BERT を用いた日本語文のボトムアップ語順整序とその評価

山添 壮登 [1,a)]　大野 誠寛 [1,b)]　松原 茂樹 [2,c)]
[1] 東京電機大学大学院未来科学研究科　[2] 名古屋大学情報連携推進本部
[a)] 21fmi21@ms.dendai.ac.jp　[b)] ohno@mail.dendai.ac.jp
[c)] matsubara.shigeki.z8@f.mail.nagoya-u.ac.jp

概要

文法的な誤りがないだけでなく，読みやすい語順の文を生成する技術は文生成において重要な技術となる．本稿では，係り受け関係が既知という前提のもと，任意の受け文節に係る文節集合内の文節を適切な順序に並べ，1 文全体をボトムアップに語順整序する手法を提案する．本手法では，同一文節に係る 2 文節間の前後関係を BERT により推定し，その確率値を用いて語順を決定する．語順整序実験及び主観的評価を行い，本手法の有効性を確認した．

1　はじめに

日本語は語順が比較的自由であるといわれているが，語順に関する選好がないわけではない [1]．そのため，文法的な誤りがないだけでなく，読みやすい語順であることも文生成において重要となる．

語順整序に関する研究は，推敲支援や文生成などへの応用を目的に，これまでにもいくつか行われている [2, 3, 4, 5, 6, 7, 8, 9, 10, 11]．このうち，内元ら [2] は，日本語の語順の傾向をコーパスから学習する手法として，日本語における語順決定に関わる様々な要因を素性として用いて，統計的に語順を整える手法を提案している．また，高須ら [3] は，文生成への応用を目的に，内元らの素性に加えて RNN 言語モデル (RNNLM) を用いた語順整序手法を提案している．内元ら [2] や高須ら [3] は，同じ文節に係るもの同士でまとめた各文節集合を対象として語順整序を行っているが，1 文全体の語順整序は行っていない．また，Kuribayashi ら [4] は，日本語文の語順を評価するために言語モデルを利用することを提案し，語順分析において言語モデルを用いることの有用性を示した．しかし，1 文中の動詞や文節数を制限しており，複雑な文を対象としていない．

本稿では，文生成のための要素技術として，1 文を構成する全ての文節の集合を読みやすく並べる手法を提案する．具体的には，文節間の係り受け関係は既知であることを前提として，1 文の係り受け構造を表す木（以下，係り受け木）に対して，BERT[12] に基づくモデルをボトムアップに適用し，語順整序を行う手法を提案する．語順整序実験の結果，先行研究を上回る精度を達成した．

2　文生成における語順整序

本研究では，文を構成する文節集合と，それら文節間の係り受け関係は既知であるとして，それらを入力とし，その入力文節集合内の文節を読みやすく並べることを試みる．これらの入力が既知であるとの仮定は，文生成や機械翻訳への応用を念頭においたものであり，文を生成するにあたって，その文で表したい内容は決まっている状況を想定したものである．この仮定は，内元ら [2] や高須ら [3] の先行研究の問題設定においても見られるものである．

内元ら [2] は，1 文の係り受け構造は既知であるとして，任意の受け文節 b_r に係る文節の集合 $B_r = \{b_{r_1}, b_{r_2}, \cdots, b_{r_n}\}(n \geq 2)$ に対して，B_r から考えられる順列 $\{\mathbf{b}^k | 1 \leq k \leq n!\}$ の中で最も読みやすい順列を求める問題として語順整序を定義している．ここで，$\mathbf{b}^k$ は k 番目の順列とする．内元らは，構文情報を中心とした素性に基づいた最大エントロピーのモデルを用いて語順整序する手法を提案している．

