

概要

統計翻訳は、異なる文法構造間の翻訳において、低い翻訳精度を得る傾向がある。異なる文法構造間の翻訳は、翻訳する言語から目的言語へ大きく文法構造が変化するためである。異なる文法構造間の翻訳精度を向上するために、例えば、日本語の文法構造を英語の文法構造に変換して統計翻訳を行う方法が存在する。

岡崎 [1] 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) の順に並び替えただけでは翻訳精度は向上しなかったことが報告されている。一方、星野ら [2] によると、述語項構造に基づく事前並べ替えを行った後で、句に基づく統計翻訳 (PSMT) を行うことで、翻訳精度が向上したことが報告されている。星野らは特許文を使用している。そして、人手評価を行っていない。しかし、特許文は文法構造が複雑で、人手評価が困難である。

そこで、本研究では簡単な文法構造である単文コーパス [3] を使用し、動詞および目的語および助詞の文法構造変換を行い、翻訳精度の変化を調べた。

実験の結果、動詞の文法構造変換と助詞の文法構造変換の組み合わせは有効だが、目的語の文法構造変換は有効でなかった。また、HSMT において動詞の文法構造変換と助詞の文法構造変換の組み合わせが、最も高い翻訳精度であることが示された。

今後、提案手法の翻訳精度を向上するために、複合動詞や連続した助詞を処理する必要がある。

目次

第1章	はじめに	1
第2章	日英統計翻訳システム	2
2.1	統計翻訳とは	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	7
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	デコーダー	12
2.6	パラメータチューニング	12
第3章	日本語文法構造の変換による翻訳	14
3.1	過去の研究	14
3.1.1	動詞の文法構造変換 [1]	14
3.1.2	述語項構造に基づく事前並べ替え [2]	14
3.2	提案手法	15
3.2.1	提案手法の概要	15

3.2.2	本研究で用いる文法	17
3.2.3	主語の定義	17
3.2.4	動詞の定義	17
3.2.5	目的語の定義	18
3.2.6	助詞の定義	19
第4章	実験環境	20
4.1	対訳データ	20
4.1.1	通常 of データ	20
4.1.2	提案手法 of データ	21
4.2	翻訳モデル of 学習	21
4.2.1	翻訳モデル学習 of 概要	21
4.3	言語モデル of 学習	21
4.4	デコーダーに関するパラメータ	22
4.5	評価方法	22
第5章	翻訳実験	23
5.1	実験内容	23
5.2	実験条件	25
5.3	実験結果	25
5.3.1	人手評価	25
5.3.2	自動評価	26
5.3.3	実験結果 of まとめ	26
5.3.4	PSMT での翻訳例	27
5.3.5	HSMT での翻訳例	35
第6章	考察	42
6.1	星野らの研究 [2] との比較	42
6.2	PSMT と HSMT の対比較	42
6.2.1	対比較 of 結果	42
6.2.2	対比較 of まとめ	43
6.2.3	文法構造変換 of 精度確認	43
6.2.4	翻訳例	43

6.3	実験の問題点	47
6.3.1	複合動詞	47
6.3.2	連続した助詞	47
6.4	OSV(目的語, 主語, 動詞) への文法構造変換	47
6.4.1	概要	47
6.4.2	実験内容	48
6.4.3	人手評価	49
6.4.4	自動評価	50
6.4.5	評価結果のまとめ	50
6.4.6	PSMT での翻訳例	50
6.4.7	HSMT での翻訳例	53
第7章 おわりに		55

目 次

2.1	日英統計翻訳の枠組	3
2.2	アライメントの例	5
2.3	デコーダーにおける翻訳候補の生成例	12
3.1	動詞の文法構造変換の変換例	14
3.2	述語項構造に基づく事前並べ替えの変換例	15
3.3	目的語の文法構造変換の変換例	16
3.4	助詞の文法構造変換の変換例	16
3.5	動詞+目的語+助詞の文法構造変換の変換例	16
6.1	OSV(目的語, 主語, 動詞) への文法構造変換の翻訳	49

表 目 次

3.1	動詞の定義	18
4.1	対訳データの文数	20
5.1	PSMT での人手評価	25
5.2	HSMT での人手評価	25
5.3	PSMT の翻訳精度	26
5.4	HSMT の翻訳精度	26
5.5	動詞単独が優れている翻訳例	28
5.6	ベースラインが優れている翻訳例	28
5.7	目的語単独が優れている翻訳例	29
5.8	ベースラインが優れている翻訳例	29
5.9	助詞単独が優れている翻訳例	30
5.10	ベースラインが優れている翻訳例	30
5.11	動詞+目的語が優れている翻訳例	31
5.12	ベースラインが優れている翻訳例	31
5.13	動詞+助詞が優れている翻訳例	32
5.14	ベースラインが優れている翻訳例	32
5.15	目的語+助詞が優れている翻訳例	33
5.16	ベースラインが優れている翻訳例	33
5.17	動詞+目的語+助詞が優れている翻訳例	34
5.18	ベースラインが優れている翻訳例	34
5.19	動詞単独が優れている翻訳例	35
5.20	ベースラインが優れている翻訳例	35
5.21	目的語単独が優れている翻訳例	36
5.22	ベースラインが優れている翻訳例	36
5.23	助詞単独が優れている翻訳例	37

5.24	ベースラインが優れている翻訳例	37
5.25	動詞+目的語が優れている翻訳例	38
5.26	ベースラインが優れている翻訳例	38
5.27	動詞+助詞が優れている翻訳例	39
5.28	ベースラインが優れている翻訳例	39
5.29	目的語+助詞が優れている翻訳例	40
5.30	ベースラインが優れている翻訳例	40
5.31	動詞+目的語+助詞が優れている翻訳例	41
5.32	ベースラインが優れている翻訳例	41
6.1	PSMT と HSMT の対比較	42
6.2	文法構造変換の精度	43
6.3	正しく出力されなかった文	43
6.4	PSMT が優れている翻訳例	44
6.5	HSMT が優れている翻訳例	44
6.6	PSMT が優れている翻訳例	45
6.7	HSMT が優れている翻訳例	45
6.8	PSMT が優れている翻訳例	46
6.9	HSMT が優れている翻訳例	46
6.10	複合動詞の一部のみが並び替えされる例	47
6.11	連続した助詞の出力例	47
6.12	主語の文法構造変換単独の人手評価	49
6.13	主語+目的語+助詞の文法構造変換単独の人手評価	49
6.14	PSMT の翻訳精度	50
6.15	HSMT の翻訳精度	50
6.16	主語単独が優れている翻訳例	51
6.17	ベースラインが優れている翻訳例	51
6.18	主語+目的語+助詞が優れている翻訳例	52
6.19	ベースラインが優れている翻訳例	52
6.20	主語単独が優れている翻訳例	53
6.21	ベースラインが優れている翻訳例	53
6.22	主語+目的語+助詞が優れている翻訳例	54

6.23	ベースラインが優れている翻訳例	54
------	---------------------------	----

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欠点がある。この欠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、あまり発展されなかった。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、語彙や文法構造が似ているため翻訳精度が高い傾向がある。しかし、日本語から英語へ翻訳する場合、語彙や文法構造が異なるため、翻訳精度が低い傾向がある。

その問題を解決するために、日本語の単語を並び替えて、英語の文法構造に近づけてから、統計翻訳を行なう研究がされている。

岡崎 [1] 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) の順に並び替えただけでは翻訳精度は向上しなかったことが報告されている。一方、星野ら [2] によると、述語項構造に基づく事前並べ替えを行った後で、句に基づく統計翻訳 (PSMT) を行うことで、翻訳精度が向上したことが報告されている。星野らは特許文を使用している。そして、人手評価を行っていない。しかし、特許文は文法構造が複雑で、人手評価が困難である。

そこで、本研究では簡単な文法構造である単文コーパス [3] を使用し、星野らの論文の検証を行う。

第2章 日英統計翻訳システム

2.1 統計翻訳とは

統計翻訳とは、翻訳する言語の文とその対訳文を大量に収集した対訳データを用いて統計的モデルを学習し、自動的に翻訳システムを構築する機械翻訳の一つである。統計翻訳が提案される前の機械翻訳には、ルールベース翻訳がある。これは、例えば、「S → NP VP」という英語のルールがあれば、「S → NP は VP」という日本語のルールに翻訳するように、翻訳のルールを定義する必要がある。このため、多言語化が困難であり、開発コストが高いという問題がある。しかし、統計翻訳において、統計的モデルを学習するアルゴリズムは、言語に依存しないため、大量の対訳データがあれば、多言語化が容易であり、開発コストを抑えることが可能である。

統計翻訳は、現在は句に基づく統計翻訳 (Phrase-based Stastical Machine Translation, PSMT)[4] が一般的である。句に基づく統計翻訳は、文脈に基づく訳語の選択や局所的な語順の並び替えが、語に基づく統計翻訳 (Word-based Stastical Machine Translation, WSMT) より優れているためである。しかし、句に基づく統計翻訳は、局所的な句の相対位置に依存しているため、大局的な語順の並び替えは困難であり、日英統計翻訳をはじめとした異なる文法構造間の翻訳精度は低い傾向がある。

