

## 入出力文間の同義判定による語順整序の精度向上

吉田 嵐<sup>†, a)</sup> 大野 誠寛<sup>†, b)</sup> 松原 茂樹<sup>‡</sup>  
 東京電機大学 大学院未来科学研究科<sup>†</sup> 名古屋大学 情報基盤センター<sup>‡</sup>

## 1 はじめに

語順整序 [1, 2, 3] とは、意味は伝わるものの読みにくい語順を持った文 (例文 I) を入力とし、その文の文節を並び替えることにより、読みやすい文 (例文 O1) に整えるタスクである。

例文 I : 可能性を子供のことを考えて話した。  
 例文 O1 : 子供のことを考えて可能性を話した。

しかし、語順整序の結果、入力文と異なる意味に解釈される文 (例文 O2) や、不自然な文 (例文 O3) が間違っ出て出力されることがある。

例文 O2 : 可能性を考えて子供のことを話した。  
 例文 O3 : ことを考えて子供の可能性を話した。

このような出力をいかに回避するかが、語順整序タスクの課題となっている。

そこで本稿では、既存の語順整序手法に対して、入力文と出力文の文意が同一か否かを判定する機構を導入することにより、語順整序の精度向上を試みる。

## 2 同義判定

同義判定は、単語と単語 [4]、文と文 [5]、テキストとテキスト [6] など、各種言語単位のペアが同一の意味を持つか否かを判定するタスクである。このうち、本研究と同様に文間の同義判定を行う研究として、言い換え表現の検出のために、異なる語彙を持った文間で同義判定を行う研究 [5] が一部存在する。本研究では、語順整序の入出力を対象とするため、語順のみが異なり、同一の語彙で構成された文間で同義判定を行うが、そのような研究は調査の限り存在しない。

一方、同義判定の類似タスクに含意関係認識があるが、その中では、同一の語彙で構成された文間で含意関係を認識する研究 [7] が存在する。この研究では、文中に登場する名詞のみに注目し、それらの語順を入れ替えた文間を対象としている。本研究では、名詞に限定することなく、語順を入れ替えた文間を対象とする。

## 3 提案手法

提案手法の概要を図1に示す。本研究では、既存の語順整序手法 [3] と同様に、意味は伝わるものの読みにくい語順を持った文の文節列を入力とする。提案手法では、まず、入力文に対して既存手法 [3] を適用し、この手法におけるスコアでランキングされた  $n$  個の出力文 (Top-1 から Top- $n$ ) を得る。次に既存手法 [3] の出力文 Top-1 から順に、入力文と出力文 Top- $i$  のペアを作成し、各ペアに対して同義判定を適用することを繰り返す。入力文と出力文 Top- $i$  が同義と判定された時点

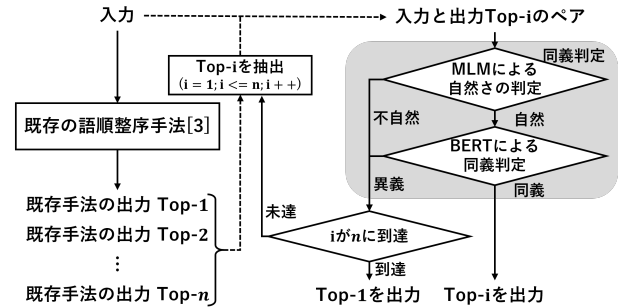


図1: 提案手法の概要

で、その Top- $i$  を最終的な語順整序結果として出力する。全  $n$  文が入力文と同義でないと判定された場合は、出力文 Top-1 を最終的な語順整序結果として出力する。

## 3.1 既存の語順整序手法 [3]

本研究では、既存の語順整序手法として、係り受け解析・語順整序・読点挿入を同時実行する手法 [3] を採用する。この手法は、Shift-Reduce 法において、Shift と Reduce の他に3つの操作と、語順変更用スタックを追加することにより、同時実行を実現している。また、ビームサーチを導入しており、各時刻での操作を分類モデルで選択する際の分類スコアに基づいてランキングした上位  $n$  個の語順整序文を出力できる。

## 3.2 同義判定

同義判定の概要を図1の灰色部分に示す。同義判定では、既存の語順整序手法 [3] における入力文と出力文 Top- $i$  の2文が同義か否かを2段階で判定する。

