

複文における言い換え文の生成

駿河樹¹ 村上仁一²¹ 鳥取大学工学部電気情報系学科 ² 鳥取大学工学部

b20t2057u@edu.tottori-u.ac.jp

murakami@tottori-u.ac.jp

概要

本論文は、複文における日本語の言い換え文の生成を行う。

言い換え文の生成には様々な手法がある。松本 [1] が行った手法では、モデルの学習・言い換え文の生成に単文対訳文を使用した。本論文では、文長が長く複雑になりがちである複文対訳文を使用して同様の実験を行った。また実験結果の比較・検討を行った。その結果、複文は単文と比べると、意味を正しく保持した質の良い言い換え文が作成されやすい事が分かった。

1 はじめに

自然言語処理の分野ではこれまで言い換え文を作成する研究が多く行われている [2][3]。その方法の一例として機械翻訳を用いる手法がある。しかし言い換え文の精度は良くない。この原因として機械翻訳機の翻訳精度が悪いことが挙げられる。

この欠点を解決するため、松本 [1] は、折り返し翻訳を用いて対訳英文と折り返し翻訳した英文の2つを比較し、一致した場合に対訳日本語と出力日本語は言い換え文であるとした。この方法により正解率は約 80% となった。

松本は、実験データに比較的短い文である単文を使用している。しかし実際に日常で使用される文は複文が多い。そこで本研究では、松本の折り返し翻訳の実験方法を基に、複文を使用する。そして本研究と松本の研究結果との比較を行う。

1.1 単文

単文とは、日本語の述語が 1 語のみで構成された文である [4]。

例：信号が青になった。(The signal turned green.)

1.2 複文

複文とは、日本語の述語が 2 語以上で構成された文である [4]。

例：この詩はその詩を書いた人が実際に何を感じたかを私達に伝えている。(This poem conveys to us what its author actually felt.)

2 先行研究

松本は、「2つの文をそれぞれ翻訳し、翻訳結果が一致したとき、2つの文は言い換えである」という仮説を立てた。そして日英の対訳文に対して折り返し翻訳を行い、英語文が一致した場合、日本語文は言い換え文であるとした。以下に松本の手法の実験手順を示す。図 1 に概要を示す。

- OpenNMT で単文の日英対訳文約 16 万文を用いて学習する
- テストデータとして、1 の学習に使用した単文の英文を用いて英日翻訳を行う
- 2 の結果を基に日英翻訳を行う
- 3 の結果とテストデータ英文を比較し、完全一致した場合のみ、対訳日本語と 2 の結果が言い換え文であるとして出力する

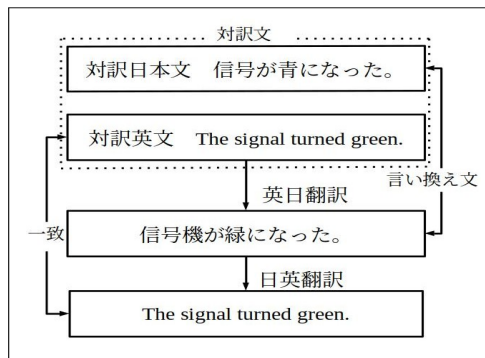


図 1 松本の例 (単文)

3 問題点

松本は、実験に使用する学習データに、日本語・英語の単文対訳文を使用していた。しかし日常で使われる文は多くが複文である。

4 実験

本研究でも、松本の仮説に基づいて、同様の手順で実験を行う。相違点として、実験に複文を使用する。複文は単語や構文構造が複雑になりがちである。そのため、機械翻訳の結果が一致する確率が低くなることが予想される。図2に概要を示す。

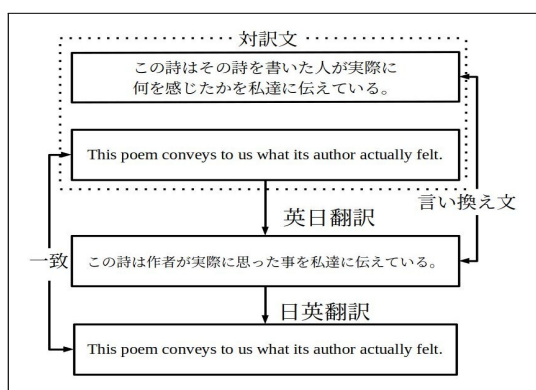


図2 本実験の例（複文）

4.1 実験条件

本研究では、松本と同様に OpenNMT[5] を用いて機械翻訳を行う。実験には、研究室が所有する、電子辞書データを基に作成した対訳コーパス [4] を使用する。以下の表1に使用するデータを示す。モデルの学習と実験には同じデータを使用する。また、以降は本研究の実験結果を「複文」、松本の実験結果を「単文」として表記する。

表1 実験データ

	複文	単文
学習データ・実験データ	92,427(複文)	163,188(単文)