2.2 日英統計翻訳における基本モデル

日英統計翻訳システムの枠組を図 2.1 に示す。

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。

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

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

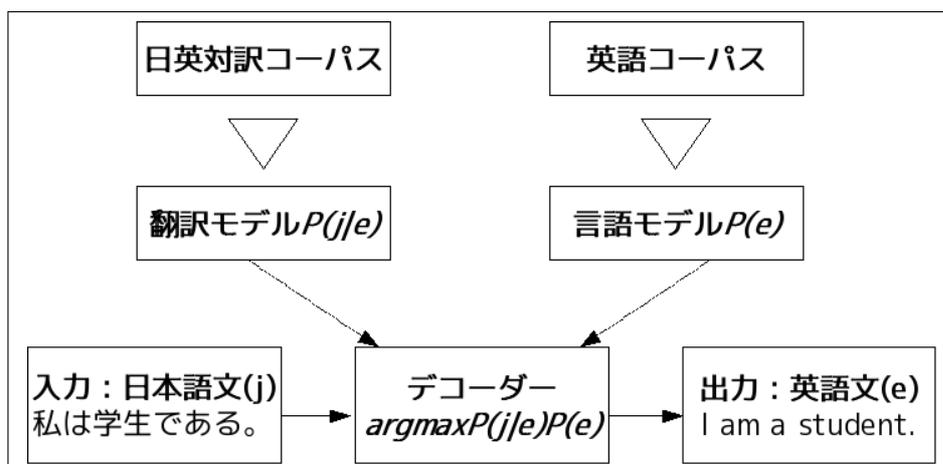


図 2.1: 日英統計翻訳の枠組

$P(j|e)$ は翻訳モデル, $P(e)$ は言語モデル, $\operatorname{argmax}_e P(j|e)P(e)$ はデコーダーにより求められる確率値が最大となる翻訳候補である. $P(e|j)$ を直接求めるには, 2 言語の確率モデルを正確に推定することが困難であることが挙げられる. そこで, 翻訳モデルを用いて翻訳候補を生成した後, 言語モデルを用いて, デコーダーより適切な翻訳を得る可能性を高めている. また, パラメータチューニングを行うことで, より適切な翻訳を得る可能性を高めている.

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 [5] を用いた GIZA++ [6] 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

```

0 ||| , ||| (0) ||| (0) ||| 0.000963391 0.0017823 0.0196078 0.0047559 2.718
0 ||| 10 ||| (0) ||| (0) ||| 0.0540541 0.350932 0.117647 0.0358275 2.718
0 ||| 100 ||| (0) ||| (0) ||| 0.0555556 0.512 0.0588235 0.0405834 2.718
0 ||| 20 ||| (0) ||| (0) ||| 0.0114943 0.306818 0.0196078 0.0256817 2.718
1 0 ||| 10 ||| (0) (0) ||| (0,1) ||| 0.679245 0.118474 0.257143 0.0346384 2.718
1 0 ||| 10th ||| (0) (0) ||| (0,1) ||| 0.5 0.0355998 0.00357143 0.00298755 2.718
1 0 ||| Ten ||| (0) (0) ||| (0,1) ||| 0.375 0.120536 0.0107143 0.00585105 2.718
1 0 ||| ten ||| (0) (0) ||| (0,1) ||| 0.685841 0.18536 0.576208 0.0733759 2.718
...

```

左から、日本語フレーズ、英語フレーズ、日英方向のフレーズ内単語対応、英日方向のフレーズ内単語対応、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、翻訳精度を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。

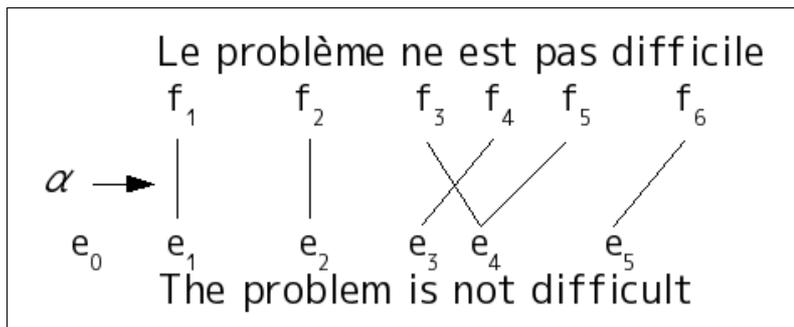


図 2.2: アライメントの例

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 における対応関係を表す。

2.3.3 Model1

(2.3) 式は次のように書き換えることが出来る。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 単語目まで、 $e = e_1^l$ 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.

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

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 $ID(e_id)$ 、翻訳する言語の単語 $ID(f_id)$ 、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 ($P(f_id|e_id)$) 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 $ID(e_id)$ 、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。

—— 言語モデルの例 ——

```
-2.210231 American family -0.1563026
-1.961283 American football -0.06711379
-0.6271492 Americans . 3.252243
-0.7939654 Aomori . 3.678212
-0.8880116 Article 15 -0.1335633
-0.9993331 Asia and -0.1789583
-1.696126 Asian affairs 0.01086971
-0.9044766 Asian countries -0.006818749
-1.696126 Asian markets -0.1379688
...
```

一番上の行は、左から、American の後に family が来る確率を常用対数 \log_{10} で表した値 $\log_{10}(P(\text{family}|\text{American}))$ 、2-gram で表された単語列 American family、バックオフスムージングにより、American の後に family が来る確率を常用対数 \log_{10} で表した値 $\log_{10}(P(\text{family}|\text{American}))$ 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、文法構造の情報を持たない。例えば日英統計翻訳のような異なる文法構造間の翻訳は、日本語から英語へ文法構造が変化するため、同じ文法構造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 $N(w_{i-2}^i)$ 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。

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

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。

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

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 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 2.2 に示す。

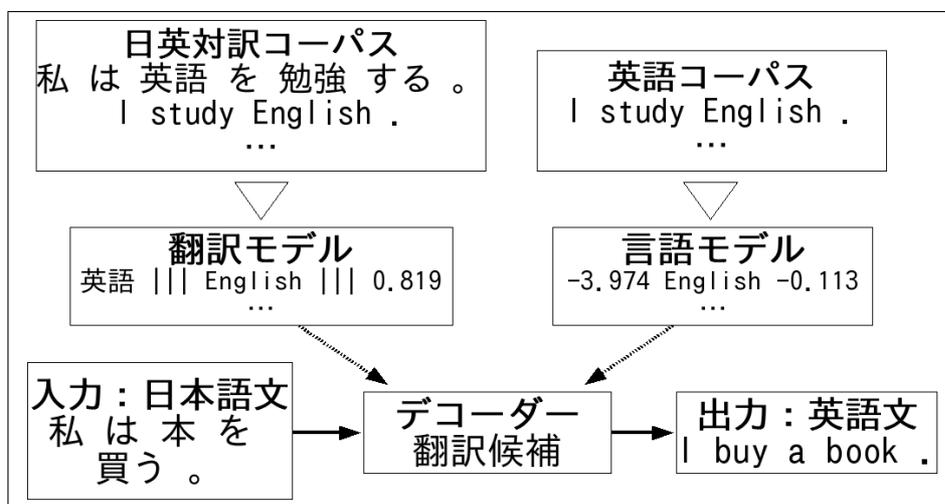


図 2.3: デコーダーにおける翻訳候補の生成例

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。

2.6 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)[7] 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、develop データと呼ばれる試し翻訳を行うデータを用いる。

develop データの各文から、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、develop データに依存するため、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、文法構造が単純な単文 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文法構造が複雑な文 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 日本語文法構造の変換による翻訳

3.1 過去の研究

文法構造変換を行い，日本語文を英語の文法構造に近づけてから，統計翻訳を行う研究がされている．以下に文法構造変換を用いた研究について説明する．

3.1.1 動詞の文法構造変換 [1]

岡崎による動詞の文法構造変換の手法は日本語文を英語の文法構造に近づけるために，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．動詞の文法構造変換の変換例を図 3.1 に示す．

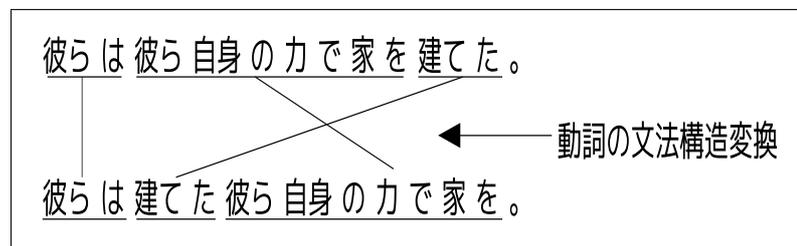


図 3.1: 動詞の文法構造変換の変換例

3.1.2 述語項構造に基づく事前並べ替え [2]