また高須ら [3] は，内元ら [2] の問題設定を引継ぎ，内元らの素性から主な素性[1)]を選択して学習した SVM によるモデルと，自然な語順をとらえることが期待できる RNN 言語モデルを併せて用いた．具体的には，B_r から考えられる全ての順列 $\{\mathbf{b}^k | 1 \leq k \leq n!\}$ の中から，式 (1) の $Score(\mathbf{b}^k)$ が最大となる順列 $\mathbf{b}^k$ を求めている．

$$Score(\mathbf{b}^k) = \alpha S_{RNNLM}(\mathbf{b}^k) + (1-\alpha) S_{SVM}(\mathbf{b}^k) \quad (1)$$

ここで，$S_{RNNLM}(\mathbf{b}^k)$ と $S_{SVM}(\mathbf{b}^k)$ はそれぞれ

1)　係り受け関係によって接続された各文節の形態素情報．

RNNLM と SVM によるモデルを用いて求めた $\mathbf{b}^k$ のスコアを意味する．

しかし，上記の両研究はともに，1 文を構成する全ての文節の集合の部分集合である B_r のみを語順整序の対象にしており，実際には 1 文全体の語順整序を行っておらず，1 文単位での評価も行っていない．また高須らの研究では，各文節を修飾する文節列を考慮せず，B_r 内の各文節を単に並べ替えた単語列（すなわち，1 文全体を並び替えた際には実際には出現すると限らない文字列）に対して RNNLM を適用しスコアを計算している．

3 BERT を用いたボトムアップ語順整序

本手法では，1 文全体を構成する文節集合と，その係り受け構造を入力とし，入力文節集合内の文節を読みやすく並べたものを出力する．その際，1 文の文節集合とその係り受け構造を表す係り受け木を作成し，その木に対してボトムアップに処理を施す．また，BERT を用いて 2 文節間の前後関係の判定を繰り返し，1 文全体を語順整序する．

3.1 ボトムアップ処理

本手法におけるボトムアップ処理を以下に示す．

1. 入力の文節集合と，その係り受け構造を表す係り受け木を作る．具体的には，入力文の各文節を 1 つのノードとして配置する．係り受け関係を表すエッジを用いて，それらの間を結ぶ．なお，以下の手順 2 と手順 3 において，複数のノードが 1 つのノードにまとめ上げられる操作がある．したがって，ノードは文節列（長さ 1 も含む）を表し，エッジは子ノード（の最終文節）が親ノードに係る係り受け関係を表すものとする．
2. 葉ノードのみを子に 1 つもつ親ノードと，その子ノードとをまとめ上げ 1 つのノードにする．その際，係り受けの後方修飾性を考慮し，子と親とをこの順に連接した文節列を新たなノードとする．この手順は，葉ノードのみを子に 1 つもつ親ノードがなくなるまで繰り返す．
3. 葉ノードのみを子に複数もつ親ノードと，その子ノード集合とをまとめ上げ 1 つのノードにする．その際，BERT によるモデルを用いて子ノード集合内の適切な語順を求め，その語順で連接した文節列の後ろに親ノードを繋げた文節列を新たなノードとする．

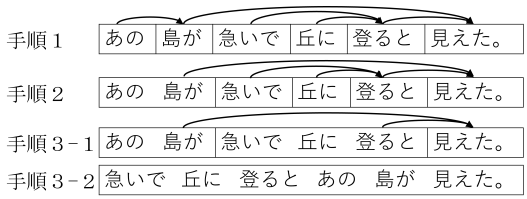


図 1 ボトムアップな語順整序の例

4. 手順 2 と手順 3 を，係り受け木が根ノード 1 つになるまで繰り返す．

上記手順の具体例を図 1 に示す．まず手順 1 では，入力の 6 文節とそれらの間の 5 つの係り受け関係から，図 1 の最上部の係り受け木を作る．手順 2 では，葉ノードのみを子に 1 つ持つ親ノードである「島が」と，その子である「あの」が「あの島が」にまとめ上げられる．手順 3-1 では，葉ノードである「急いで」と「丘に」の語順が計算され，双方の親ノードである「登ると」と共に，計算結果の語順に従い「急いで丘に登ると」にまとめ上げられる．手順 3-2 では，共通する親ノード「見えた.」の子ノードであり葉ノードの「あの島が」と「急いで丘に登ると」の語順が計算され，それら 3 つのノードが「急いで丘に登るとあの島が見えた.」という 1 つのノードにまとめ上げられる．係り受け木が根ノードのみになったため，語順整序が完了する．

