

対訳句と対訳文を用いた日本語文の語順変更の効果

畑中勇輝 *1 村上仁一 *1 徳久雅人 *1

*1 鳥取大学大学院 工学研究科 情報エレクトロニクス専攻

*2 {s102039, murakami, 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、統計翻訳を行う研究が行われている。

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 [2] によると、述語項構造に基づいた語順変更を行った後で、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、本研究では日本語の語順が柔軟であることを利用して、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。

2 日本語の語順変更による翻訳

2.1 動詞の語順変更 [1]

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。

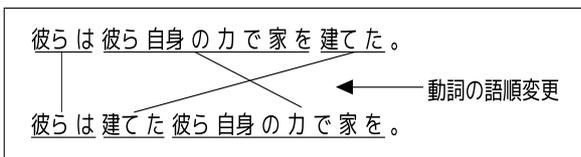


図1 動詞の語順変更の例

2.2 述語項構造に基づく語順変更 [2]

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、本研究では日本語の語順が柔軟であることを利用して、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。

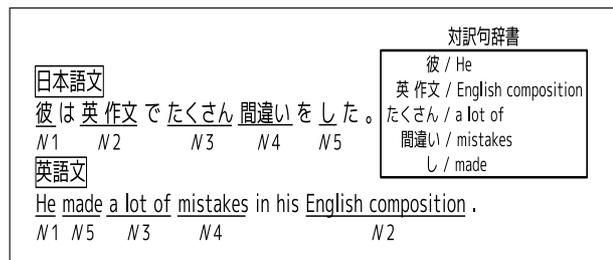


図2 変数化

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。

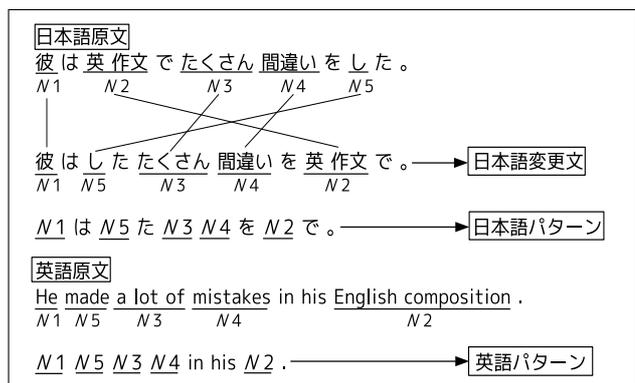


図3 語順変更

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，実験は句に基づく統計翻訳 (PSMT) を行う。なお，日英統計翻訳を行う。

4.2 実験データ

本研究では，電子辞書の例文から抽出した対訳データの単文コーパス [3] を使用する。実験に用いる文数を表 1 に示す。

表 1 実験に用いる文数

学習文	99,914 文
テスト文	9,995 文
ディベロップメント文	1,000 文

4.3 対訳句辞書

本研究では対訳句として，鳥バンク [4] 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。結果，213,511 句の対訳句を抽出する。以下に抽出した対訳句辞書の例を示す。

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>)

4.4 統計翻訳システム

4.4.1 デコーダー

デコーダーには *moses*[5] を用いる。

4.4.2 言語モデルの学習

N-gram モデルの学習には *SRILM*[6] を用いる。また，*N*-gram モデルは 5-gram とする。

4.4.3 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。しかし，“*ttable-limit*”の値は“40”とする。また，“*distortion-limit*”の値は，提案手法において，“6”に設定する。

4.5 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。また，“*distortion-limit*”の値は“-1”に設定する。

4.6 語順変更

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。

4.7 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*[7]，*NIST*[7]，*METEOR*[8]，*RIBES*[9]，*TER*[10]，*WER*[10] 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，ベースラインとの対比較実験を行う。

5 実験結果

5.1 語順変更の精度

本研究で用いる語順変更のプログラムの精度を調べるため，ディベロップメント文からランダムに 30 文抽出して，正確に語順変更されているか調査した。語順変更の精度を表 2 に示す。30 文中 4 文が語順変更失敗したことから，誤り率は約 13% である。

表 2 語順変更の精度

変更○	変更×
26	4

語順変更失敗した例を表 3 に示す。なお，下線部は間違いと判断した箇所を表す。入力文では“父”は“病氣”に係るのに対して，日本語変更文では“父”は“兄”に係り，語順変更失敗している。

表 3 語順変更失敗した例

入力文	彼は父の病氣を手紙で兄に知らせた。
参照文	He informed his brother of their father's illness in a letter .
変数情報	<i>N1</i> :彼/He <i>N2</i> :父/his <i>N3</i> :病氣/illness <i>N4</i> :手紙/a letter <i>N5</i> :兄/brother <i>N6</i> :知らせ/informed
日本語変更文	彼は知らせた父の兄に病氣を手紙で。
日本語パターン	<i>N1</i> は <i>N6</i> た <i>N2</i> の <i>N5</i> に <i>N3</i> を <i>N4</i> で。
英語パターン	<i>N1</i> <i>N6</i> <i>N2</i> <i>N5</i> of their father's <i>N3</i> in <i>N4</i> .