星野らによる述語項構造に基づく事前並べ替えの手法を以下に示す．

1. 係り受け・述語項構文解析

日本語文に対して，係り受け・述語項構文解析を行う．

2. 文の文法構造変換

係り受け関係を用いて，深さ優先に基づき，並び替える．

3. 助詞の文法構造変換

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事前並べ替え [2] の変換例を図 3.2 に示す．

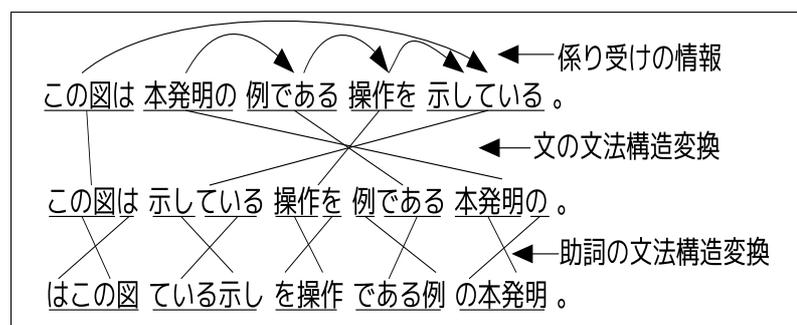


図 3.2: 述語項構造に基づく事前並べ替えの変換例

3.2 提案手法

3.2.1 提案手法の概要

星野らの研究 [2] は特許文を使用している．そして，人手評価を行っていない．しかし，特許文は文法構造が複雑で，人手評価が困難である．

そこで，本研究では簡単な文法構造である単文コーパス [3] を使用する．動詞および目的語および助詞の 3 種類の文法構造変換を組み合わせた実験を行う．それぞれの手法を以下に示す．

1. 動詞の文法構造変換

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．なお，岡崎の動詞の文法構造変換 (図 3.1) と同様の手法である．

2. 目的語の文法構造変換

目的語の句を後ろから順に並び替える．なお，この文法構造変換は星野らの述語項構造に基づく事前並べ替えの手法 (図 3.2) と，ほぼ同等である．

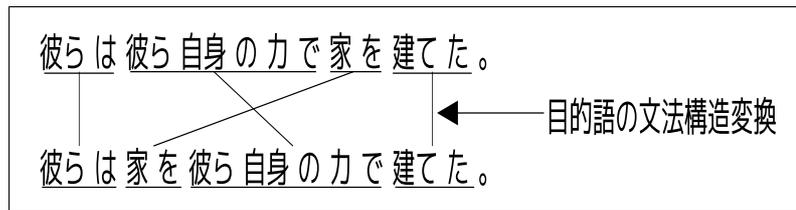


図 3.3: 目的語の文法構造変換の変換例

3. 助詞の文法構造変換

目的語内の文節の内容語と機能語を並び替える。

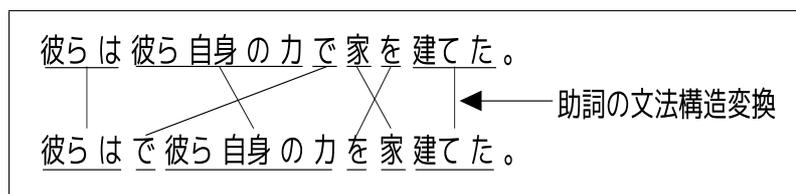


図 3.4: 助詞の文法構造変換の変換例

動詞の文法構造変換と目的語の文法構造変換と助詞の文法構造変換を組み合わせたと
 きの変換例を図 3.5 に示す。

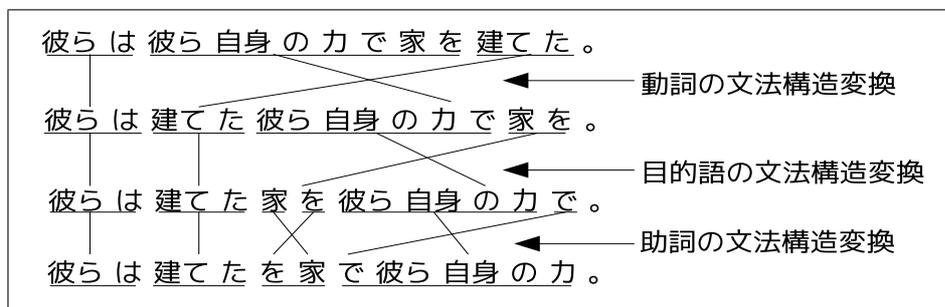


図 3.5: 動詞+目的語+助詞の文法構造変換の変換例

3.2.2 本研究で用いる文法

提案方法では，形態素解析を行うために Cabocha[8] を用いる．本研究で用いる主語・動詞・目的語・助詞の定義を以下に示す．

3.2.3 主語の定義

主語は文頭から係助詞「は」，格助詞「が」，係助詞「も」が最初に出現するまでの単語列とする．主語の例を以下の下線部に示す．

主語の例

私の母は 部屋の掃除を始めた。
少年の快活さが 私たちの悲しみをやわらげた。
彼も 春の目覚めの年頃になった。

3.2.4 動詞の定義

動詞は文末から最初に，動詞の定義に該当した単語までの単語列とする．動詞の定義の一部を表 3.1 に示す．動詞の例を以下の下線部に示す．

動詞の例

私の母は部屋の掃除を始めた。
少年の快活さが私たちの悲しみをやわらげた。
彼も春の目覚めの年頃になった。

表 3.1: 動詞の定義

「名詞-サ変接続」 + 「名詞-サ変接続」 + 「動詞-自立」 例：償還請求します
「名詞-サ変接続」 + 「動詞-自立」または「助動詞」 + 「助動詞」 例：反対しなかった
「動詞-自立」 + 「助動詞」 + 「助動詞」 例：生まなかった, すぎなかった
「名詞-サ変接続」 + 「名詞-接尾-形容動詞語幹」または「助動詞」 + 「動詞-非自立」 例：保守的すぎる
「名詞-サ変接続」 + 「動詞-自立」または「動詞-非自立」 例：支持した
「名詞-形容動詞語幹」 + 「動詞-自立」, 「動詞-非自立」または「助動詞」 例：だ文, 馬鹿です, 馬鹿する
「名詞-固有名詞-人名-姓」 + 「動詞-自立」または「助動詞」 例：山田です, 山田する
「形容詞-自立」 + 「動詞-非自立」または「助動詞」 例：速かった
「名詞-一般」 + 「助動詞」 「副詞-助詞類接続」 + 「動詞-自立」
「動詞-自立」 + 「動詞-自立」または「動詞-非自立」 例：吹き倒された, 言い負かす (複合動詞)
「動詞-自立」 例：言う, 歩く
「形容詞-自立」 例：なかった, 多い, 大きい

3.2.5 目的語の定義

目的語は主語と動詞に挟まれた単語列とする。目的語の例を以下の下線部に示す。

——— 目的語の例 ———

私の母は 部屋の掃除 を始めた。
 少年の快活さが 私たちの悲しみ をやわらげた。
 彼も 春の目覚めの年頃 になった。

3.2.6 助詞の定義

助詞は目的語内の文節の中で「の」、「と」以外の助詞とする。助詞の例を以下の下線部に示す。

— 助詞の例 —

私の母は部屋の掃除を始めた。

少年の快活さが私たちの悲しみをやわらげた。

彼も春の目覚めの年頃になった。

第4章 実験環境

4.1 対訳データ

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 [3] から，主語のない文と体言止めの文を削除する．対訳データの文数を表 4.1 に示す．

単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[8] 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 を用いた英字全小文字化を行わない．

表 4.1: 対訳データの文数

train 文	82,261
test 文	8,189
development 文	825

4.1.1 通常 of データ

通常 of データは，4.1 章 of データを使用する．以下に日英対訳コーパス of 例を示す．

日本語コーパス

その講座は 1 月に終わる。
彼はおじさんの口利きで商社に就職した。
新聞は社説で物価高を論じている。

英語コーパス

The course finishes in January .
He obtained a position in a trading company by the good offices of his uncle .
The paper comments editorially on the rise of prices .

4.1.2 提案手法のデータ

提案手法のデータは，日本語コーパスが3.3章にて文法構造変換されたデータを用いる．以下に動詞の文法構造変換と目的語の文法構造変換と助詞の文法構造変換を組み合わせたときの変換例を示す．

日本語コーパス

その講座は終わるに1月。
彼は就職したに商社でおじさんの口利き。
新聞は論じているを物価高で社説。

英語コーパス

The course finishes in January .
He obtained a position in a trading company by the good offices of his uncle .
The paper comments editorially on the rise of prices .