なお，高須らの手法 [3] における非ボトムアップ処理では，根ノード「見えた。」に係る子ノードを語順整序をする際に，子ノード「島が」と「登ると」は子孫のノードとまとめ上げられることはなく，これらの文節のみからなる「島が登ると」と「登ると島が」のどちらがより読みやすいかを，機械学習モデルを用いて判断することになる．

3.2 BERT を用いた語順整序

3.1 節で説明したように，BERT[12] に基づくモデルを用いて同一の親を持つ兄弟ノードを語順整序する．ここでは，ノード v_r の子ノード集合 $V_r = \{v_{r_1}, v_{r_2}, \cdots, v_{r_n}\}$ を語順整序する際の計算について説明する．親ノード v_r の子ノード集合 $V_r = \{v_{r_1}, v_{r_2}, \cdots, v_{r_n}\}$ が与えられたとき，本手法は V_r から考えられる全ての順列 $\{\mathbf{v}^k | 1 \leq k \leq n!\}$ の中から，式 (2) の $S(\mathbf{v}^k)$ が最大となる順列 $\mathbf{v}^k$ を求めている．ここで，$S(\mathbf{v}^\mathbf{k})$ は順列 $\mathbf{v}^\mathbf{k}$ の読みやすさを示すスコアを表す．順列 $\mathbf{v}^\mathbf{k}$ はノード列 $v_1^k v_2^k \cdots v_n^k$ であり，v_i^k は順列 $\mathbf{v}^k$ における i 番目のノードを意味す

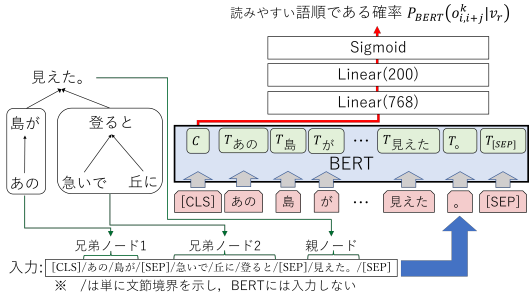


図 2 BERT に基づく $P_{BERT}(o^k_{i,i+j}|v_r)$ の推定モデル

る．このとき，$S(\mathbf{v^k})$ は以下のように算出される．

$$S(\mathbf{v^k}) = \prod_{i=1}^{n-1} \prod_{j=1}^{n-i} P_{BERT}(o^k_{i,i+j}|v_r) \quad (2)$$

ここで $o^k_{i,i+j}$ は，ノード v^k_i が v^k_{i+j} より文頭側に現れる前後関係を意味する．$P_{BERT}(o^k_{i,i+j}|v_r)$ は，ノード v^k_i および v^k_{i+j} が親ノード v_r に係るとき，語順 $o^k_{i,i+j}$ が適切である確率を BERT に基づくモデルによって推定した値である．

図 2 に本手法における BERT モデルの概要を示す．BERT への入力は，兄弟である葉ノード v^k_i と v^k_{i+j} とそれらの親ノードとの 3 者を結合し，先頭に [CLS] を，それぞれの後に [SEP] を付与したうえで，サブワード分割を施したものとする．本手法では，BERT の出力のうち [CLS] に対応する出力のみを取り出し，2 層の Linear 層と Sigmoid を介し，入力が読みやすい語順である確率を出力する．