## 3.2.1 MLM による自然さの判定

1段階目では、入力文と出力文 Top- $i$  のペアを入力とし、出力文の自然さを判定する。具体的には、MLM scoring [8] を参考に、事前学習済みの BERT を用いて、入力文と出力文 Top- $i$  の各々に対して自然さのスコア ( $S_I, S_{O_i}$ ) を算出し<sup>\*1</sup>、両者の差 ( $S_{O_i} - S_I$ ) が閾値以上であれば、出力文 Top- $i$  は自然であると判定する。本研究では、意味は伝わるものの読みにくい語順を持った文を入力とし、入出力文の間で語彙の変化はなく語順のみが変化するという問題設定を考慮し、 $S_{O_i}$  の値のみで判定するのではなく、 $S_{O_i}$  が  $S_I$  よりも著しく低下した場合に出力文 Top- $i$  を不自然と判定する。

## 3.2.2 BERT による同義判定

2段階目では、1段階目で自然であると判定された入出力文ペアを入力とし、ファインチューニングした BERT を用いて、そのペアの入出力文が同義であるか否かの2値の判定を行う。BERT への入力は、入力文と出力文 Top- $i$  をサブワード分割し、[CLS] 入力文 [SEP] 出力文 Top- $i$  [SEP] の順に結合したものとする。BERT の出力のうち、[CLS] に対応する出力のみを取り出し、

<sup>\*1</sup>各文の各トークンを1つずつマスクし、MLMにより元のトークンの出現スコアを算出することを繰り返し、その和平均をその文の自然さのスコアとする。

Improving Japanese Word Reordering by Recognizing Semantic Equivalence between Input and Output Sentences

Ran Yoshida<sup>†, a)</sup>, Tomohiro Ohno<sup>†, b)</sup>, Shigeki Matsubara<sup>‡</sup>

<sup>†</sup> Graduate School of Science and Technology for Future Life, Tokyo Denki University.

<sup>‡</sup> Information Technology Center, Nagoya University

a) 25fmi35@ms.dendai.ac.jp b) ohno@mail.dendai.ac.jp

2 層の Linear 層と Softmax を介し両ラベルの分類確率を得て、その確率が高いラベルを出力とする。

学習には、語順整序の入出力文が両者とも自然で、かつ、同義であるペア（正例）と、両者とも自然で、かつ、異義となるペア（負例）が必要になる。本研究では、京大コーパスの各文（以下、元文）から、機械的に正例と負例を大量に作成する。まず、元文に対して、語順をランダムに変更した文を機械的に作成する。その中で、元文の係り受け構造が維持されるとした場合に、語順が変更されても、係り受けの非交差性と後方修飾性を満たす文を入力・同義出力候補、満たさない文を異義出力候補とした。これらには、不自然な語順の文が存在し得るため、それぞれに 3.1 節の MLM フィルタリングを適用し、自然と判定されたもののみを候補として残す。入力・同義出力候補、異義出力候補、元文の 3 つからペアを作成し、入力・同義出力候補同士のペア、または、入力・同義出力候補と元文のペアを正例、入力・同義出力候補と異義出力候補のペアを負例とした。

## 4 評価実験

語順整序において同義判定を導入することの有効性を検証するために、語順整序実験を行った。

### 4.1 実験概要

テストデータと開発データには、京大テキストコーパス Ver.4.0[9] を元に人手を介して疑似的に作成された読みにくい語順の文、各 1,000 文（既存研究 [3] と同一）を使用した。同義判定モデルの学習データには、京大テキストコーパスのうち、上記 2,000 文とは異なる 31,798 文に対して、3.2.2 節の手順を適用し得られた正例 275,882 ペアと負例 183,276 ペアを用いた。

比較手法として、同義判定を導入せず、既存手法 [3] のみを適用する手法を用意した。また、ビーム幅の影響を確認するため、提案手法と比較手法ともに、ビーム幅を 5 および 10 に設定したものを用意した。

同義判定モデルは、Pytorch を用いて実装し、東北大学が公開しているモデル<sup>\*1</sup>を使用した。学習はミニバッチ学習（バッチサイズ 32、学習率  $1e-6$ ）で行い、エポック数は 10 とした。MLM フィルタの閾値は  $-1.824$  とした。これらハイパーパラメータは開発データを用いて決定した。