4.2 翻訳モデルの学習

4.2.1 翻訳モデル学習の概要

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[5]に基づくGIZA++[6]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，max-phrase-lengthは20と設定する．なお，“train-factored-phrase-model.perl”はリオーダーリングモデルを作成することが可能であるが今回は使用しないため，学習法はdistanceとする．

4.3 言語モデルの学習

言語モデルの学習には，“SRILM (Stanford Research Institute Language Model)”[9]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．

4.4 デコーダーに関するパラメータ

デコーダーは、moses[10]を使用する。mosesはパラメータを設定することが出来る。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、過去の研究から、翻訳モデルは日英翻訳確率と英日翻訳確率の共起確率を使用したほうがよい結果が得られる傾向があるため、“0.5 0.0 0.5 0.0 0.0”とする[11]。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“0.1”，フレーズの並び替え制限“distortion-limit”の値は“-1”，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、リオーダーリングモデル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。しかし、本研究では、リオーダーリングモデルを用いずに実験を行う。

4.5 評価方法

評価方法に人手評価と自動評価を用いる。人手評価は翻訳文からランダムに100文取り出し、文法構造変換を行わない実験(ベースライン)との対比較実験を行う。自動評価にはBLEU[12]，NIST[12]，METEOR[13]，RIBES[14]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 score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。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。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。

第5章 翻訳実験

5.1 実験内容

本研究では動詞および目的語および助詞の3種類の文法構造変換を用いる。そして、3種類の文法構造変換を組み合わせた計8種類の実験を行う。実験の種類を以下に示す。

1. 文法構造変換を行わない実験 (ベースライン)

—— ベースラインの例 ——

その講座は1月に終わる。
彼はおじさんの口利きで商社に就職した。
新聞は社説で物価高を論じている。

2. 動詞の文法構造変換単独

—— 変換例 ——

その講座は終わる1月に。
彼は就職したおじさんの口利きで商社に。
新聞は論じている社説で物価高を。

3. 目的語の文法構造変換単独

—— 変換例 ——

その講座は1月に終わる。
彼は商社におじさんの口利きで就職した。
新聞は物価高を社説で論じている。

4. 助詞の文法構造変換単独

変換例

その講座はに1月終わる。
彼はでおじさんの口利きに商社就職した。
新聞はで社説を物価高論じている。

5. 動詞の文法構造変換+目的語の文法構造変換

変換例

その講座は終わる1月に。
彼は就職した商社におじさんの口利きで。
新聞は論じている物価高を社説で。

6. 動詞の文法構造変換+助詞の文法構造変換

変換例

その講座は終わるに1月。
彼は就職したでおじさんの口利きに商社。
新聞は論じているで社説を物価高。

7. 目的語の文法構造変換+助詞の文法構造変換

変換例

その講座はに1月終わる。
彼はに商社でおじさんの口利き就職した。
新聞はを物価高で社説論じている。

8. 動詞の文法構造変換+目的語の文法構造変換+助詞の文法構造変換

変換例

その講座は終わるに1月。
彼は就職したに商社でおじさんの口利き。
新聞は論じているを物価高で社説。

また、実験は句に基づく統計翻訳(PSMT)、階層型統計翻訳(HSMT)の2種類で行う。したがって、計16種類の実験を行う。

5.2 実験条件

辞書の例文から抽出した対訳データの単文コーパスから、主語のない文と体言止めの文を削除したデータを使用して、実験を行う。人手評価は翻訳文からランダムに100文取り出し、文法構造変換を行わない実験(ベースライン)との対比較実験を行う。自動評価にはBLEUとNISTとMETEORとRIBESを用いる。

5.3 実験結果

5.3.1 人手評価

人手評価のPSMTでの結果を表5.1に、HSMTでの結果を表5.2に示す。

表 5.1: PSMTでの人手評価

文法構造変換	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	13	70	11
目的語+助詞	6	10	73	10
動詞+目的語+助詞	6	14	69	11

表 5.2: HSMTでの人手評価

文法構造変換	ベース ライン○	提案 手法○	差なし	同一出力
動詞	6	8	74	12
目的語	4	6	68	22
助詞	5	6	73	16
動詞+目的語	7	5	78	10
動詞+助詞	3	12	71	14
目的語+助詞	6	6	71	17
動詞+目的語+助詞	7	14	69	10

5.3.2 自動評価

自動評価の PSMT での結果を表 5.3 に，HSMT での結果を表 5.4 に示す。

表 5.3: PSMT の翻訳精度

文法構造変換	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 5.4: HSMT の翻訳精度

文法構造変換	BLEU	NIST	METEOR	RIBES
動詞	0.1348	4.7889	0.4593	0.7266
目的語	0.1339	4.7849	0.4574	0.7237
助詞	0.1340	4.7707	0.4572	0.7224
動詞+目的語	0.1323	4.8200	0.4585	0.7269
動詞+助詞	0.1342	4.7824	0.4617	0.7267
目的語+助詞	0.1320	4.7797	0.4480	0.7229
動詞+目的語+助詞	0.1333	4.8274	0.4638	0.7294
ベースライン	0.1331	4.7383	0.4557	0.7193

5.3.3 実験結果のまとめ

表 5.1, 5.2 から以下のことが示せる。

1. PSMT と HSMT において，動詞の文法構造変換と助詞の文法構造変換を組み合わせた実験は翻訳精度が高い。
2. 一部の例外を除き，文法構造変換を行うことで，翻訳精度が向上する。
3. PSMT において，助詞の文法構造変換単独は有効である。しかし，動詞の文法構造変換単独および目的語の文法構造変換単独は有効でない。

4. HSMTにおいて、動詞単独、目的語単独、助詞単独の文法構造変換は有効でない。
5. 表5.1と表5.3, 表5.2と表5.4を比べると、自動評価と人手評価の実験結果は一致しない。

5.3.4 PSMTでの翻訳例

動詞および目的語および助詞の文法構造変換を組み合わせた実験の翻訳例を以下に示す。なお、表中の下線は誤りと判断した箇所を表す。

1. 動詞の文法構造変換単独

表 5.5: 動詞単独が優れている翻訳例

例 1	
入力文	The police caught the author of bank robbers .
動詞	The boy came to life .
ベースライン	The boy <u>breath</u> .
正解文	The boy came back to life .
例 2	
入力文	ケネディ 大統領 は 彼 に 国防 長官 の いす を 申し出 た 。
動詞	President Kennedy's offered him to the post of the secretary of defense .
ベースライン	President Kennedy's <u>him to the defense of chief of the chair</u> .
正解文	Kennedy offered him the defense post .
例 3	
入力文	それは 確かに 1 0 ドル は した 。
動詞	It is a certainly ten dollars .
ベースライン	<u>It is a ten dollars</u> .
正解文	It cost me ten dollars , if a cent .

表 5.6: ベースラインが優れている翻訳例

例 1	
入力文	多くの 会社 が 休業 した 。
動詞	<u>The closed for a lot of our company</u> .
ベースライン	Many companies closed for .
正解文	Many businesses closed down .
例 2	
入力文	被害 者 たち は 警察 に 保護 を 求め た 。
動詞	<u>We asked the police protection victims</u> .
ベースライン	The victims asked the police protection .
正解文	The sufferers asked for police protection .
例 3	
入力文	警察 は 銀行 強盗 の 張本人 を 捕まえた 。
動詞	The police <u>have not</u> caught the author of the bank robber .
ベースライン	The police caught the author of bank robbers .
正解文	The police arrested the ringleader of the bank robbery .

2. 目的語の文法構造変換単独

表 5.7: 目的語単独が優れている翻訳例

例 1	
入力文	ふつうは子供が学校へ行く時間だ。
目的語	usually time when children go to school .
ベースライン	<u>We,in is the time when children go to school .</u>
正解文	Normally , it is the time when children go to school .
例 2	
入力文	今朝はいつもより早く起きた。
目的語	I got up earlier than usual this morning .
ベースライン	<u>This morning got up earlier than usual .</u>
正解文	I got up earlier than usual this morning .
例 3	
入力文	本が倒れた。
目的語	The book fell .
ベースライン	<u>There are books .</u>
正解文	The book fell down .

表 5.8: ベースラインが優れている翻訳例

例 1	
入力文	靴は 5 足持っている。
目的語	<u>has five pairs of shoes .</u>
ベースライン	I have five pairs of shoes .
正解文	I have five pairs of shoes .
例 2	
入力文	ひとつの問題が別の問題を生じた。
目的語	<u>One of the problem of the problem .</u>
ベースライン	A problem has caused another problem .
正解文	One problem led to another .
例 3	
入力文	彼は両手を頭の上に上げた。
目的語	He raised <u>their hands</u> on the <u>top</u> of his head .
ベースライン	He put his hands on the head .
正解文	He raised his hands above his head .

3. 助詞の文法構造変換単独

表 5.9: 助詞単独が優れている翻訳例

例 1	
入力文	私は新しい事業を計画している。
助詞	I have a new business project .
ベースライン	I have a <u>gunpowder factory</u> .
正解文	I am planning another enterprise .
例 2	
入力文	被告に有利な判決が下された。
助詞	The decided in favor of the defendant was made .
ベースライン	<u>decided in favor of the defendant</u> .
正解文	A decision was given for the defendant .
例 3	
入力文	最近、学生の気持ちが弛んでいる。
助詞	Recently , the feelings of the students is slackening .
ベースライン	In recent years , however , has flabby feelings of the students .
正解文	Students <u>have recently become too relaxed</u> .

表 5.10: ベースラインが優れている翻訳例

例 1	
入力文	東京の都心部でも同様の効果が出ている。
助詞	Tokyo is the effects of the in the center of the city .
ベースライン	There is a similar effect in the center of Tokyo .
正解文	The situation is reported to be similar in Tokyo .
例 2	
入力文	総会は 1 年に 1 回開かれる。
助詞	The <u>general are held</u> once a year .
ベースライン	The general meeting is held once a year .
正解文	The general assembly is held once a year .
例 3	
入力文	彼女にいつもの笑顔が戻った。
助詞	The smile came back to her <u>in the</u> .
ベースライン	The smile came back to her usual .
正解文	She got back her usual smile .