5.2 翻訳実験

5.2.1 自動評価

PSMT での自動評価の結果を表 4 に示す。なお，太字はスコアが高い方を表す。

表 4 自動評価結果

	提案手法	ベースライン
BLEU	0.1418	0.1377
NIST	4.9383	4.8628
METEOR	0.4625	0.4566
RIBES	0.7371	0.7106
TER	0.6798	0.6969
WER	0.6951	0.7230

5.2.2 人手評価

PSMT での人手評価の結果を表 5 に示す。

表 5 人手評価結果

ベースライン○	提案手法○	差なし	同一出力
7	10	74	9

5.3 実験結果のまとめ

表 4, 5 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. 人手評価では大きな差は見られなかった。

6 翻訳例

提案手法の方が優れている翻訳例を表 6, ベースラインの方が優れている翻訳例を表 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 提案手法の方が優れている翻訳例

例 1	
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。
提案手法○	The company is important in the local economy .
ベースライン	The local economy is important in <u>the presence of the company</u> .
参照文	The company bulks large in the local economy .
例 2	
入力文	我々 は 準備 が 非常 に 後 れ て いる 。
提案手法○	We are very behind in preparation .
ベースライン	We are ready for very underdeveloped .
参照文	We are badly behind in our preparations .
例 3	
入力文	煙 突 が もくもくと 煙 を 吐 いて いた 。
提案手法○	A chimney was emitting smoke .
ベースライン	A chimney <u>smoke</u> .
参照文	The chimney was sending up volumes of smoke .

表 7 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 。
提案手法	Our company has a brilliant engineers is 居り .
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	人 の 一 生 は し ば し ば ろ う そ く に た と え ら れ る 。
提案手法	Man's life is often <u>even if you're is the candle</u> .
ベースライン○	Man's life is often likened to candle .
参照文	Man's life is often compared to a candle .
例 3	
入力文	夜 明 け 前 に 海 は 穏 や か に な っ た 。
提案手法	The sea <u>quietly</u> before dawn .
ベースライン○	The sea smoothed down before dawn .
参照文	The sea had gone down before daylight came .

7 考察

7.1 自動評価と人手評価

人手評価では評価文数が少ないため、信頼性が低い。今後、評価文数を増やし、再び人手評価を行う。また、自動評価と人手評価で結果が異なるので、原因の調査を行う。

7.2 重文複文

人手評価で翻訳精度が向上しなかった原因として、単文では“reordering model”と語順変更の効果が重なったためだと考えている。今後、語順変更の効果を明確にするために、重文複文で実験を行う必要がある。

7.3 語順変更の精度

本研究で行った語順変更の誤り率は 13% であった。今後、語順変更の精度を向上させる必要がある。

7.4 distortion-limit

7.4.1 実験内容

提案手法では“distortion-limit”の値は“6”に設定した。しかし、岡崎 [1] によると、語順変更では“distortion-limit”は“-1”に設定した方が、高い翻訳精度が得られた。そこで、提案手法の“distortion-limit”の値を“-1”に設定し、実験を行った(以下、追加実験とする)。なお、その他の実験環境は 4 節と同じ環境で行った。

7.4.2 実験結果

PSMT での自動評価の結果を表 8 に示す。なお、太字はスコアが高い方を表し、括弧の中は“distortion-limit”の値を表す。

表 8 自動評価結果

	追加実験 (-1)	提案手法 (6)
BLEU	0.1413	0.1418
NIST	4.7984	4.9383
METEOR	0.4556	0.4625
RIBES	0.7268	0.7371
TER	0.7027	0.6798
WER	0.7206	0.6951

7.4.3 翻訳例

追加実験 (-1) の方が優れている翻訳例を表 9, 提案手法 (6) の方が優れている翻訳例を表 10 に示す。なお、下線部は間違いと判断した箇所を表す。

表 9 追加実験 (-1) の方が優れている翻訳例

入力文	何物も彼を満足させなかった。
追加実験 (-1) ○	He did not satisfy nothing .
提案手法 (6)	<u>did not</u> satisfy him for nothing .
参照文	Nothing contented him .

