

# 前処理に文節単位の語順変更を用いた日英統計翻訳

畑中勇輝 村上仁一 坂田純 徳久雅人  
鳥取大学大学院 工学研究科 情報エレクトロニクス専攻  
{s102039, murakami, d112004, tokuhisa} @ ike.tottori-u.ac.jp

## 1 はじめに

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 [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、述語項構造に基づいた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、語順変更の有効性を調査する。

## 2 日本語の語順変更による翻訳

### 2.1 動詞の語順変更 [3]

岡崎による動詞の語順変更の手法は日本語文を英語文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1に示す。

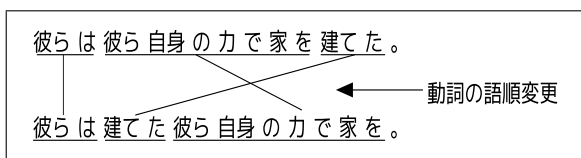


図1 動詞の語順変更の例

### 2.2 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。

## 3 提案手法

岡崎や星野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、本研究ではテスト文に対訳文があると仮定し、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、語順変更の有効性を調査する。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(3データ)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。図2に変数化の手順を示す。

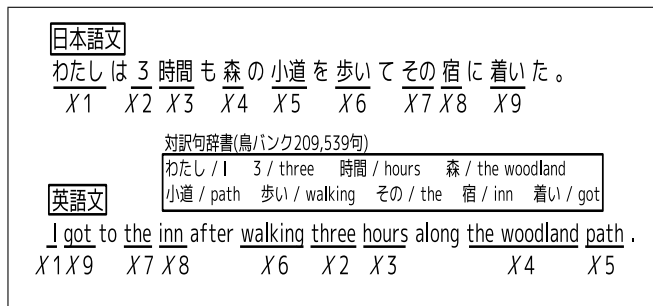


図2 変数化

### 手順2 語順変更

対訳文を用いて、“3データ”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。図3に語順変更の手順を示す。

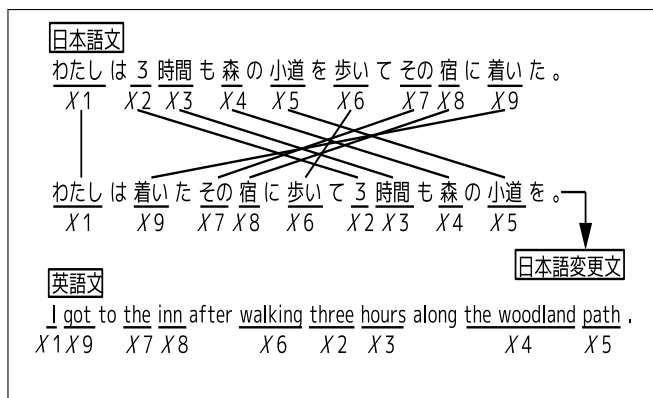


図3 語順変更

### 手順3 翻訳実験

“3データ”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(PSMT)を行う。

## 4 実験環境

### 4.1 実験内容

対訳句辞書と対訳文を用いて，“3 データ”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(PSMT)を行う。なお、翻訳は日英統計翻訳で行う。

### 4.2 実験データ

本研究では以下の2種類の実験データを用いて、実験を行う。

- 辞書の例文から抽出した対訳データの単文 [5]
- 辞書の例文から抽出した対訳データの重文複文 [5]

実験データの文数と平均単語数を表1に示す。

表1 実験データの文数

	単文	重文複文
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)

### 4.3 対訳句辞書

本研究では対訳句として、鳥バンク [6] 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), ADVP(形容詞句) の対訳句を抽出し、対訳句辞書を作成する。その結果、209,539 句の対訳句を抽出した。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 (AJ)
思わ	expected (V)
彼女	She (N)
いつも	always (ADV)
とても静かな	very quiet (AJV)
ぼくにあいさつし	greeted me (VP)
わたしの好きな	my favorite (AJVP)
長い間	for a long time (ADVP)

鳥バンクの対訳句の精度を調べるため、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2に示す。

表2 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、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

俺	father
国民	a people
手術	a surgeon
作れ	cultivate
はっきりし	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

### 4.4 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、単文と重文複文のテスト文から、それぞれランダムに 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3、重文複文の語順変更の精度を表4に示す。

表3 単文の語順変更の精度

語順変更○	語順変更×
90	10

表4 重文複文の語順変更の精度

語順変更○	語順変更×
87	13

単文では誤り率は 10%、重文複文では誤り率は 13% であった。単文の語順変更成功した例を表5、失敗した例を表6に示す。また、重文複文の語順変更成功した例を表7、失敗した例を表8に示す。なお、下線部は間違いと判断した箇所を表す。

u>

表5 単文の語順変更の成功例

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 .

表6 単文の語順変更の失敗例

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 ほど背が X3 に。
英語パターン	X1 is so X5 that his X2 X4 the X3 .

表7 重文複文の語順変更の成功例

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 .

表8 重文複文の語順変更の失敗例

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	彼は体の失って釣り合いを倒れたその場に。
日本語パターン	X1 は体の 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 .