4. 動詞＋目的語の文法構造変換

表 5.11: 動詞＋目的語が優れている翻訳例

例 1	
入力文	彼らのチームは最初から試合の主導権を握った。
動詞＋目的語	They took the initiative from the start of the game .
ベースライン	<u>Their team from the start the initiative of the game .</u>
正解文	Their team controlled the match from the start .
例 2	
入力文	今朝はいつもより早く起きた。
動詞＋目的語	I woke up this morning earlier than usual .
ベースライン	<u>This morning got up earlier than usual .</u>
正解文	I got up earlier than usual this morning .
例 3	
入力文	試験方法の変更は遅れている。
動詞＋目的語	The change of the test method is behind .
ベースライン	test method is <u>a change of late .</u>
正解文	Changes in the way exams are run are overdue .

表 5.12: ベースラインが優れている翻訳例

例 1	
入力文	被害者たちは警察に保護を求めた。
動詞＋目的語	<u>We asked the police for protection victims .</u>
ベースライン	The victims asked the police protection .
正解文	The sufferers asked for police protection .
例 2	
入力文	ひとつの問題が別の問題を生じた。
動詞＋目的語	<u>One of the problem of the problem .</u>
ベースライン	A problem has caused another problem .
正解文	One problem led to another .
例 3	
入力文	総会は 1 年に 1 回開かれる。
動詞＋目的語	<u>The meeting will be held once a year .</u>
ベースライン	The general meeting is held once a year .
正解文	The general assembly is held once a year .

5. 動詞＋助詞の文法構造変換

表 5.13: 動詞＋助詞が優れている翻訳例

例 1	
入力文	私は新しい事業を計画している。
動詞＋助詞	I plan to the new business .
ベースライン	<u>I have a gunpowder factory .</u>
正解文	I am planning another enterprise .
例 2	
入力文	彼の意見に文句はありません。
動詞＋助詞	I don't have any objection to his opinion .
ベースライン	I don't have <u>the opinions of the with him .</u>
正解文	I have no quarrel with his opinion .
例 3	
入力文	君の行為は全く我慢できない。
動詞＋助詞	I can't stand his actions of you .
ベースライン	<u>Your act is completely .</u>
正解文	Your conduct is beyond bearing .

表 5.14: ベースラインが優れている翻訳例

例 1	
入力文	ひとつの問題が別の問題を生じた。
動詞＋助詞	<u>One of the problem of the problem .</u>
ベースライン	A problem has caused another problem .
正解文	One problem led to another .
例 2	
入力文	東京の都心部でも同様の効果が出ている。
動詞＋助詞	<u>The in the center of Tokyo are similar effect .</u>
ベースライン	There is a similar effect in the center of Tokyo .
正解文	The situation is reported to be similar in Tokyo .
例 3	
入力文	被害者たちは警察に保護を求めた。
動詞＋助詞	<u>We asked the police protection of the victims .</u>
ベースライン	The victims asked the police protection .
正解文	The sufferers asked for police protection .

6. 目的語+助詞の文法構造変換

表 5.15: 目的語+助詞が優れている翻訳例

例 1	
入力文	その 訓練 は ただちに 効果 が あった。
目的語+助詞	The training was a effect immediately .
ベースライン	The training <u>at once</u> .
正解文	The training was of immediate benefit .
例 2	
入力文	彼は 人の 目を 気に し 過ぎる。
目的語+助詞	He is too sensitive to the eyes of the people .
ベースライン	He is too sensitive to the eye .
正解文	He is too sensitive about how he appears to people .
例 3	
入力文	最近、 学生 の 気持ち が 弛ん で いる。
目的語+助詞	Recently , the feelings of the students is slackening .
ベースライン	In recent years , <u>however</u> , has flabby feelings of the students .
正解文	Students have recently become too relaxed .

表 5.16: ベースラインが優れている翻訳例

例 1	
入力文	多くの 会社 が 休業 した。
目的語+助詞	A lot of the company shut down .
ベースライン	Many companies closed for .
正解文	Many businesses closed down .
例 2	
入力文	総会 は 1 年 に 1 回 開か れる。
目的語+助詞	The general held once a year .
ベースライン	The general meeting is held once a year .
正解文	The general assembly is held once a year .
例 3	
入力文	彼女 に いつも の 笑顔 が 戻っ た。
目的語+助詞	She <u>always</u> has returned to the smile .
ベースライン	The smile came back to her usual .
正解文	She got back her usual smile .

7. 動詞＋目的語＋助詞の文法構造変換

表 5.17: 動詞＋目的語＋助詞が優れている翻訳例

例 1	
入力文	彼らのチームは最初から試合の主導権を握った。
動詞＋目的語＋助詞	Their team took the initiative in the match from the start .
ベースライン	<u>Their team from the start the initiative of the game .</u>
正解文	Their team controlled the match from the start .
例 2	
入力文	試験方法の変更は遅れている。
動詞＋目的語＋助詞	The change of the test method is behind .
ベースライン	<u>test method is a change of late .</u>
正解文	Changes in the way exams are run are overdue .
例 3	
入力文	被告に有利な判決が下された。
動詞＋目的語＋助詞	There was decided in favor of the defendant .
ベースライン	<u>decided in favor of the defendant .</u>
正解文	A decision was given for the defendant .

表 5.18: ベースラインが優れている翻訳例

例 1	
入力文	被害者たちは警察に保護を求めた。
動詞＋目的語＋助詞	We asked the police protection of the victims .
ベースライン	<u>The victims asked the police protection .</u>
正解文	The sufferers asked for police protection .
例 2	
入力文	ひとつの問題が別の問題を生じた。
動詞＋目的語＋助詞	One of the problem of another problem .
ベースライン	<u>A problem has caused another problem .</u>
正解文	One problem led to another .
例 3	
入力文	この鉱山は見込みがある。
動詞＋目的語＋助詞	<u>This mine .</u>
ベースライン	The mine prospects well .
正解文	The mine prospects well .

5.3.5 HSMT での翻訳例

動詞および目的語および助詞の文法構造変換を組み合わせた実験の翻訳例を以下に示す。なお、表中の下線は誤りと判断した箇所を表す。

1. 動詞の文法構造変換単独

表 5.19: 動詞単独が優れている翻訳例

例 1	
入力文	少年は息を吹き返した。
動詞	The boy came to life .
ベースライン	The boy <u>was a breath</u> .
正解文	The boy came back to life .
例 2	
入力文	彼の意見に文句はありません。
動詞	I have no words of his opinion .
ベースライン	<u>He is with the opinions of the</u> .
正解文	I have no quarrel with his opinion .
例 3	
入力文	その数は 20 パーセントだけ減らされた。
動詞	The number was reduced by 20% .
ベースライン	The number was <u>only a 20 percent</u> .
正解文	The number was lessened by 20% .

表 5.20: ベースラインが優れている翻訳例

例 1	
入力文	靴は 5 足持っている。
動詞	<u>My shoes has five feet</u> .
ベースライン	I have five pairs of shoes .
正解文	I have five pairs of shoes .
例 2	
入力文	彼はなかなか仕事ができる。
動詞	<u>He is a man of work</u> .
ベースライン	He can work .
正解文	He is no slouch at the job .
例 3	
入力文	警察は銀行強盗の張本人を捕まえた。
動詞	The police <u>have not</u> caught the author of the bank robbery .
ベースライン	The police caught the author of the bank robbery .
正解文	The police arrested the ringleader of the bank robbery .

2. 目的語の文法構造変換単独

表 5.21: 目的語単独が優れている翻訳例

例 1	
入力文	君の行為は全く我慢できない
目的語	I can't bear an act of you .
ベースライン	I can't <u>act of you</u> .
正解文	Your conduct is beyond bearing .
例 2	
入力文	それは確かに 10 ドルはした。
目的語	It is certainly ten dollars .
ベースライン	<u>It is ten dollars</u> .
正解文	It cost me ten dollars , if a cent .
例 3	
入力文	彼は人の目を気にし過ぎる。
目的語	He is too sensitive to the eyes of people .
ベースライン	He is too <u>concerned</u> in the eyes of the people .
正解文	He is too sensitive about how he appears to people .

表 5.22: ベースラインが優れている翻訳例

例 1	
入力文	靴は 5 足持っている。
目的語	<u>My shoes have five feet</u> .
ベースライン	I have five pairs of shoes .
正解文	I have five pairs of shoes .
例 2	
入力文	東京の都心部でも同様の効果が出ている。
目的語	There is a similar effect in the center <u>of the department</u> of Tokyo .
ベースライン	There is a similar effect in the center of Tokyo .
正解文	The situation is reported to be similar in Tokyo .
例 3	
入力文	彼は車を私に預けた。
目的語	He <u>confided</u> her car to me .
ベースライン	He left a car on me .
正解文	He entrusted me with his car .