4 評価実験

本手法の有効性を示すために，新聞記事文を用いた語順整序実験を実施した．なお，本研究では新聞記事文は読みやすい語順であるとみなす．

4.1 実験概要

実験には，新聞記事文に形態素情報及び文節境界情報，構文情報が人手で付与された京都大学テキストコーパス Ver. 4.0[13] を用いた．1 月 1 日から 8 日までと 1 月 10 日から 6 月 9 日までの 25,388 文を学習データとし[2]，1 月 9 日と 6 月 10 日から 6 月 30 日までの 2,368 文から，1,050 文を開発データ，1,164 文をテストデータとした[3]．

評価指標は二文節単位一致率（2 つずつ文節を取り上げ，それらの前後関係が元の文と一致しているものの割合[4]）[2] と文単位一致率（元の文の語順と完全に一致している文の割合）[14] を採用した．

比較のため，以下の 4 つの手法を用意した．

[BERT⁻]: 本手法においてボトムアップ処理の代わりに非ボトムアップ処理を用いる手法．

[RNNLM⁻+SVM⁻]: 高須ら [3] による手法．非ボトムアップ処理を採用．式 (1) の $\alpha = 0.15$.

[RNNLM+SVM]: [RNNLM⁻+SVM⁻] においてボトムアップ処理を採用した手法．式 (1) の $\alpha = 0.19$.

[SVM⁻]: [RNNLM⁻+SVM⁻] において SVM を単体で用いる手法．式 (1) の $\alpha = 0$．内元らの手法 [2] の ME を SVM に差し替えた再実装手法とみなす．

モデルは PyTroch[5] を用いて実装した．BERT の事前学習モデルには，京都大学の公開モデル（BASE WWM 版）[6] を用いた．Linear 層 2 層の次元数はそれぞれ 768 と 200 とし，それぞれの入力を 0.1 の確率でドロップアウトさせた．学習アルゴリズムは AdamW を用い，パラメータの更新はミニバッチ学習 (学習率 1e-6，バッチサイズ 16) により行った．損失関数には BCELoss を使用した．同一パラメータで 5 つのモデルを作成し，それらの各一致率の平均を評価値とした．エポック数は，開発データにおいて文単位一致率が最良だった 6 とした．

表 1 実験結果

	二文節単位一致率	文単位一致率
[BERT]	92.01%	71.53%
[BERT⁻]	89.48%	65.36%
[RNNLM+SVM]	85.56%	58.68%
[RNNLM⁻+SVM⁻]	85.84%	58.59%
[SVM⁻]	85.49%	57.47%

4.2 実験結果

表 1 に実験結果を示す．本手法 [BERT] は，両一致率において他の各手法を有意に上回っており ($p < 0.01$)，本手法の有効性を確認した．

図 3 に，本手法 [BERT] と 2 番目に高精度だった [BERT⁻] の各手法において，文の長さごとに文単位一致率を集計したグフラを示す．ほとんどの文長において [BERT] が [BERT⁻] を上回った．[BERT] は，ボトムアップに処理することにより，各文節の子孫の文節の情報を適切に考慮できるため，特に長い文

2) 同じ親を持つ複数の子ノード集合から 2 つずつノードを取り出したときに作られる前後関係のうち，正解の語順と同じ語順を正例，異なる語順を負例とし，学習データを作成した．
3) 残りの 154 文は 3.1 節の手順 2 を単純に繰り返すことで構文情報のみから 1 文全体の語順が確定するため，除外した．
4) 冗長な計測を避けるため，それぞれの祖先文節の語順によって語順が定まるような文節の組は除外した．
5) https://pytorch.org/
6) https://nlp.ist.i.kyoto-u.ac.jp/?ku_bert_japanese/

