

文章執筆支援のための項省略の可否判断予測モデル

久保田智天^{1,2} 松林優一郎^{1,2} 乾健太郎^{3,1,2}

¹ 東北大学 ² 理化学研究所 ³ MBZUAI

kubota.tomotaka.s4@dc.tohoku.ac.jp y.m@tohoku.ac.jp

kentaro.inui@mbzuai.ac.ae

概要

本研究では、文章執筆における「**省略された主語や目的語(項)を表出すべきか**」の可否判断の支援を目的に、読み手の項省略可否判断を予測するモデルを構築する。項省略の可否判断予測タスクの設計とデータセットの構築を行い、追加学習を実施したLLMの予測性能を評価する。実験の結果、モデルは一定の判断能力を持つ一方で、適切な位置への指摘や正確な格の使い分けに関しては改善の余地があることが明らかになった。

1 はじめに

日本語では、しばしば項の省略が発生する。項は述語に関連した主語や目的語を表す名詞句を指し、文中における述語との関係を格で表現する。

- 太郎は友人から紅茶をもらった。家に帰ってから家族と**紅茶**を飲んだ。
- それには、こんなきみの悪いことが書いてあったのです。今夜、**ほうせき**をもらいに行く。

文1の項「紅茶を」は、省略しても文脈から容易に補完できるため、省略が可能である。一方、文2の項「ほうせきを」は、省略すると読み手は文脈から項を補完できず文意の解釈が困難になるため、省略できない。このように、話者は読み手に文意が伝わるよう適切に項を省略する必要があり、母語話者は省略の可否を経験的に体得している。省略可否の要因分析や、省略可否を判断する計算モデルの構築によって、文脈に応じた判断要因の説明による省略判断の体系的学習や、文章執筆時における適切な省略判断の支援が期待される。

そこで、本研究では、読み手に意図が伝わりやすい文章を執筆するための、適切な項省略の可否判断支援を見据えたモデルの構築を目指す。特に、与えられた文章中で、補完が困難なために読み手が表出

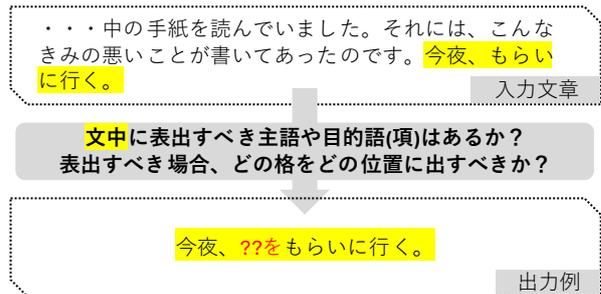


図1 省略可否判断の予測タスクの概観

を必要とする、文意の伝達を妨げる不適切な項省略を予測し、項の表出を指摘するモデルを構築する。

項省略判断を対象とした関連研究として、Ishizukiら[1]は、日本語母語話者の判断を収集し、要因分析と言語モデルを用いた判断のモデリングを行っている。Ishizukiらの設定では、話者が意図する本来の項を明示し、その項が省略可能かを判断する。

これに対し、本研究で想定する執筆支援の場面では、不適切な項省略がすでに生じており、省略された本来の項をモデルは知り得ない状況で、省略後の文のみを手掛かりに、読み手にとって不適切な項省略を検出する必要がある。また、指摘のためには、項省略の可否だけでなく「どの格を、どの位置に表出すべきか」を含めた判断の予測が必要となる。

これを踏まえ、本研究では、(1)文中で省略された項を表出すべきかどうかを判断し、表出すべき格と位置を指摘する省略可否判断予測タスク(図1)を設計し、(2)述語項構造が付与されたコーパスからデータセットを構築し、(3)LLMに対する追加学習と評価を行う。データセットの構築では、人手アノテーションの代替にサプライザル[2,3]を用いて、表出の必要度が高い項が省略された事例を作成する。

実験の結果、モデルは、項省略の可否判断について、ベースラインよりも高い予測性能を示した。一方で、適切な位置への指摘や正確な格の使い分けでは課題があることが示された。