3. 助詞の文法構造変換単独

表 5.23: 助詞単独が優れている翻訳例

例 1	
入力文	その訓練はただちに効果があった。
助詞	The training was a effect immediately .
ベースライン	<u>The training at once .</u>
正解文	The training was of immediate benefit .
例 2	
入力文	多くの会社が休業した。
助詞	Many company has closed .
ベースライン	<u>Many closed for the company .</u>
正解文	Many businesses closed down .
例 3	
入力文	被害者たちは警察に保護を求めた。
助詞	The victims asked the police protection .
ベースライン	<u>The victims of the police protection .</u>
正解文	The sufferers asked for police protection .

表 5.24: ベースラインが優れている翻訳例

例 1	
入力文	私は新しい事業を計画している。
助詞	I <u>ordered</u> a new business project .
ベースライン	I plan to the new business .
正解文	I am planning another enterprise .
例 2	
入力文	彼はなかなか仕事ができる。
助詞	<u>He is good at her job .</u>
ベースライン	He can work .
正解文	He is no slouch at the job .
例 3	
入力文	世界の生産高は年に約 10 万トンと推定される。
助詞	<u>Production is estimated at a year about ten tons of the world .</u>
ベースライン	Production of the world in a year is estimated at one hundred thousand tons .
正解文	World production is put at some 100,000 tons a year .

4. 動詞＋目的語の文法構造変換

表 5.25: 動詞＋目的語が優れている翻訳例

例 1	
入力文	彼は人の目を気にし過ぎる。
動詞＋目的語	He is too sensitive to the eyes of the people .
ベースライン	He is too <u>concerned</u> in the eyes of the people .
正解文	He is too sensitive about how he appears to people .
例 2	
入力文	少年は息を吹き返した。
動詞＋目的語	The boy came to life .
ベースライン	The boy <u>was a breath</u> .
正解文	The boy came back to life .
例 3	
入力文	彼の意見に文句はありません。
動詞＋目的語	There is no words of his opinion .
ベースライン	<u>He is with the opinions of the</u> .
正解文	I have no quarrel with his opinion .

表 5.26: ベースラインが優れている翻訳例

例 1	
入力文	雪の重みで木の枝が曲がっていた。
動詞＋目的語	The branches of the tree was bent under <u>way</u> under the weight of the snow .
ベースライン	The branches of the tree are bent under the weight of the snow .
正解文	The tree branches have bent under the snow .
例 2	
入力文	警察は銀行強盗の張本人を捕まえた。
動詞＋目的語	The police <u>have not</u> caught the author of a robbery at the bank .
ベースライン	The police caught the author of the bank robbery .
正解文	The police arrested the ringleader of the bank robbery .
例 3	
入力文	数学ではとても彼にかなわない。
動詞＋目的語	<u>I am not nearly so strong in mathematics as he</u> .
ベースライン	I am no match for him in mathematics .
正解文	I am nowhere near as good as he in math .

5. 動詞＋助詞の文法構造変換

表 5.27: 動詞＋助詞が優れている翻訳例

例 1	
入力文	教育にはたいへんな金がかかる。
動詞＋助詞	Education is very expensive .
ベースライン	<u>The money will take a very education .</u>
正解文	Education is very expensive .
例 2	
入力文	君の行為は全く我慢できない。
動詞＋助詞	I can't stand this act of you .
ベースライン	<u>I can't act of you .</u>
正解文	Your conduct is beyond bearing .
例 3	
入力文	この学校にはあの学校の 2 倍の生徒がいます。
動詞＋助詞	This school has twice the students of that school .
ベースライン	<u>That will be twice as much as of the students of the school in this school .</u>
正解文	There are twice as many students in this school as in that one .

表 5.28: ベースラインが優れている翻訳例

例 1	
入力文	靴は 5 足持っている。
動詞＋助詞	<u>The shoes have five feet .</u>
ベースライン	I have five pairs of shoes .
正解文	I have five pairs of shoes .
例 2	
入力文	雪の重みで木の枝が曲がっていた。
動詞＋助詞	The branches of the tree was bent <u>under way</u> under the weight of the snow .
ベースライン	The branches of the tree are bent under the weight of the snow .
正解文	The tree branches have bent under the snow .
例 3	
入力文	世界の生産高は年に約 1 0 万トンと推定される。
動詞＋助詞	Production is estimated at the age of about ten thousand tons <u>of the world</u> .
ベースライン	Production of the world in a year is estimated at one hundred thousand tons .
正解文	World production is put at some 100,000 tons a year .

6. 目的語+助詞の文法構造変換

表 5.29: 目的語+助詞が優れている翻訳例

例 1	
入力文	その 訓練 は ただちに 効果 が あった。
目的語+助詞	The training was effect immediately .
ベースライン	<u>The training at once .</u>
正解文	The training was of immediate benefit .
例 2	
入力文	それは 確かに 1 0 ドル は した。
目的語+助詞	It is certainly ten dollars .
ベースライン	<u>It is ten dollars .</u>
正解文	It cost me ten dollars , if a cent .
例 3	
入力文	彼の 意見 に 文句 は ありません。
目的語+助詞	I have no words to his opinion .
ベースライン	<u>He is with the opinions of the .</u>
正解文	I have no quarrel with his opinion .

表 5.30: ベースラインが優れている翻訳例

例 1	
入力文	私は 新しい 事業 を 計画 している。
目的語+助詞	I <u>ordered</u> a new business plans .
ベースライン	I plan to the new business .
正解文	I am planning another enterprise .
例 2	
入力文	警察 は 銀行 強盗 の 張本人 を 捕まえた。
目的語+助詞	The author of the bank robbery and <u>caught him</u> .
ベースライン	The police caught the author of the bank robbery .
正解文	The police arrested the ringleader of the bank robbery .
例 3	
入力文	彼は 車 を 私 に 預けた。
目的語+助詞	He <u>confided her car in front of me</u> .
ベースライン	He left a car on me .
正解文	He entrusted me with his car .

7. 動詞＋目的語＋助詞の文法構造変換

表 5.31: 動詞＋目的語＋助詞が優れている翻訳例

例 1	
入力文	彼らのチームは最初から試合の主導権を握った。
動詞＋目的語＋助詞	Their team took the initiative of the match from the start .
ベースライン	<u>Their team from the start of the initiative of the game .</u>
正解文	Their team controlled the match from the start .
例 2	
入力文	この学校にはあの学校の2倍の生徒がいます。
動詞＋目的語＋助詞	This school has twice the students of that school .
ベースライン	<u>That will be twice as much as of the students of the school in this school .</u>
正解文	There are twice as many students in this school as in that one .
例 3	
入力文	教育にはたいへんな金がかかる。
動詞＋目的語＋助詞	Education is very expensive .
ベースライン	<u>The money will take a very education .</u>
正解文	Education is very expensive .

表 5.32: ベースラインが優れている翻訳例

例 1	
入力文	最近、学生の気持ちが弛んでいる。
動詞＋目的語＋助詞	Recently , the feelings of the students <u>slackening</u> .
ベースライン	In recent years , feelings of the students is slackening .
正解文	Students have recently become too relaxed .
例 2	
入力文	この鉱山は見込みがある。
動詞＋目的語＋助詞	<u>This mine .</u>
ベースライン	The mine prospects well .
正解文	The mine prospects well .
例 3	
入力文	東京の都心部でも同様の効果が出ている。
動詞＋目的語＋助詞	There is a similar effect in the center of <u>the department of Tokyo</u> .
ベースライン	There is a similar effect in the center of Tokyo .
正解文	The situation is reported to be similar in Tokyo .

第6章 考察

本研究では、動詞および目的語および助詞の文法構造変換を行い、翻訳精度の変化を調べた。実験の考察を以下に示す。

6.1 星野らの研究 [2] との比較

星野らの研究との比較を以下に示す。

1. 星野らは係り受け、述語項構文解析 (目的語の文法構造変換) を用いて、文法構造変換を行った。しかし、本研究の結果では動詞の文法構造変換と助詞の文法構造変換の組み合わせは有効だが、目的語の文法構造変換は有効でなかった。
2. PSMT において、助詞の文法構造変換単独で翻訳精度が向上した。しかし、HSMT では翻訳精度が向上しない。

6.2 PSMT と HSMT の対比較

6.2.1 対比較の結果

PSMT と HSMT の対比較の結果を表 6.1 に示す。なお、実験はベースライン、動詞 + 助詞の文法構造変換、動詞 + 目的語 + 助詞の文法構造変換で行った。

表 6.1: PSMT と HSMT の対比較

	PSMT ○	HSMT ○	差なし	同一出力
動詞 + 助詞	1	8	67	24
ベースライン	5	8	68	19
動詞 + 目的語 + 助詞	7	7	60	26

6.2.2 対比較のまとめ

1. 動詞と助詞の文法構造変換を組み合わせた実験では、HSMTの翻訳精度の方がPSMTより優れている。
2. HSMTにおいて動詞と助詞の文法構造変換の組み合わせが、16種類の実験で最も良いと考える。
3. 動詞と目的語と助詞の文法構造変換を組み合わせた実験では、PSMTとHSMTの翻訳精度に差がない。

6.2.3 文法構造変換の精度確認

動詞および目的語および助詞の文法構造変換を組み合わせた7種類の実験、および6.4節でのOSVへの文法構造変換の2種類の実験の計9種類の実験の翻訳文から、それぞれランダムに100文抽出し、正しく文法構造変換が行えているかを調査した。以下に文法構造変換の精度の結果を示す。

表 6.2: 文法構造変換の精度

	出力×	出力○
9種類の実験	1	99

9種類の文法構造変換の実験において、正しく出力されなかった文を以下に示す。

表 6.3: 正しく出力されなかった文

形態素解析前	この色はながもちする。
形態素解析後	この色 はな がもち する。
正解文	この色 は な がもち する。

正しく出力されなかった理由は形態素解析を行う cabocha に原因があることが判明した。形態素解析のミスを除けば、全ての文で正しく文法構造変換が行えている。

6.2.4 翻訳例

PSMTとHSMTの対比較の翻訳例を以下に示す。

1. 動詞+助詞の文法構造変換

表 6.4: PSMT が優れている翻訳例

例 1	
入力文	雪の重みで木の枝が曲がっていた。
PSMT	The branches of the tree was bent under the weight of the snow .
HSMT	The branches of the tree was bent under <u>way</u> under the weight of the snow .
正解文	The tree branches have bent under the snow .

