と三つ年が離れている。
日本語変更文(検証)	三つ弟と離れている年が私は。
検証手法	My brother is and three years .
ベースライン○	My brother and I are three years apart .
参照文	My younger brother and I are three years apart in age .

6.5.5 重文複文の実験

自動評価の結果を表 6.27 に示す。

表 6.27: 重文複文の実験の自動評価結果

	BLEU	NIST	METEOR	RIBES	TER	WER
検証手法	0.1016	4.2077	0.3978	0.6613	0.7848	0.8191
ベースライン	0.1128	4.3727	0.4075	0.6651	0.7830	0.8174

検証手法とベースラインの対比較評価の結果を表 6.28 に示す。また、検証手法の方が優れている翻訳例を表 6.29, ベースラインの方が優れている翻訳例を表 6.30 に示す。なお、下線部は間違いと判断した箇所を表す。

表 6.28: 重文複文の実験の対比較評価結果

検証手法○	ベースライン○	差なし	同一出力
6	6	86	2

表 6.29: 検証手法の方が優れている翻訳例

例 1	
入力文	彼は彼女と行くと自発的に申し出た。
日本語変更文(検証)	彼は彼女と行くと自発的に申し出た。
検証手法○	He volunteered to go with her .
ベースライン	He voluntarily going with her .
参照文	He volunteered to go with her .
例 2	
入力文	またお会いできて光栄です。
日本語変更文(検証)	お会いできて光栄ですまた。
検証手法○	It is a great honor to see you again .
ベースライン	It is a great honor and see you again .
参照文	Great to see you again .
例 3	
入力文	これは僕が設計して建てた家だ。
日本語変更文(検証)	建てた僕が家だこれは設計して。
検証手法○	This is the house where I was built design .
ベースライン	This house was built to design of me .
参照文	I planned and built this house .

表 6.30: ベースラインの方が優れている翻訳例

例 1	
入力文	彼女の帽子は実にへんてこなものだ。
日本語変更文 (検証)	彼女の帽子は実にへんてこなものだ。
検証手法	The へんてこ of her hat .
ベースライン○	Her hat is really weird .
参照文	Her hat is a horror .
例 2	
入力文	日が出て霧を払う。
日本語変更文 (検証)	日が出て霧を払う。
検証手法	The day for the fog .
ベースライン○	The sun dispels the fog .
参照文	The sun dispels the mist .
例 3	
入力文	私は失敗して目が覚めた。
日本語変更文 (検証)	私は失敗して目が覚めた。
検証手法	I woke up , I failed .
ベースライン○	I woke with my failure .
参照文	My failure brought me to my senses .

6.5.6 実験のまとめ

単文、重文複文ともに、自動評価では検証手法は他の2手法よりスコアが低下しているが、人手評価では差がないという結果になった。この結果より、日本語文を英語の語順に並び替えても、ランダムに並び替えても、翻訳精度に差がなく、語順変更の影響はあまり見られないことが分かった。

6.6 鳥バンクのパターンを使用した実験

6.6.1 実験内容

本研究は自分で作成したプログラムで日本語パターンを作成し、語順変更を行った。しかし、鳥バンクには人手で作成された日本語パターンが存在する。そこで、自分で作成したプログラムの日本語パターン(以下、自動作成パターン)と鳥バンクの日本語パター

ン(以下, 人手作成パターン)では, どちらの方が語順変更を行うのに適しているのか調査した. なお, 実験は重文複文コーパスを使用する.

6.6.2 人手作成パターンの例

鳥バンクの人手作成パターンの例を表 6.31 に示す.

表 6.31: 人手作成パターンの例

例 1	
入力文	その知らせを聞いて彼女の顔は明るくなった。
人手作成パターン	その N1 を N2 て N3 の N4 は N5 くなった。
参照文	Her face lit up when she heard the news.
例 2	
入力文	ぼくの趣味はいろいろな国の切手を集めることです。
人手作成パターン	N1 の N2 は N3 N4 の N5 を N6 ことです。
参照文	My hobby is collecting stamps from different countries.
例 3	
入力文	これは人々に愛唱されている古い民謡の一つです。
人手作成パターン	N1 は N2 に N3 れている N4 N5 の N6 です。
参照文	This is one of the old folk songs people love to sing .

6.6.3 実験結果

鳥バンクの重文複文コーパスからランダムに 50 文抽出し, 自動作成パターンと人手作成パターン(単語レベル)で対比較評価を行った. なお, 2つの日本語変更文を比べ, 参照文の語順により近づいた方を優れていると判断した. 対比較評価の結果を表 6.32 に示す.