2 タスク設定

項省略の可否判断予測タスクの概観を図 1 に示す。モデルは、与えられた文章の最終文中において、文脈上省略されている可能性のある項を表出すべきかどうかを、一般的な日本語母語話者の読み手の立場から予測し、その結果に応じて最終文を編集して出力する。予測対象は、「が・を・に・は・も」の表層格で表される項である。

入力として、先行 k 文と予測対象文からなる文章を与える。本研究では先行文脈を $k = 3$ 以内に定める。先行文脈の設定は、省略項の先行詞の 80%以上が先行 3 文以内に現れる知見に基づく日本語省略解析(ゼロ照応解析)の設定 [4, 5, 6] に従った。予測対象文は、文意の解釈に必要な項がどれか一つ省略¹⁾された文、あるいは、項が十分に表出された文を与える。また、読み手が文章を読解する実場面では、省略された項の情報は明示されない事を考慮し、省略された項の格と省略位置の情報は与えない。

項を表出する必要があると予測した場合は、「どの位置に何の格を表出すべきか」を明示するマーカーを予測対象文に一つだけ挿入して出力する。マーカーは??+ 格(が・を・に・は・も)で構成し、必要な項の表層格に応じて格を選択する。例えば、図 1 中の「今夜、もらいに行く。」について表層格「を」で表す項の表出を指摘する場合は、「今夜、??をもらいに行く。」とマーカーを挿入して出力する。これは、対話場面で聞き手が話者に対し「何をもらいに行くの?」と聞き返す行為を再現した設定である。

項の表出が不要であると予測した場合は、マーカーを挿入せずにそのまま繰り返し出力する。

3 データセット

BCCWJ(現代日本語書き言葉均衡コーパス) [7] に対して述語項構造・依存構造情報が付与された BCCWJ-DepParaPAS [8] の PB(書籍)・PM(雑誌)ドメイン 163 文書の各文から事例を作成する。

ここで、人手アノテーションを必要とせずに学習事例を構築するために、熟達した執筆者は読み手に応じて適切に省略可否を判断しており、文意伝達に必要な項を文中で表出しているという仮定を置く。

1) 実際には、表出が必要な項が複数省略されている状況も考えられる。本研究では、単純な設定下における項の省略可否判断の予測能力評価を目的とし、表出が必要な項が1つだけ省略された文に限定する。文意の解釈上で表出が必要な複数の項が同時に省略された文への拡張は今後の課題とする。

3.1 事例集合の作成

項の表出を指摘する (Insertion) / 指摘しない (No Insertion) 各事例を作成する。詳細な作成手順は附録 A に示す。

Insertion 事例 文中に表出された項を省略することで、文意の解釈に必要な項が不足した事例を作成する。本研究では、省略の結果、項共有が発生し自然に補完できる事例の混入を避けるため、文中で一度だけ出現する表層格で表される項を省略対象とする。入力文は文中で表出された項を一つ省略して作成し、正答文は省略位置にマーカーを挿入して作成する。省略対象の項が複数存在する場合は、各項をそれぞれ一つずつ省略した事例を作成する。

No Insertion 事例 熟達した執筆者が執筆した作成元の文は情報が十分に記述されているため、項の表出を必要ではないとみなし、作成元の文をそのまま入力・正答文に利用して作成する。事例は、文中に表出していない表層格の個数分作成する。

3.2 Insertion 事例のフィルタリング

3.1 節で作成した Insertion 事例には、読み手側から見た場合に項を表出する必要性を認識できない事例が含まれる。例えば、以下の例

- 僕は、瞳子さんが何を言いたそうと、百パーセント彼女の意思に従うつもりでいた。・・・(中略)・・・彼女はそれだけのことを ϕ_{\perp} してくれたのだから。

は、Insertion 事例の入力文の一つであり、 ϕ_{\perp} には「僕に」が省略されている。この項は、文脈から容易に補完できるため読み手からは表出の必要性は低い。このように、話者が表出した項でも、読み手にとっては表出の必要性が低い場合がある。表出の必要性が低い項の表出を予測させる事例は、訓練・評価時にノイズとなる。そこで、Insertion 事例集合に対して、読み手にとって表出の必要性が低い項を省略した事例のフィルタリングが必要となる。