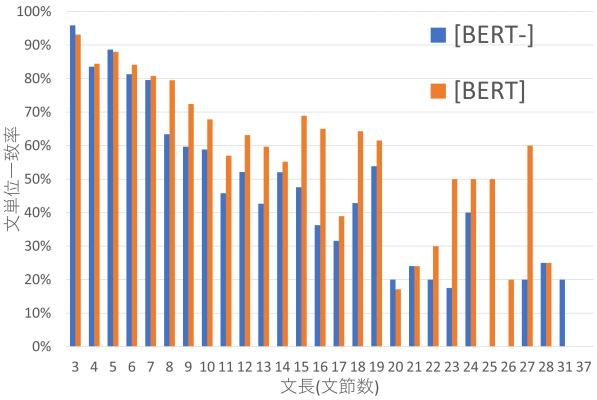


図 3　文長ごとの文単位一致率

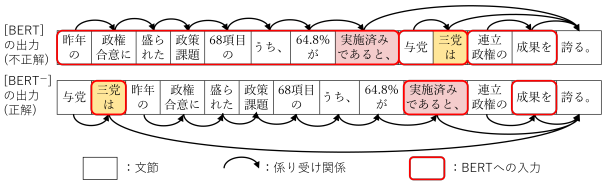


図 4　[BERT⁻] が正解し [BERT] が不正解だった例

に対して，非ボトムアップ処理を行う [BERT⁻] よりも高精度に語順整序できたと考えられる．

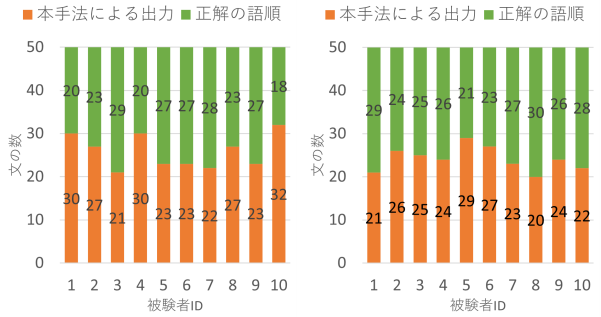
次に，[BERT⁻] が正解し，[BERT] が不正解だった例を図 4 に示す．この例では，「誇る。」の 3 つの子ノードを語順整序する際に，本手法が最長の係り句「昨年の … 実施済みであると」を文頭に配置し不正解となった．本手法は，長い係り句が文頭に表れやすいという傾向を反映し並べているが，正解は最長の係り句が文頭に表れない比較的珍しい語順であったため誤ったと考えられる．一方，1 文単独で提示されただけでは，図 4 の 2 文は同程度に読みやすいとも考えられ，文脈無の評価では，必ずしも正解だけが唯一の正しい語順とは言えないことが分かる．

5　語順整序結果に対する主観的評価

本節では，被験者実験を行い，本手法による語順整序結果の読みやすさを正解文と比較しつつ，主観的に評価した．実験では，文脈の有無による読みやすさの変化も確認するため、1 文を単独で提示する場合と、文脈を含めて提示する場合の 2 パターンで評価した．いずれの被験者も同一の 10 名である．

5.1　文単位での主観的評価

1 文を単独で提示した際の読みやすさを評価する実験では、[BERT] による出力文とその正解文の 2 文を一組として同時に提示し，被験者が読みやすいと感じた方を選択した．各被験者はランダムに抽出



(a) 文単位の場合　　(b) 文脈有の場合

図 5　主観的評価の結果

した同一の 50 組 100 文に対して評価を行った．なお実際には，ダミーをランダムに加えた計 125 組を各被験者は評価している．また，1 組中の 2 文は語順のみが異なる．

その結果を図 5a に示す．本手法を選択した割合は，最も多い人で 64.0%，最も少ない人で 42.0%だった．過半数の文に対して，正解を選択した人と，本手法を選択した人は共に 5 人おり，本手法は正解文と同程度に読みやすい語順を生成できていると考えられる．

5.2　文脈を含む場合での主観的評価

文脈を含む場合の被験者実験では，5.1 節と同一の 50 組 100 文を対象に，その各組の 2 文に加えて，文脈としてその直前の 3 文も同時に提示し、被験者が読みやすいと感じた方を選択した．なお，対象文が属する記事の中で最大 3 文とし，例えば対象文が記事の冒頭にある場合，文脈は提示されないものとした．