表 6.5: HSMT が優れている翻訳例

例 1	
入力文	東京の都心部でも同様の効果が出ている。
PSMT	The in the center of Tokyo are similar effect .
HSMT	There is a similar effect in the center of Tokyo .
正解文	The situation is reported to be similar in Tokyo .
例 2	
入力文	ひとつの問題が別の問題を生じた。
PSMT	One of the problem of the problem .
HSMT	A problem has arisen a different problem .
正解文	One problem led to another .
例 3	
入力文	警察は銀行強盗の張本人を捕まえた。
PSMT	The police caught <u>the bank robber author of</u> .
HSMT	The police caught the author of bank robbery .
正解文	The police arrested the ringleader of the bank robbery .

2. 動詞＋目的語＋助詞の文法構造変換

表 6.6: PSMT が優れている翻訳例

例 1	
入力文	靴は 5 足持っている。
PSMT	I have five pairs of shoes .
HSMT	<u>The owns five pairs of shoes .</u>
正解文	I have five pairs of shoes .
例 2	
入力文	総会は 1 年に 1 回開かれる。
PSMT	General Assembly will be held once a year .
HSMT	General meetings <u>held</u> once a year .
正解文	The general assembly is held once a year .
例 3	
入力文	最近、学生の気持ちが弛んでいる。
PSMT	Recently , the feelings of the students is slackening .
HSMT	Recently , the feelings of the students <u>slackening</u> .
正解文	Students have recently become too relaxed .

表 6.7: HSMT が優れている翻訳例

例 1	
入力文	この学校にはあの学校の 2 倍の生徒がいます。
PSMT	This school has twice as much as <u>that of the students of the school</u> .
HSMT	This school has twice the students of that school .
正解文	There are twice as many students in this school as in that one .
例 2	
入力文	彼は人の目を気にし過ぎる。
PSMT	He is too <u>much of his eyes</u> .
HSMT	He is too sensitive to the eyes of the people .
正解文	He is too sensitive about how he appears to people .
例 3	
入力文	彼の意見に文句はありません。
PSMT	<u>His opinion is not complaint</u> .
HSMT	There is no words of his opinion .
正解文	I have no quarrel with his opinion .

3. ベースライン

表 6.8: PSMT が優れている翻訳例

例 1	
入力文	多くの会社が休業した。
PSMT	Many companies closed for .
HSMT	<u>Many closed for the company .</u>
正解文	Many businesses closed down .
例 2	
入力文	ひとつの問題が別の問題を生じた。
PSMT	A problem has caused another problem .
HSMT	<u>One of the problem of another problem .</u>
正解文	One problem led to another .
例 3	
入力文	教育にはたいへんな金がかかる。
PSMT	Education is very expensive .
HSMT	<u>The money will take a very education .</u>
正解文	Education is very expensive .

表 6.9: HSMT が優れている翻訳例

例 1	
入力文	彼は車を私に預けた。
PSMT	He <u>drove the car in my trust .</u>
HSMT	He left a car on me .
正解文	He entrusted me with his car .
例 2	
入力文	最近、学生の気持ちが弛んでいる。
PSMT	In recent years , <u>however , has flabby feelings of the students .</u>
HSMT	In recent years , feelings of the students is slackening .
正解文	Students have recently become too relaxed .
例 3	
入力文	私は新しい事業を計画している。
PSMT	I have <u>a gunpowder factory .</u>
HSMT	I plan to the new business .
正解文	I am planning another enterprise .

6.3 実験の問題点

6.3.1 複合動詞

本研究では、品詞を特定するために cabocha[8] を用いた。しかし、cabocha は「気になる」といった複合動詞は「気」＋「に」＋「なる」の3つの形態素に分解される。このため、複合動詞の一部のみが並び替えされる。例を表 6.8 に示す。

表 6.10: 複合動詞の一部のみが並び替えされる例

入力文	彼は周りの目を気にしない。
動詞＋助詞	彼はしないを周りの目に気。

今後、「気になる」などの複合動詞を1つの動詞として処理する必要がある。

6.3.2 連続した助詞

本研究では、連続した助詞があった際、正しく出力されないことがある。例を表 6.9 に示す。

表 6.11: 連続した助詞の出力例

入力文	信仰は山をも動かす。
動詞＋助詞	信仰は動かすもを山。

今後、連続した助詞を1つの助詞として処理する必要がある。

6.4 OSV(目的語, 主語, 動詞)への文法構造変換

6.4.1 概要

日本語の文法構造は自由であるため、OSV(目的語, 主語, 動詞)の文法構造は不自然ではない。そこで、OSV(目的語, 主語, 動詞)への文法構造変換を行ってから、日英統計翻訳を行う手法を試みた。OSV(目的語, 主語, 動詞)への文法構造変換の手法を以下に示す。

1. 主語の文法構造変換

日本語文の SOV(主語, 目的語, 動詞) を OSV(目的語, 主語, 動詞) の順に並び替える。

2. 目的語の文法構造変換

目的語の句の後ろから順に並び替える。

3. 助詞の文法構造変換

目的語の内容語と機能語を並び替える。

6.4.2 実験内容

データは表 4.1 に用いたものを使用し、ベースラインの例と 2 種類で実験を以下に示す。

————— ベースラインの例 —————

その 講座 は 1 月 に 終わる。
彼 は おじさん の 口利き で 商社 に 就職 した。
新聞 は 社説 で 物価高 を 論じて いる。

1. 主語の文法構造変換単独

————— 主語の文法構造変換単独の例 —————

1 月 に その 講座 は 終わる。
おじさん の 口利き で 商社 に 彼は 就職 した。
社説 で 物価高 を 新聞 は 論じて いる。

2. 主語+目的語+助詞の文法構造変換

————— 主語+目的語+助詞の文法構造変換の例 —————

に 1 月 その 講座 は 終わる。
に 商社 で おじさん の 口利き 彼は 就職 した。
を 物価高 で 社説 新聞 は 論じて いる。

主語+目的語+助詞の文法構造変換の手法を図 6.1 に示す。

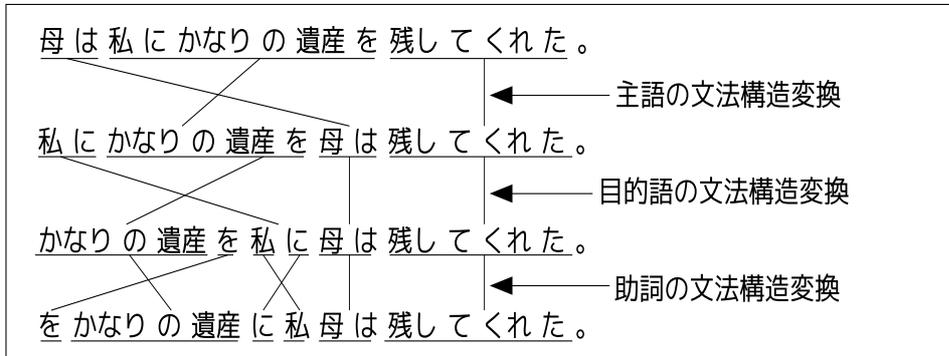


図 6.1: OSV(目的語, 主語, 動詞) への文法構造変換の翻訳

6.4.3 人手評価

PSMT と HSMT の人手評価の結果を表 6.12,6.13 に示す。

表 6.12: 主語の文法構造変換単独の人手評価

	ベースライン○	主語○	差なし	同一出力
PSMT	10	13	69	8
HSMT	10	11	72	7

表 6.13: 主語+目的語+助詞の文法構造変換単独の人手評価

	ベースライン○	主語+目的語+助詞○	差なし	同一出力
PSMT	8	10	75	7
HSMT	8	7	77	8

6.4.4 自動評価

自動評価の PSMT での結果を表 6.14 に示す。HSMT での結果は表 6.15 に示す。

表 6.14: PSMT の翻訳精度

	BLEU	NIST	METEOR	RIBES
主語	0.1399	4.8358	0.4653	0.7269
主語+目的語+助詞	0.1326	4.7590	0.4559	0.7178
ベースライン	0.1343	4.7646	0.4519	0.7196