本研究では、項の表出必要度をサプライザル [2, 3] を用いて定式化し、Insertion 事例集合の各事例について、LLM の予測確率を用いて算出したスコアが閾値より低い事例を除外して、表出の必要性が低い事例のフィルタリングを試みる。

スコアの定式化 項の表出必要度を、項の挿入による文の読解処理負荷の改善度合いで定式化す

る。サプライザル理論 [2, 3] では、単語の逐次的な処理負荷を文脈 C に対する単語 w の負の対数尤度 $-\log P(w|C)$ で定義する。これを用いて、省略された項の後続から文末までの系列について、項の表出による処理負荷の改善度合いを求める。

トークン系列 (t_1, \dots, t_n) とその文脈 C について、系列中の各トークンのサプライザル s_i を算出し、その総和を系列全体のサプライザル s とする。

$$s_i = -\log P(t_i|C, t_{<i}) \quad (1)$$

$$s = \sum_{i=1}^n s_i \quad (2)$$

省略された項の直後の位置から文末まで対象に系列サプライザル s を算出し、項の表出が後続の処理負荷をどの程度改善するかを評価する。項省略文の系列サプライザル s_{omitted} と同一文脈・同一内容で当該項のみを表出した項表出文の系列サプライザル s_{inserted} の差を項の表出の必要度 **necessity** とする。

$$\text{necessity} = s_{\text{omitted}} - s_{\text{inserted}} \quad (3)$$

necessity スコアが大きいほど、項の表出により後続の処理負荷が低くなるため、読み手にとって表出の必要性が高いと考えられる。一方で、スコアが小さいほど、項の表出の必要性は低くなり、負の数を取るときは、冗長性が増して表出文の処理負荷が相対的に高くなったと考えられる。このスコアが特定の閾値を下回る事例を、読み手にとって表出の必要度が低いものと判定し、事例集合から除外する。

閾値の設定 項の表出の必要性を読み手の側から見て検出できるかどうかを事前に人手でアノテーションしたデータを用意し、表出の必要性を検出できない事例を取りこぼしなく除外できた閾値を採用する。Insertion 事例作成時に、データセットに用いる事例と別に 40 件の事例を用意し、第一著者がアノテーションを実施した。全 40 件中、表出の必要性を認識できない事例は 7 件であった。

閾値を 2.00 から 6.50 まで 0.25 ずつ増やし探索した結果、閾値 5.25 で表出が不要な事例を最も多く除外 (7 件中 6 件) できた。ただし、除外した 13 件中 7 件は表出が必要な事例であり、過剰な除外も同時に見られた。本研究では、再現率を重視して最も取りこぼしの少ない閾値を設定した。表出が不要な事例のみを取り除ける項の必要度の定式化や閾値の探索は今後の課題とする。

フィルタリング作業 Insertion 事例集合の各事例の入力文 (項省略文) と作成元の文 (項表出文) から

表 1 Soft-case 判定における格の対応関係

| 予測/正答の格 | 正解として許容する格 |
|---------|------------|
| が | が・も・は |
| を | を・も・は |
| に | に |
| は | が・も・を・は |
| も | が・も・を・は |

算出した必要度スコアが閾値を下回る事例を除外する。スコアは、先行 3 文を加えた文脈からの各トークンに対する Llama-3.1-Swallow-8B-v0.5[9]²⁾ の予測確率に基づくサプライザルを用いて算出した。トークナイズは文節単位で実施した。

Insertion 事例集合 18848 件から 5523 事例を除外し、残りの 13325 事例をデータセットに用いる。No Insertion 事例集合 75007 件から、Insertion 事例と表層格ごとの事例分布と総事例数が同一になるようにサンプリングし、データセットを作成した。