4.2 実験手順

以下に本研究の実験手順を示す。

1. OpenNMT で複文の日英対訳文約9万文を用いて学習する
2. テストデータとして、1の学習に使用した複文の英文を用いて英日翻訳を行う
3. 2の結果を基に日英翻訳を行う
4. 3の結果とテストデータ英文を比較し、下に示

す複数の条件を満たした場合のみ、対訳日本文と2の結果が言い換え文であるとして出力する

条件

- 2の結果が対訳日本文と不一致。尚且つ、未知語<unk>を含まない
- 3の結果が対訳英文と完全一致

4.3 評価方法

本研究では、得られた言い換え文と対訳日本文が等しい意味を保っているか人手で評価を行う。ランダムに抽出した100文の言い換え文に対し、次の5項目で評価を行う。

- 正解 (T)：言い換えである
- 正解微妙 (ST)：判断し辛い为正解
- 不正解微妙 (SF)：判断し辛いが不正解
- 不正解 (F)：言い換えになっていない
- 参照間違い (R)：評価が不可能な文。原因としては、使用した対訳日本文または英文に著しい文法間違いがある・日本文と英文の意味に対応が取れていない等がある。正解率の計算時には除外

4.4 実験結果

表2に複文の実験結果を示す。単文の実験結果も比較として記載する。カバー率は、テストデータ数に対して生成言い換え文が生成された割合である。

表2 実験結果

	複文	単文
テストデータ	92,427	163,188
生成言い換え文	10,790	25,003
カバー率 (%)	12	15

複文のカバー率は単文と比べて3%減少しているが、大きくは変化していない。

4.5 出力例

表3, 4に複文の言い換え文出力結果と人手評価結果の例を示す。

表3 複文の出力例1

対訳日本文	この頃クラブ活動に忙しくて勉強がお留守になってしまった。	評価 A	評価 B
対訳英文	Lately I have been so busy with club activities that I have been neglecting my studies.		
言い換え文	この頃クラブ活動に忙しくて勉強を怠ってしまった。	T	T

表4 複文の出力例2

対訳日本語	彼は交際嫌いなのではなくて、ただ内気なだけだ。	評価 A	評価 B
対訳英文	He is not antisocial, just shy.		
言い換え文	彼は交際嫌いで、ただ内気なだけではない。	F	F

4.6 評価結果

以下の表5に人手評価の結果を示す。なお、評価者はA、Bの2人である。

表5 人手評価の結果

	複文 (A)	複文 (B)	単文 (A)	単文 (B)
テストデータ数	100			
正解 (T)	83	87	70	79
正解微妙 (TS)	5	2	10	0
不正解微妙 (FS)	3	0	6	9
不正解 (F)	6	7	13	11
参照間違い (R)	3	4	1	1
正解率 (%)	91	93	81	80

複文の正解率は平均して92%となった。言い換えの精度としては非常に高い数値と言える。

5 考察

5.1 単文との比較・検討

松本の実験結果との比較を行う。ただし、使用するデータや評価方法等、条件が異なる部分があるため、比較結果は必ずしも信頼されるものではない。

5.1.1 カバー率

表2より、単文に対し、複文の方がカバー率が低くなった。予想の通り、複文は文長が単文に比べて大きいと、折り返し翻訳後の英文が一致しにくくなるのが原因と思われる。

5.1.2 正解率

表5より、複文における人手評価の正解率は約92%、単文においては約80%である。複文は単文と比べカバー率が低いが、正解率は上回っている。これは、文長が長くなり英文一致の条件を満たす確率は落ちた一方で、条件を満たした文は、日本語の意味を保持した質の良い言い換え文である確率が高くなったためだと考える。

5.2 評価が分かれた例

以下の表6、7に複文において人手評価の結果が分かれた例をいくつか示す。

表6 評価が分かれた例1

対訳日本語	困っている人から搾り取るなんて何と浅ましいことだ。	評価 A	評価 B
対訳英文	With all his debts, he is perfectly calm and serene.		
言い換え文	困っている人に搾り取るなんて何と浅ましいことだ。	T	F

表7 評価が分かれた例2

対訳日本語	神は自らを助ける者を助ける。	評価 A	評価 B
対訳英文	God helps those who help themselves.		
言い換え文	神は自ら助ける者を助ける。	F	T

単語間に、新たな助詞等が挟まる又は置き替わることで、単語の修飾する対象が変わるケースがある。これについて、評価者によって正解・不正解の基準が大きく異なると推測する。

5.3 「質の良い言い換え文」について

生成された言い換え文において、原文（対訳日本語）と比べると、代名詞・句読点・助詞の有無やひらがなと漢字の違いといった、構文は全く一緒で一部のみが異なる文が非常に多く生成される。これを「質が悪い言い換え文」と定義する。一方で、単語の順番が入れ替わる・原文と全く異なる単語を使用しているといった文のことを「質の良い言い換え文」と定義する。