評価では、既存研究 [3] と同様に、語順整序の文単位での正解率（正解の語順と完全一致している文の割合）に加え、2 文節単位正解率（文末以外の文節を 2 つずつ取り上げたとき、前後関係が正解と一致している割合）を測定した。

### 4.2 実験結果

実験結果を表 1 に示す。いずれのビーム幅においても、提案手法は比較手法と比べて、文単位正解率、2 文節単位正解率ともに上回った。

図 2 に、提案手法が正解し、比較手法が不正解となった例を示す。Top-1 から Top-3 までを異義と判定し不正解の出力を回避したうえで、Top-4 を同義と判定し正解を出力することに成功しており、同義判定が上手く機能していることが分かる。以上より、同義判定を導入することの有効性を確認した。

表 1: 実験結果

手法	ビーム幅	文単位	2 文節単位
比較手法 (同義判定 なし)	5	61.60 (616/1,000)	90.17 (28,638/31,760)
	10	61.60 (616/1,000)	90.22 (28,654/31,760)
提案手法 (同義判定 あり)	5	62.90 (628/1,000)	90.43 (28,723/31,760)
	10	62.90 (628/1,000)	90.56 (28,763/31,760)

入出力	入出力	比較手法	提案手法	
			同義判定	最終出力
Top-1	国会議員に次いで、政令市の、各府県議会議員市議らにも結集を呼びかけていく。	出力 (不正解)	異義	--
Top-2	国会議員に次いで、政令市の、各府県議会議員市議らにも結集を呼びかけていく。	--	異義	--
Top-3	国会議員に次いで政令市の、各府県議会議員市議らにも結集を呼びかけていく。	--	異義	--
Top-4	国会議員に次いで、各府県議会議員、政令市の市議らにも結集を呼びかけていく。	--	同義	出力 (正解)

\*語順と読点は同一だが、係り受け構造が異なる。

図 2: 同義判定の成功例

次に、ビーム幅が 5 の場合と 10 の場合を比較する。両手法とも、2 文節単位正解率がわずかに向上したものの、文単位正解率は変化しなかった。一方、ビーム幅を 5 および 10 とした場合の各ビーム内に正解が含まれるものはそれぞれ、1,000 件中、773 件、808 件であった。ビーム幅を広げることで正解の語順を選択できる可能性は高まるものの、現状では最良の文をビーム内から選択できていないことが示唆される。同義と判定されるものに対するリランキングなどが考えられるが、今後の課題である。

## 5 おわりに

本稿では、BERT を用いた同義判定の導入により、語順整序の精度が向上することを確認した。今後は、同義判定モデルのさらなる精緻化、リランキングの導入などを行い、より高品質な語順整序を目指したい。

謝辞 本研究は JSPS 科研費 JP19K12127, JP24K15076 の助成を受けたものです。

## 参考文献

- [1] 内元ら、コーパスからの語順の学習。自然言語処理 7(4), pp. 163–180, 2000.
- [2] 大野ら、係り受け解析との同時実行に基づく日本語文の語順整序。信学論, J99-D(2), pp. 201–213, 2016.
- [3] 荒木ら、処理途中での非文生成の回避を考慮した日本語文に対する係り受け解析・語順整序・読点挿入の同時実行。NLP2024 発表論文集, pp. 3182–3186, 2024.
- [4] R. Yatabe, et al. Semi-supervised word sense disambiguation using example similarity graph. *Proc. TextGraphs 2020*, pp. 51–59, 2020.
- [5] R. Socher, et al. Dynamic pooling and unfolding recursive autoencoders for paraphrase detection. *Advances in neural information processing systems*, Vol. 24, pp. 801–809, 2011.
- [6] D. Bogdanova, et al. Detecting semantically equivalent questions in online user forums. *Proc. CoNLL 2015*, pp. 123–131, 2015.
- [7] 雨宮ら、名詞役割入れ替えに頑健な日本語含意関係認識。NLP2023 発表論文集, pp. 1001–1006, 2023.
- [8] J. Salazar, et al. Masked language model scoring. *Proc. ACL 2020*, pp. 2699–2712, 2020.
- [9] 黒橋ら、京都大学テキストコーパス・プロジェクト。NLP1997 発表論文集, pp.115–118, 1997

\*1 tohoku-nlp/bert-base-japanese-wholeword-masking