4 評価手法

正答文とモデルの出力文を比較し、マーカの有無、挿入位置 (position)、および挿入した表層格 (case) について、以下のルールで正誤を判定する。

位置の判定 挿入位置の一致を考慮する (Exact-position) / しない (Free-position³⁾) 判定を設定する。Exact-position では、正答と出力でマーカの前 5 文字以内の文字列が一致する場合を正解とする。不正解の場合は、本来表出すべき項を指摘しておらず、また、別の項を誤って指摘したものと判定する。

格の判定 格の一致を厳密 (Strict-case) / 緩やか (Soft-case) に見る判定を設定する。Strict-case では、モデルの挿入した表層格と、挿入すべき表層格が一致する場合を正解とする。Soft-case では、表 1 の対応関係に基づき、適合率計算時にはモデルが指摘した格に対して許容される正答集合を、再現率計算時には表出すべき格に対して許容される予測集合を用い、それぞれに含まれる場合を正解と判定する。

これらの判定を組み合わせて、以下の判定基準

- Exact-position, Strict-case matching (E-p, Strict-c)
- Exact-position, Soft-case matching (E-p, Soft-c)
- Free-position, Strict-case matching (F-p, Strict-c)

2) <https://huggingface.co/tokyotech-llm/Llama-3.1-Swallow-8B-v0.5>

3) Free-position では、挿入位置は一致しないものの、指摘すべき格が不足している事を理解できているかを評価する。ただし、文中に述語が複数ある場合、どの述語の格に対して指摘しているかは曖昧性が残ることは注意されたい。

表 2 各セットの事例分布 (%)

| 表層格 | 訓練 | | 開発 | | 評価 | |
|-----|-------|--------|------|--------|------|--------|
| | Ins. | NoIns. | Ins. | NoIns. | Ins. | NoIns. |
| が | 13.4 | 13.6 | 12.7 | 14.8 | 13.4 | 12.9 |
| を | 14.3 | 14.4 | 13.3 | 15.7 | 15.4 | 13.3 |
| に | 8.6 | 8.5 | 8.0 | 9.3 | 8.7 | 9.2 |
| は | 10.1 | 9.9 | 9.2 | 8.8 | 11.0 | 9.0 |
| も | 3.7 | 3.5 | 4.5 | 3.7 | 3.2 | 3.9 |
| 事例数 | 21308 | | 2646 | | 2696 | |

• Free-position, Soft-case matching (F-p, Soft-c)

を設定する。各評価事例におけるモデルの予測に対して表層格別に判定した各基準の TP, FP, FN を集約し、Micro F1 値を算出する。モデルが複数のマーカーを挿入した場合は、各マーカーを一つずつ残した文を作成し、それぞれを独立に評価する。

5 実験・議論

5.1 実験設定

Llama-3.1-Swallow-8B-instruct-v0.5[9]⁴⁾ を用いる。LoRA-tuning [10] を実施し (学習設定は付録 B に示す)、各エポック終了時に開発セット上で算出した Exact-position, Strict-case 判定の Micro F1 値が最大となるモデルを採用する。プロンプトにはチャットテンプレートを適用し、system にはタスク形式の指示 (詳細は付録 C に示す) を与え、user には入力文章のみを与える。データは、同一文書から作成した事例がセット間で混合しないように訓練：検証：テスト = 8 : 1 : 1 に分割した。各セットの事例分布⁵⁾ を表 2 に示す。評価は、各判定基準における Micro F1 値で行う。また、Free-position 判定のベースラインには次の方法を用いる、まず、項の表出可否を等確率 (0.5) で選択した上で、項の表出を選択した場合には、訓練データに含まれる Insertion 事例の分布に従って、未表出の格からランダムに格を選択する。

5.2 実験結果

実験結果を表 3 に示す。ベースライン (Random) と Free-position 判定の Micro F1 値を比較すると、全ての case 判定においてベースラインを上回っており、指摘すべき格の不足を一定程度指摘できてい

4) <https://huggingface.co/tokyotech-llm/Llama-3.1-Swallow-8B-Instruct-v0.5>

5) 表層格ごとの分布の算出に際して、No Insertion 事例では、事例作成時に文中に表出していない格のうちの一つを「表出の必要がない格」としてラベル付けがされている。