以下の表8が質の悪い言い換え文とした例である。質の良い言い換え文の例として表3などがある。

表8 質の悪い言い換え文の例

対訳日本語	そのような遠い信号を確認することは難しい。	評価 A	評価 B
対訳英文	It is difficult to identify a distant signal like that.		
言い換え文	そのような遠い信号を確認するのは難しい。	T	T

人手評価を行った100文について、質の良い言い換え文の数を調査した。以下の表9に複文・単文における質の良い言い換え文の数を示す。なお、定義が曖昧であり基準の統一は難しいため本調査は1人のみで行った。

表9 質の良い言い換え文の数

	複文	単文
テストデータ数	100	
質の良い言い換え文	27	9

単文と比べて複文は質の良い言い換え文の数が多いことが分かる。

それぞれ出力された言い換え文を調査すると、単文を学習したモデルでは、'the' や'they', 'it' 等の、

日本語訳では通常省略する英単語を忠実に訳しているケースが非常に多かった。一方複文を学習したモデルでは同様の問題は軽減されていた。

5.4 参照間違い (R) について

人手評価の項目の一つに参照間違い (R) を適用した。本実験で使用したコーパスには 2023 年現在では使用されない・意味が通じない文法や単語が使用された日本語が一定数含まれる。本実験において参照間違いにあたる文を全て事前に取り除くことは出来なかったため、評価項目に追加した。

6 N-best 及び M-system

6.1 N-best

N-best とは翻訳の出力結果を第 N 候補まで出力を得る手法である。本章では、一つの文に対する言い換え文の出力数を増やすことで、言い換えのバリエーションを増やすことが出来ることを期待する。なお、本実験では N=4 として実験を行う。

以下に 4-best における実験手順を示す。

1. OpenNMT に複文の日英対訳文約 9 万文を用いて学習する
2. テストデータとして 1 の学習に使用した複文の英文約 9 万文を用いて、英日翻訳を 1 文に対して 4 候補出力する
3. 2 のそれぞれの候補に日英翻訳を 4 候補出力する
4. 3 の結果とテストデータ英文を比較し、完全一致した場合のみ、対訳日本語と 2 の結果が言い換えであるとして出力する

6.2 M-system

M-system とは、M 個の異なるモデルを作成し、言い換え文の出力数を増やす手法である。翻訳モデル OpenNMT において、モデルの学習の際にシード値を変化させることで、同一の学習データを用いながら異なる翻訳結果を出力するモデルを作成することが出来る。

本節では、M=4 として 4 つの翻訳モデルを作成し、前章 4-best の実験を 4 回行うことで言い換え文の数をさらに増やす。6.1 節の 4-best の実験手順を異なるモデルで 4 回行う。

6.3 実験結果

以下の表 10 に 4-best・4-system の実験結果を示す。なお、4 章の実験結果の一部を 1-best として記載する。

表 10 4-best・4-system の実験結果

	4-system	4-best	1-best
テストデータ		92,427	
生成言い換え文	377,511	140,136	10,790

1-best と比較して、4-best・4-system では非常に多くの言い換え文を生成することが出来た。

6.4 評価結果

以下の表 11 に 4-best・4-system の評価結果を示す。評価者は 1 人である。

表 11 4-system の人手評価の結果

	4-system	4-best	1-best
テストデータ数		100	
正解 (T)	75	88	83
正解微妙 (TS)	6	2	5
不正解微妙 (FS)	3	2	3
不正解 (F)	14	7	6
参照間違い (R)	2	1	3
正解率 (%)	83	91	91

1-best と比べると、4-best の正解率はほぼ変動しなかった。4-system は 83 % となり、約 8 % 低下した。

7 おわりに

本研究では、複文を用いた言い換え文の生成を行った。松本が提案した手法で、実験データに複文を使用した言い換え文を作成した。その結果、高い正解率の言い換え文を多く生成することが出来た。また実験結果について単文と比較・検討を行った。カバー率は複文は 12 % となり、単文の 15 % と比べると低下した。一方正解率は複文は 92 % となり、単文の 80 % を大きく上回った。質の良い言い換え文の数についても、複文は単文を大きく上回る結果となった。今後、複文言い換え文を使用したコーパスの拡張により翻訳機の性能向上を検討する。

謝辞

人手評価に参加して下さった以下の方に感謝します。

宮本 歩

参考文献

- [1] 松本武尊. 折り返し翻訳を用いた言い換え生成. 自然言語処理学会, 2023.
- [2] Jianing Zhou and Suma Bhat. Paraphrase generation: A survey of the state of the art. In **Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing**, pp. 5075–5086, 2021.
- [3] 福田治輝, 綱川隆司, 大島純, 大島律子, 西田昌史, 西村雅史ほか. 逆翻訳による言い換え生成を用いた協調学習自動評定の性能改善. 第 82 回全国大会講演論文集, Vol. 2020, No. 1, pp. 535–536, 2020.
- [4] 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.
- [5] OpenNMT. <https://opennmt.net/>.