失敗した原因として、単文では入力文の“背”，重文複文では入力文の“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。

#### 4.5 統計翻訳システム

本研究で用いる統計翻訳システムについて説明する。

##### 4.5.1 デコーダー

デコーダーには moses[7] を用いる。

##### 4.5.2 言語モデルの学習

N-gram モデルの学習には SRILM[8] を用いる。ただし、N-gram モデルは 5-gram に設定する。

##### 4.5.3 デコーダーに関するパラメータ

本研究ではデコーダーのパラメータは default 値とする。ただし、“ttable-limit” の値は “40” に設定する。

#### 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、“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、“3 データ” 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。自動評価には約 10,000 文のテスト文を用いて、BLEU[9], METEOR[10], RIBES[11], TER[12] を用いる。人手評価はテスト文からランダムに 100 文取り出し、ベースラインとの対比較評価を行う。

### 5 単文の実験

#### 5.1 自動評価

PSMT での自動評価の結果を表 9 に示す。なお、太字はスコアが高い方を表す。

表 9 単文の実験の自動評価結果

	BLEU	METEOR	RIBES	TER
提案手法	<b>0.1430</b>	<b>0.4617</b>	<b>0.7345</b>	<b>0.6820</b>
ベースライン	0.1340	0.4513	0.7024	0.7178

#### 5.2 人手評価

提案手法とベースラインの対比較評価の結果を表 10 に示す。

- 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

表 10 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

#### 5.3 翻訳例

提案手法の方が優れている翻訳例を表 11、ベースラインの方が優れている翻訳例を表 12 に示す。なお、下線部は間違いと判断した箇所を表す。

表 11 提案手法の方が優れている翻訳例

入力文	彼はわがチームの主将としてふさわしい。
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 .

表 12 ベースラインの方が優れている翻訳例

入力文	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 .

#### 5.4 単文の実験のまとめ

表 9, 10 から以下のことが示せる。

1. 自動評価では全ての評価において、翻訳精度が向上した。
2. 人手評価では差は見られなかった。

### 6 重文複文の実験

#### 6.1 自動評価

PSMT での自動評価の結果を表 13 に示す。なお、太字はスコアが高い方を表す。

表 13 重文複文の実験の自動評価結果

	BLEU	METEOR	RIBES	TER
提案手法	<b>0.1295</b>	<b>0.4255</b>	<b>0.7181</b>	<b>0.7412</b>
ベースライン	0.1175	0.4125	0.6749	0.7693

#### 6.2 人手評価

提案手法とベースラインの対比較評価の結果を表 14 に示す。

表 14 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

#### 6.3 翻訳例

提案手法の方が優れている翻訳例を表 15、ベースラインの方が優れている翻訳例を表 16 に示す。なお、下線部は間違いと判断した箇所を表す。

表 15 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。
提案手法○	These are the fact that we can not ignore .
ベースライン	These are not our ignored .
参照文	These are facts to which we cannot close our eyes .

表 16 ベースラインの方が優れている翻訳例

入力文	私なんぞがチームの代表だなんて分不相応だ。
提案手法	It is absurd that I am the representative of the team 不相応 minutes .
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 .

#### 6.4 重文複文の実験のまとめ

表 13, 14 から以下のことが示せる。

1. 自動評価では全ての評価において、翻訳精度が向上した。
2. 人手評価では差は見られなかった。

## 7 考察

### 7.1 “reordering-model” の効果

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。単文の実験の自動評価の結果を表 17, 対比較評価の結果を表 18, 重文複文の実験の自動評価の結果を表 19, 対比較評価の結果を表 20 に示す。

表 17 単文の実験の自動評価結果

	BLEU	METEOR	RIBES	TER
“reordering-model あり”	0.1340	0.4513	0.7024	0.7178
“reordering-model なし”	0.1336	0.4511	0.7044	0.7080

表 18 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

表 19 重文複文の実験の自動評価結果

	BLEU	METEOR	RIBES	TER
“reordering-model あり”	0.1175	0.4125	0.6749	0.7693
“reordering-model なし”	0.1140	0.4092	0.6694	0.7778

表 20 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

単文、重文複文ともに、自動評価では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。

### 7.2 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。しかし、これらの研究では翻訳精度の大きな向上が見られなかった。本研究においても、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ても、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。

## 8 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、語順変更の有効性を調査した。

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。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。

## 参考文献

- [1] Chenchen Ding, etc: Word Order Does NOT Differ Significantly Between Chinese and Japanese. Proceedings of the 1st Workshop on Asian Translation (WAT2014), pages 77-82, Tokyo, Japan, 4th October 2014
- [2] 村松航平, 他: 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: 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, 他: 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] 村上仁一, 他: 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
- [6] 鳥バンク: <http://unicorn.ike.tottori-u.ac.jp/toribank/>
- [7] Philipp Koehn, etc: 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, June 2007
- [8] Andreas Stolcke: <http://www.speech.sri.com/projects/srilm>, 7th International Conference on Spoken Language Processing, pp.901-904, 2002
- [9] Papineni Kishore, etc: BLEU,NIST: a method for automatic evaluation of machine translation. 40th Annual meeting of the Association for Computational Linguistics, pp.311-318, 2002
- [10] Banerjee Satanjeev, etc: 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
- [11] 平尾努, 他: 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
- [12] Gregor Leusch, etc: A Novel String-to-String Distance Measure with Applications to Machine Translation Evaluation. In Proc. of MT Summit IX, 240-247, TRANSLATION ERROR RATE (TER) 7.0 <http://www.cs.umd.edu/snover/tercom/> (2003)