表 3 実験結果 (p:position, c:case)

| | E-p, Strict-c | E-p, Soft-c | F-p, Strict-c | F-p, Soft-c |
|----------|---------------|-------------|---------------|-------------|
| が | 52.20 | 66.50 | 60.30 | 79.10 |
| を | 73.90 | 76.30 | 81.30 | 86.30 |
| に | 60.30 | 60.30 | 66.30 | 66.30 |
| は | 45.90 | 61.70 | 52.00 | 75.60 |
| も | 27.60 | 61.50 | 29.30 | 73.60 |
| Micro F1 | 56.33 | 69.53 | 63.03 | 80.22 |
| Random | - | - | 18.55 | 45.75 |

ることが分かる。一方で、case 判定が同一条件の場合の position 判定別の Micro F1 値を比較すると、Exact-position でスコアが約 10 ポイントほど低下しており、項を表出すべき位置の特定が難しいことが示唆される。表層格ごとの結果では、Strict-case 判定で「は」と「が」は「に」よりも訓練事例数が多いにも関わらず F1 値が低いことが特徴的である。これらの格は Soft-case 判定で F1 値が Strict-case 判定より 10 ポイント以上高く、「に」のスコアを上回る。指摘すべき格の不足を一定是指摘できるが、正確な表層格の選択は難しいことが示唆される。

以上の結果から、モデルは項省略の可否について一定の判断能力を持つ一方で、文脈に応じた表層格の正確な使い分けや、適切な位置への表出の指摘については改善の余地があることが示唆される。なお、本研究では、執筆者が項を表出した位置を正解の位置として、その位置に表出を指摘できるかを評価している。読み手が表出を必要とする位置に対する指摘性能の評価は今後の課題とする。

6 おわりに

本研究は、「文中の省略された主語や目的語 (項) を表出すべきか」を読み手の視点から予測するモデルを構築した。項省略の可否判断予測タスクの設定とデータセットの構築を行い、追学習した LLM の予測性能を評価した。実験の結果、モデルは指摘すべき格の不足の予測でベースラインより高い性能を示す一方で、適切な位置への指摘や、正確な表層格の使い分けが課題として残る。予測性能向上の手段の一つには、訓練事例の拡張が考えられる。

今後の展望として、既存の省略解析を統合した、与えられた文脈で読み手が補完する項の提案が挙げられる。これにより、読み手は項の補完が可能であるものの、話者の意図と異なる項を補完してしまうような、文意の誤伝達を招く不適切な項省略も指摘できると期待される。

参考文献

- [1] Yukiko Ishizuki, Tatsuki Kuribayashi, Yuichiroh Matsubayashi, Ryohei Sasano, and Kentaro Inui. To drop or not to drop? predicting argument ellipsis judgments: A case study in Japanese. In **Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)**, 2024.
- [2] John Hale. A probabilistic Earley parser as a psycholinguistic model. In **Second Meeting of the North American Chapter of the Association for Computational Linguistics**, 2001.
- [3] Roger Levy. Expectation-based syntactic comprehension. **Cognition**, Vol. 106, No. 3, pp. 1126–1177, 2008.
- [4] Ryohei Sasano and Sadao Kurohashi. A discriminative approach to Japanese zero anaphora resolution with large-scale lexicalized case frames. In **Proceedings of 5th International Joint Conference on Natural Language Processing**, 2011.
- [5] Masatsugu Hangyo, Daisuke Kawahara, and Sadao Kurohashi. Japanese zero reference resolution considering exophora and author/reader mentions. In **Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing**, 2013.
- [6] Souta Yamashiro, Hitoshi Nishikawa, and Takenobu Tokunaga. Neural Japanese zero anaphora resolution using smoothed large-scale case frames with word embedding. In **Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation**, 2018.
- [7] Kikuo Maekawa, Makoto Yamazaki, Toshinobu Ogiso, Takehiko Maruyama, Hideki Ogura, Wakako Kashino, Hanae Koiso, Masaya Yamaguchi, Makiro Tanaka, and Yasuharu Den. Balanced corpus of contemporary written Japanese. **Language resources and evaluation**, Vol. 48, pp. 345–371, 2014.
- [8] 浅原正幸, 大村舞. Bccwj-depparapas: 『現代日本語書き言葉均衡コーパス』係り受け・並列構造と述語項構造・共参照アノテーションの重ね合わせと可視化. 言語処理学会第22回年次大会発表論文集, 2016.
- [9] Kazuki Fujii, Taishi Nakamura, Mengsay Loem, Hiroki Iida, Masanari Ohi, Kakeru Hattori, Hirai Shota, Sakae Mizuki, Rio Yokota, and Naoaki Okazaki. Continual pre-training for cross-lingual LLM adaptation: Enhancing Japanese language capabilities. In **First Conference on Language Modeling**, 2024.
- [10] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, et al. Lora: Low-rank adaptation of large language models. **ICLR**, Vol. 1, No. 2, p. 3, 2022.