表 6.32: 対比較評価結果

自動作成パターン	人手作成パターン	差なし	同一出力
19	4	13	14

自動作成パターンの方が優れている例を表 6.33, 人手作成パターンの方が優れている例を表 6.34 に示す.

表 6.33: 自動作成パターンの方が優れている例

例 1	
入力文	わたしは食事のマナーを知らなかったので赤恥をかいた。
参照文	I was humiliated for having no knowledge of table manners.
日本語パターン(自動)	X1はX2のX3をX4なかったのでX5た。
日本語変更文(自動)	わたしは赤恥をかいた知らなかったので食事のマナーを。
日本語パターン(人手)	X1はX2のX3をX4なかったので赤恥をかいた。
日本語変更文(人手)	わたしは知らなかったので赤恥をかいた食事のマナーを。
例 2	
入力文	詩を味わうためには声を出して読むべきだ。
参照文	In order to appreciate a poem, you should read it aloud.
日本語パターン(自動)	X1をX2ためにはX3を出してX4べきだ。
日本語変更文(自動)	味わうためには詩を読むべきだ声を出して。
日本語パターン(人手)	X1をX2ためには声を出してX3べきだ。
日本語変更文(人手)	味わうためには声を出して詩を読むべきだ。
例 3	
入力文	この本はロシアの民話を集めたものです。
参照文	This is a collection of Russian folktales.
日本語パターン(自動)	X1本はX2のX3をX4たものです。
日本語変更文(自動)	この本は集めたものですロシアの民話を。
日本語パターン(人手)	X1本はX2のX3を集めたものです。
日本語変更文(人手)	この本はロシアの民話を集めたものです。

表 6.34: 人手作成パターンの方が優れている例

例 1	
入力文	会社にいる間は愛想よくするようにしている。
参照文	I try to make myself agreeable while in the office.
日本語パターン(自動)	X1 にいる間は愛想よくするようにしている。
日本語変更文(自動)	会社にいる間は愛想よくするようにしている。
日本語パターン(人手)	X1 にいる X2 は愛想 X3 X4 ように X5 ている。
日本語変更文(人手)	しているするように愛想よく間は会社にいる。
例 2	
入力文	これは人々に愛唱されている古い民謡の一つです。
参照文	This is one of the old folk songs people love to sing.
日本語パターン(自動)	X1 は X2 に X3 れて X4 X5 X6 の X7 です。
日本語変更文(自動)	これはいる一つです古い民謡の人々に愛唱されて。
日本語パターン(人手)	X1 は X2 に X3 れている X4 X5 の X6 です。
日本語変更文(人手)	これは一つです古い民謡の人々に愛唱されている。
例 3	
入力文	彼は青信号になるとアクセルをぐいと踏み込んだ。
参照文	He stomped on the gas pedal when the green light came on.
日本語パターン(自動)	X1 は青信号になると X2 をぐいと踏み込んだ。
日本語変更文(自動)	彼は青信号になるとアクセルをぐいと踏み込んだ。
日本語パターン(人手)	X1 は X2 になると X3 をぐいと踏み込んだ。
日本語変更文(人手)	彼はアクセルをぐいと踏み込んだ青信号になると。

6.6.4 実験のまとめ

自動作成パターンと人手作成パターンを比べると、自動作成パターンの方が、より英語文の語順に近づけるのに、優れていることが分かった。その理由として、自動作成パターンの方が変数の数が多いため、語順変更を行うには、自動作成パターンの方が効果的であった。

6.7 “distortion-limit”の値を変えた実験

6.7.1 実験内容

日本語から英語への翻訳では動詞の位置が大きく変化する．そのため，フレーズの並び替えの重みである“distortion-limit”の値を変えて，最適な“distortion-limit”の値を調査した．翻訳には句に基づく統計翻訳を行う．なお，“reordering-model”を用いる．

6.7.2 単文の実験

提案手法とベースラインの自動評価の結果を表6.35に示す．なお，括弧内の数字は”distortion-limit”の値を表し，下線部は評価法ごとの上位3スコアを表している．