表 6.15: HSMT の翻訳精度

	BLEU	NIST	METEOR	RIBES
主語	0.1393	4.8579	0.4614	0.7253
主語+目的語+助詞	0.1306	4.7174	0.4488	0.7098
ベースライン	0.1331	4.7383	0.4557	0.7193

6.4.5 評価結果のまとめ

主語の文法構造変換単独，主語+目的語+助詞の文法構造変換ともに，ベースラインと比べると，翻訳精度は向上しない。

6.4.6 PSMT での翻訳例

主語の文法構造変換単独，主語+目的語+助詞の文法構造変換の実験の翻訳例を以下に示す。なお，表中の下線は誤りと判断した箇所を表す。

1. 主語の文法構造変換単独

表 6.16: 主語単独が優れている翻訳例

例 1	
入力文	私は新しい事業を計画している。
主語	I plan to the new business .
ベースライン	I have <u>a gunpowder factory</u> .
正解文	I am planning another enterprise .
例 2	
入力文	夏の夜空に花火が揚がった。
主語	Fireworks were set off in the night sky of summer .
ベースライン	<u>Fireworks the summer in the night sky</u> .
正解文	The fireworks shot up into the summer night sky .
例 3	
入力文	試験方法の変更は遅れている。
主語	The change of the test method is slow .
ベースライン	<u>test method is a change of late</u> .
正解文	Changes in the way exams are run are overdue .

表 6.17: ベースラインが優れている翻訳例

例 1	
入力文	霜が消えた。
主語	<u>The frost</u> .
ベースライン	The frost has gone out .
正解文	The frost has gone .
例 2	
入力文	数学ではとても彼にかなわない。
主語	I am no match for him in mathematics <u>as he</u> .
ベースライン	I am no match for him in mathematics .
正解文	I am nowhere near as good as he in math .
例 3	
入力文	この鉱山は見込みがある。
主語	There is <u>no prospect of this mine</u> .
ベースライン	This mine prospers well .
正解文	The mine prospers well .

2. 主語＋目的語＋助詞の文法構造変換

表 6.18: 主語＋目的語＋助詞が優れている翻訳例

例 1	
入力文	それは確かに 10 ドルはした。
主語＋目的語＋助詞	It was a certainly ten dollars .
ベースライン	<u>It is a ten dollars .</u>
正解文	It cost me ten dollars , if a cent .
例 2	
入力文	彼の意見に文句はありません。
主語＋目的語＋助詞	I don't have any objection of his opinion .
ベースライン	I don't have the opinions of <u>the with him .</u>
正解文	I have no quarrel with his opinion .
例 3	
入力文	本が倒れた。
主語＋目的語＋助詞	The book fell .
ベースライン	<u>There are books .</u>
正解文	The book fell down .

表 6.19: ベースラインが優れている翻訳例

例 1	
入力文	被害者たちは警察に保護を求めた。
主語＋目的語＋助詞	<u>We urge the victims of the police protection .</u>
ベースライン	The victims asked the police protection .
正解文	The sufferers asked for police protection .
例 2	
入力文	ひとつの問題が別の問題を生じた。
主語＋目的語＋助詞	<u>There was a different problem is one of the problem .</u>
ベースライン	A problem has caused another problem .
正解文	One problem led to another .
例 3	
入力文	教育にはたいへんな金がかかる。
主語＋目的語＋助詞	<u>A great money on education .</u>
ベースライン	Education is very expensive .
正解文	Education is very expensive .

6.4.7 HSMT での翻訳例

主語の文法構造変換単独，主語＋目的語＋助詞の文法構造変換の実験の翻訳例を以下に示す．なお，表中の下線は誤りと判断した箇所を表す．

1. 主語の文法構造変換単独

表 6.20: 主語単独が優れている翻訳例

例 1	
入力文	彼らのチームは最初から試合の主導権を握った。
主語	They took the initiative in the match from the start .
ベースライン	<u>They team from the start of the initiative of the game .</u>
正解文	Their team controlled the match from the start .
例 2	
入力文	私は列車速度に注目する。
主語	I attention to the train speed .
ベースライン	<u>I will of the speed of a train .</u>
正解文	I pay attention to train speed .
例 3	
入力文	被告に有利な判決が下された。
主語	The decision was sentenced in favor of the defendant .
ベースライン	<u>The decided in favor ofthe accused was passed .</u>
正解文	A decision was given for the defendant .

表 6.21: ベースラインが優れている翻訳例

例 1	
入力文	雪の重みで木の枝が曲がっていた。
主語	The branches of the tree was bent under <u>way</u> under the weight of the snow .
ベースライン	The branches of the tree are bent under the weight of the snow .
正解文	The tree branches have bent under the snow .
例 2	
入力文	現場の責任者が交替する。
主語	The responsibility for the scene <u>of the hands</u> .
ベースライン	The responsibility for the scene changed .
正解文	The on-the-job superintendent is changed .
例 3	
入力文	警察は銀行強盗の張本人を捕まえた。
主語	The police caught <u>in a bank robber invisible author of all</u> .
ベースライン	The police caught the author of the bank robbery .
正解文	The police arrested the ringleader of the bank robbery .

2. 主語＋目的語＋助詞の文法構造変換

表 6.22: 主語＋目的語＋助詞が優れている翻訳例

例 1	
入力文	多くの会社が休業した。
主語＋目的語＋助詞	Many companies closed .
ベースライン	<u>Many closed for the company .</u>
正解文	Many businesses closed down .
例 2	
入力文	君の行為は全く我慢できない。
主語＋目的語＋助詞	I can't stand at your actions .
ベースライン	I can't <u>act of you .</u>
正解文	Your conduct is beyond bearing .
例 3	
入力文	彼は人の目を気にし過ぎる。
主語＋目的語＋助詞	He is too sensitive to the eyes of the people .
ベースライン	He is too <u>concerned</u> in the eyes of the people .
正解文	He is too sensitive about how he appears to people .

表 6.23: ベースラインが優れている翻訳例

例 1	
入力文	私は新しい事業を計画している。
主語＋目的語＋助詞	I plan <u>is</u> a new business .
ベースライン	I plan to the new business .
正解文	I am planning another enterprise .
例 2	
入力文	靴は 5 足持っている。
主語＋目的語＋助詞	<u>The shoes .</u>
ベースライン	I have five pairs of shoes .
正解文	I have five pairs of shoes .
例 3	
入力文	彼は車を私に預けた。
主語＋目的語＋助詞	<u>He deposited the car .</u>
ベースライン	He left a car on me .
正解文	He entrusted me with his car .

第7章 おわりに

星野ら [2] は述語項構造に基づく事前並べ替えを行った後で，句に基づく統計翻訳 (PSMT) を行うことで，翻訳精度が向上したことを報告している．しかし，星野らは特許文を使用している．そして，人手評価を行っていない．だが，特許文は文法構造が複雑で，人手評価が困難である．そこで，本研究では簡単な文法構造である単文コーパス [3] を使用し，動詞および目的語および助詞の文法構造変換を行い，翻訳精度の変化を調べた．

実験の結果，動詞の文法構造変換と助詞の文法構造変換の組み合わせは有効だが，目的語の文法構造変換は有効でなかった．また，HSMT において動詞の文法構造変換と助詞の文法構造変換の組み合わせが，最も高い翻訳精度であることが示された．

今後，提案手法の翻訳精度を向上するために，複合動詞や連続した助詞を処理する必要がある．

参考文献

- [1] 岡崎弘樹: 日本語文法構造変換による日英統計翻訳. 卒業論文, 2009.
- [2] 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, 2013.
- [3] 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, Mar. 2012.
- [4] 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), page 48-54, 2003.
- [5] Peter F Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, Robert L. Mercer: The Mathematics of Statical Machine Translation, Parameter Estimation. Computational Linguistics, 1993.
- [6] GIZA++: <http://www.fjoch.com/GIZA++>
- [7] Franz Josef Och: Minimum Error Rate Training in Statistical Machine Translation. Proceeding of the 41st Annual Meeting of the Association for Computational Linguistics, page 160-167, 2003.
- [8] 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
- [9] Andreas Stolcke: SRILM - an Extensible Language Modeling Toolkit. <http://www.speech.sri.com/projects/srilm> 7th International Conference on Spoken Language Processing pp.901-904 2002

- [10] 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 pages 177–180 June 2007
- [11] Jin'ichi Murakami, Masato Tokuhisa, Satoru Ikehara: Stastical Machine Translation using Large J/E Parallel Corpus and Long Phrase Tables. International Workshop on Spoken Language Translation 2007, pp.151-155, 2007.
- [12] Papineni Kishore, Salim Roukos, Todd Ward, Wei-Jing Zhu: BLEU,NIST: a method for automatic evaluation of machine translation. 40th Annual meeting of the Association for Computational Linguistics, pp.311–318, 2002.
- [13] 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
- [14] 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 年次大会発表論文集 pp.1111–1114 2011