A 事例の作成手法

コーパス中の各文に対して、以下の処理を実行する。事例の作成例を表 4 に示す。

はじめに、文中の用言・体言述語に対して直接係り受け関係にあるガ・ヲ・ニ格の項を述語項構造の情報に基づき特定し、各項がどの表層格（格助詞「が・を・に」および係助詞「は・も」）により表現されるかを品詞情報を用いてカウントする。加えて、品詞情報を用いて文中に格助詞「が・を・に」と係助詞「は・も」が出現する回数を独立にカウントし、文全体での助詞の表層的な出現回数を集計する。

Insertion 事例は、表層格の出現回数と助詞の表層的な出現回数の双方において、文全体で表層格が 1 回のみ出現する項を省略して作成する。省略の際には、依存構造情報を用い、省略対象の項に係る修飾語を含むフレーズ全体を削除する。また、会話文で文頭に項が出現する場合には、省略後の文が不自然とならないよう、括弧等の記号を適宜補完する。

No Insertion 事例は、文中において、表層格の出現回数と、助詞の表層的な出現回数の双方において、当該表層格が出現しない場合、その格ごとに事例をずつ 1 件作成する。

表 4 作成例

| |
|------------------------------------|
| Source Text : |
| 知識の量的な「爆発」と同時に、知識の複合化が起 こっています。 |
| Insertion 事例 |
| Input Text : |
| 知識の量的な「爆発」と同時に、起こっています。 |
| Ground Truth : |
| 知識の量的な「爆発」と同時に、??が起こって います。 |
| No Insertion 事例 |
| Input Text : |
| 知識の量的な「爆発」と同時に、知識の複合化が 起こっています。 |
| Ground Truth : |
| 知識の量的な「爆発」と同時に、知識の複合化が 起こっています。 |

B ハイパーパラメータ

学習時のハイパーパラメータを表 5 に示す。

表 5 ハイパーパラメータ

| パラメータ | 値 |
|----------------|--|
| 学習率 | 1e-4 |
| バッチサイズ | 4 |
| 勾配累積ステップ数 | 4 |
| エポック数 | 15 |
| fp16 | False |
| bf16 | True |
| lora_r | 8 |
| lora_alpha | 16 |
| lora_dropout | 0.1 |
| target_modules | q_proj, k_proj, v_proj, o_proj, gate_proj, up_proj, down_proj |

C プロンプト

モデルの訓練・評価時に利用したシステムプロンプトを表 6 に示す。

表 6 システムプロンプト

あなたは日本語の母語話者です。ユーザーが与える文章を読んで、文章の最後の一文について、省略された主語や目的語などの述語項（名詞句+助詞「が／を／に／は／も」）を明示すべきかどうかを判断してください。
出力は必ず 1 文のみとし、理由や説明は一切書かないでください。
明示が必要な場合は、最終文の適切な位置に、明示すべき項の助詞に合わせて「??が」「??を」「??に」「??は」「??も」のいずれかを 1 つだけ挿入して出力してください。
明示が不要な場合は、最終文をそのまま出力してください。