表 6.35: 単文の自動評価結果

	BLEU	NIST	METEOR	RIBES	TER	WER
提案手法 (0)	0.1235	4.7473	0.4459	0.7273	0.6961	0.7115
提案手法 (1)	0.1235	4.7473	0.4459	0.7273	0.6961	0.7115
提案手法 (2)	0.1271	4.7613	0.4475	0.7286	0.6964	0.7117
提案手法 (4)	0.1338	4.7533	0.4509	0.7291	0.7048	0.7201
提案手法 (6)	0.1418	<u>4.9383</u>	<u>0.4625</u>	<u>0.7371</u>	<u>0.6798</u>	<u>0.6951</u>
提案手法 (8)	<u>0.1445</u>	<u>4.8564</u>	<u>0.4624</u>	<u>0.7323</u>	<u>0.6896</u>	<u>0.7062</u>
提案手法 (10)	<u>0.1435</u>	4.8510	0.4599	<u>0.7320</u>	0.6939	0.7112
提案手法 (12)	<u>0.1441</u>	<u>4.9375</u>	<u>0.4637</u>	0.7315	<u>0.6833</u>	<u>0.7012</u>
提案手法 (-1)	0.1413	4.7984	0.4556	0.7268	0.7027	0.7206
ベースライン (6)	0.1233	4.7121	0.4396	0.7005	0.7114	0.7405
ベースライン (-1)	0.1377	4.8628	0.4566	0.7106	0.6969	0.7230

6.7.3 重文複文の実験

提案手法とベースラインの自動評価の結果を表6.36に示す。なお、括弧内の数字は”distortion-limit”の値を表し、下線部は評価法ごとの上位3スコアを表している。

表 6.36: 自動評価結果

	BLEU	NIST	METEOR	RIBES	TER	WER
提案手法 (0)	0.1176	4.3081	0.4100	0.7098	0.7491	0.7623
提案手法 (1)	0.1176	4.3081	0.4100	0.7098	0.7491	0.7623
提案手法 (2)	0.1192	4.3151	0.4089	0.7082	0.7534	0.7661
提案手法 (4)	0.1245	4.3282	0.4135	0.7101	0.7525	0.7668
提案手法 (6)	0.1295	4.4601	0.4255	<u>0.7181</u>	<u>0.7412</u>	<u>0.7565</u>
提案手法 (8)	<u>0.1310</u>	<u>4.4977</u>	<u>0.4274</u>	<u>0.7196</u>	<u>0.7401</u>	<u>0.7560</u>
提案手法 (10)	<u>0.1314</u>	<u>4.5233</u>	<u>0.4291</u>	<u>0.7123</u>	<u>0.7449</u>	<u>0.7644</u>
提案手法 (12)	<u>0.1297</u>	4.4569	0.4260	0.7112	0.7547	0.7747
提案手法 (-1)	0.1289	<u>4.4643</u>	<u>0.4271</u>	0.7076	0.7496	0.7695
ベースライン (6)	0.0991	4.1837	0.3885	0.6493	0.8028	0.8437
ベースライン (-1)	0.1175	4.4218	0.4125	0.6749	0.7693	0.8027

6.7.4 実験のまとめ

提案手法において、単文では”distortion-limit”の値を”6”または”8”，重文複文では”distortion-limit”の値を”8”または”10”に設定して、実験を行うことが最適であることを確認した。また、ベースラインでは単文，重文複文ともに，”distortion-limit”の値を”-1”に設定して，実験を行うことが最適である。

6.8 語順変更の研究の今後

統計翻訳において，語順変更の研究が盛んに行われてきた [1][2]。しかし，これらの研究は目標とする英語文の最適な語順に並び替えられていない。また，翻訳精度の大きな向上が見られなかった。そこで，本研究では目標とする英語文の最適な語順に並び替えてから，句に基づく統計翻訳を行った。しかし，翻訳精度は向上しなかった。このことから，語順変更では翻訳精度の大きな向上が見られないと考える。

第7章 おわりに

統計翻訳において、語順変更をしてから、統計翻訳をする研究が盛んに行われてきた。しかし、これらの研究は翻訳される英語文の語順を推定して、日本語文の語順変更を行っている。そのため、目標とする英語文の最適な語順に並び替えられていない。

そこで、本研究では語順変更の手法の限界を調査するため、テスト文に参照文があると仮定し、日本語文を英語の参照文の語順に並び替える。その後、句に基づく統計翻訳を行い、語順変更の効果を調査した。

実験の結果、人手評価において、単文では提案手法の方が優れている文が100文中7文、ベースラインの方が優れている文が100文中5文という結果になった。また、重文複文では提案手法の方が優れている文が100文中2文、ベースラインの方が優れている文が100文中5文という結果になり、単文と重文複文ともに差は見られなかった。このことから、語順変更では翻訳精度の大きな向上が見られないと考える。今後は語順変更以外の手法で、翻訳精度を向上させる手法を調査する。