表 10 提案手法 (6) の方が優れている翻訳例

入力文	警官は男の手からさっとピストルを奪った。
追加実験 (-1)	The policeman <u>of the man who was</u> the gun from his hand .
提案手法 (6) ○	The policeman stole the gun from the man .
参照文	The policeman snatched the pistol from his hand .

7.4.4 実験結果のまとめ

表8から、語順変更を行う際は“distortion-limit”の値を“-1”ではなく、“6”に設定した方が翻訳精度が高いことが分かった。

7.5 テスト文

本研究ではテスト文に対訳文があることを前提とし、実験を行った。しかし、実際にはテスト文に対訳文は存在しない。この問題に対して、レーベンシュタイン距離を用いて、入力文に類似した文を検索して、語順変更を行う手法を提案する。以下、手法を説明する。

手順1 変数化

3節の手順1と同様に行う。

手順2 語順変更とパターン化

3節の手順2と同様に行う。同時に、日本語と英語のパターンを得る。

手順3 類似文検索

レーベンシュタイン距離を用いて、日本語原文から、入力文に類似した文を検索する。類似文検索の手順を図4に示す。

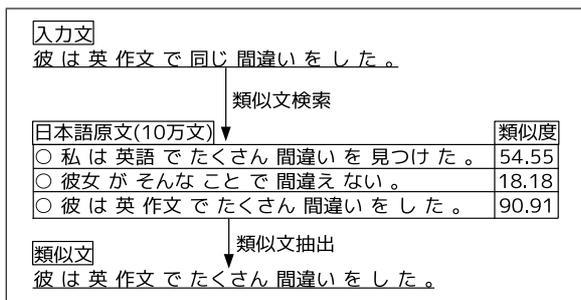


図4 類似文検索

手順4 テスト文の語順変更

手順3で得た類似文のパターンを用いて、語順変更を行い、出力をテスト文とする。テスト文の語順変更の手順を図5に示す。

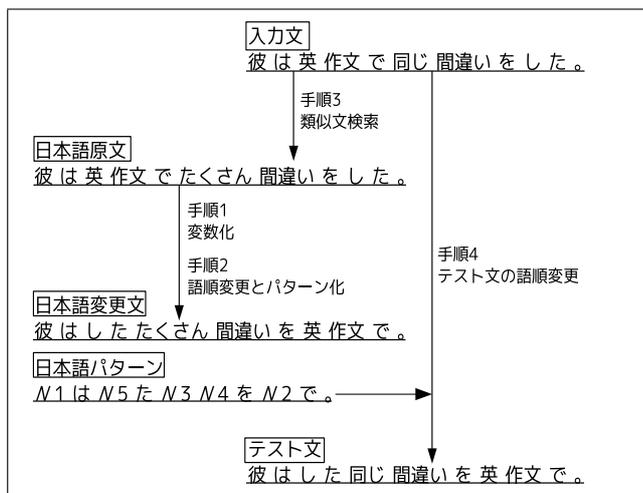


図5 テスト文の語順変更

今後、レーベンシュタイン距離を用いて、入力文に類似した文を検索して、語順変更を行う手法で実験を行い、翻訳精度の調査を行う予定である。

8 おわりに

岡崎や星野らの研究では、語順変更を行った後で、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、本研究では日本語の語順が柔軟であることを利用して、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した。

実験の結果、日本語文を英語文の語順に並び替えることで、自動評価では翻訳精度が向上した。しかし、人手評価では大きな差は見られなかった。

本研究ではテスト文に対訳文があることを前提とし、実験を行った。しかし、実際にはテスト文に対訳文が存在することは無い。したがって、今後の研究の予定として、7.4節に示したレーベンシュタイン距離を用いて、入力文に類似した文を検索して、語順変更を行う手法により、翻訳精度の変化を調べる。

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, 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, 2013
- [3] 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
- [4] 鳥バンク: <http://unicorn.ike.tottori-u.ac.jp/toribank/>
- [5] 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
- [6] Andreas Stolcke: am Extensible Language Modeling Toolkit. <http://www.speech.sri.com/projects/srilm>, 7th International Conference on Spoken Language Processing, pp.901-904, 2002
- [7] 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
- [8] 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
- [9] 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回年次大会発表論文集, pp.1111-1114, 2011
- [10] 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)