その結果を図 5b に示す．本手法を選択した割合は，最も多い人で 58.0%，最も少ない人で 40.0%だった．過半数の文に対して，正解を選択した人が 6 人，本手法を選択した人が 3 人だった．文脈有で評価する場合，文単位でのみ学習した本手法の結果は，文脈を含めて推敲された新聞記事文と比べて読みやすいとは言えないことが分かった．

6　おわりに

本稿では，文を構成する文節集合とその係り受け木に対して，ボトムアップに BERT を適用することにより，1 文全体の適切な語順を同定する手法を提案した．実験の結果，BERT 及びボトムアップ処理の有効性を確認した．また主観的評価の結果，語順整序において文脈を考慮する必要性を確認した．

謝辞

本研究は，一部，科学研究費補助金基盤研究 (C) No. 19K12127 により実施した．

参考文献

[1] 日本語記述文法研究会. 現代日本語文法 7. くろしお出版, 2009.

[2] 内元清貴, 村田真樹, 馬青, 関根聡, 井佐原均. コーパスからの語順の学習. 自然言語処理, Vol. 7, No. 4, pp. 163–180, 2000.

[3] 高須恵, 大野誠寛, 松原茂樹. RNNLM と SVM を用いた日本語文の語順整序. 情報処理学会第 82 回全国大会講演論文集, 第 2020 巻, pp. 453–454, 2020.

[4] Tatsuki Kuribayashi, Takumi Ito, Jun Suzuki, and Kentaro Inui. Language models as an alternative evaluator of word order hypotheses: A case study in Japanese. In **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics**, pp. 488–504, 2020.

[5] 横林博, 菅沼明, 谷口倫一郎. 係り受けの複雑さの指標に基づく文の書き換え候補の生成と推敲支援への応用. 情報処理学会論文誌, Vol. 45, No. 5, pp. 1451–1459, 2004.

[6] Katja Filippova and Michael Strube. Generating constituent order in German clauses. In **Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics**, pp. 320–327, 2007.

[7] Karin Harbusch, Gerard Kempen, Camiel van Breugel, and Ulrich Koch. A generation-oriented workbench for performance grammar: Capturing linear order variability in German and Dutch. In **Proceedings of the 4th International Natural Language Generation Conference**, pp. 9–11, 2006.

[8] Geert-Jan M. Kruijff, Ivana Kruijff-Korbayovà, John Bateman, and Elke Teich. Linear order as higher-level decision: Information structure in strategic and tactical generation. In **Proceedings of the 8th European Workshop on Natural Language Generation**, pp. 74–83, 2001.

[9] Eric Ringger, Michael Gamon, Robert C. Moore, David Rojas, Martine Smets, and Simon Corston-Oliver. Linguistically informed statistical models of constituent structure for ordering in sentence realization. In **Proceedings of the 20th International Conference on Computational Linguistics**, pp. 673–679, 2004.

[10] James Shaw and Vasileios Hatzivassiloglou. Ordering among premodifiers. In **Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics**, pp. 135–143, 1999.

[11] Allen Schmaltz, Alexander M. Rush, and Stuart Shieber. Word ordering without syntax. In **Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing**, pp. 2319–2324, 2016.

[12] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)**, pp. 4171–4186, 2019.

[13] Daisuke Kawahara, Sadao Kurohashi, and Kôiti Hasida. Construction of a Japanese relevance-tagged corpus. In **Proceedings of the 3rd International Conference on Language Resources and Evaluation**, pp. 2008–2013, May 2002.

[14] 山添壮登, 大野誠寛, 松原茂樹. 言語モデルと構文情報を用いた日本語文のボトムアップ語順整序. 情報科学技術フォーラム講演論文集, 第 20 巻, pp. 295–296, 2021.