謝辞

最後に、本研究の御指導をいただきました鳥取大学工学研究科情報エレクトロニクス専攻計算機工学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ともに厚くお礼を申し上げます。また，ご多忙の中，助言を頂きました徳久雅人講師に厚くお礼申し上げます。そして，計算機工学C講座研究室の皆様，参考にさせて頂いた論文の著者の方々に対して，深く感謝します。

参考文献

- [1] Chenchen Ding, Masao Utiyama, Eiichiro Sumita, Mikio Yamamoto: Word Order Does NOT Differ Significantly Between Chinese and Japanese. WAT2014, pp.77-82, 2014
- [2] 村松航平, Kevin Duh, 松本裕治: 事前並べ替えを利用したヒンディー語—英語統計的機械翻訳. 言語処理学会第 21 回年次大会, pp.968-971, 2015
- [3] 岡崎弘樹, 村上仁一, 徳久雅人, 池原悟: 日本語文法構造の変換による日英統計翻訳. 言語処理学会第 15 回年次大会, pp.240-243, 2009
- [4] 星野翔, 宮尾祐介, 須藤克仁, 永田昌明: 日英統計的機械翻訳のための述語項構造に基づく事前並べ替え. 言語処理学会第 19 回年次大会, pp.394-397, 2013
- [5] Philipp Koehn, Franz Josef Och, Daniel Marcu: Statistical Phrase-Based Translation. Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology (HLT-NAACL 2003), pp.48-54, 2003
- [6] Peter F Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, Robert L. Mercer: The Mathematics of Statistical Machine Translation, Computational Linguistics, 1993
- [7] GIZA++, <http://www.fjoch.com/GIZA++>
- [8] Tillmann Christoph and Zhang Tong: A Localized Prediction Model for Statistical Machine Translation, Annual Meeting of the Association for Computational Linguistics, pp.557-564, 2005
- [9] Philipp Koehn, Franz Josef Och, Daniel Marcu: Statistical Phrase-Based Translation, Association for Computational Linguistic, pp.48-54, 2003

- [10] Franz Josef Och: Minimum Error Rate Training in Statistical Machine Translation. Proceeding of the 41st Annual Meeting of the Association for Computational Linguistics, pp.160-167, 2003
- [11] 畑中勇輝, 村上仁一, 徳久雅人: 文法構造変換および句構造変換を用いた統計翻訳. 言語処理学会第20回年次大会, pp.804-807, 2014
- [12] 鳥バンク: <http://unicorn.ike.tottori-u.ac.jp/toribank/>
- [13] 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
- [14] Taku Kudo, Yuji Matsumoto: Japanese Dependency Analysis using Cascaded Chunking. CoNLL 2002: Proceedings of the 6th Conference on Natural Language Learning 2002 (COLING 2002 Post-Conference Workshops), pp.63-69, 2002
- [15] Andreas Stolcke: SRILM - an Extensible Language Modeling Toolkit. Proceedings of the 7th International Conference on Spoken Language Processing, pp.901-904, 2002
- [16] Philipp Koehn, Marcello Federico, Brooke Cowan, Richard Zens, Chris Dyer, Alexandra Constantin, Evan Herbst: Moses: Open Source Toolkit for Statistical Machine Translation. moses.2007-05-29.tgz, <http://www.statmt.org/moses/>, Proceedings of the ACL 2007 Demo and Poster Sessions, pp.177-180, 2007
- [17] Papineni Kishore, Salim Roukos, Todd Ward, Wei-Jing Zhu: BLEU, NIST: A Method for Automatic Evaluation of Machine Translation. Proceedings of the 40th Annual meeting of the Association for Computational Linguistics, pp.311-318, 2002
- [18] Banerjee Satanjeev, Lavie Alon: METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments. Proceedings of Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization at the 43th Annual Meeting of the Association of Computational Linguistics (ACL-2005), pp.65-72, 2005
- [19] 平尾努, 磯崎秀樹, Kevin Duh, 須藤克仁, 塚田元, 永田昌明: RIBES: 順位相関に基づく翻訳の自動評価法. 言語処理学会第17年次大会, pp.1111-1114, 2011

- [20] Gregor Leusch, Nicola Ueffing, Hermann Ney: A Novel String-to-String Distance Measure with Applications to Machine Translation Evaluation. Proceeding of MT Summit IX, pp.240-247, 2003