

見出し意味具体化に向けた日本語ベンチマークの構築

白井穂乃¹ 石原祥太郎¹¹ 株式会社日本経済新聞社

{hono.shirai,shotaro.ishihara}@nex.nikkei.com

概要

新聞・雑誌などの記事の見出しには独特の文体があり、音声読み上げや機械翻訳の品質を低下させる懸念がある。本研究では、見出しの内容を補完し一般的な文体に変換する**見出し意味具体化**に焦点を当て、日本語の新聞記事96本を用いて、ルールベースや大規模言語モデルの性能を評価する。最初に、人手によるアノテーションを通じて、助詞の省略や略語に関する見出しと本文の文体の違いを明らかにした。次に、大規模言語モデルの性能を計測したところ、助詞の補完では人手評価で誤りがほとんど確認されないが略語の展開に関しては改善の余地があり、自動評価の検討の必要性も示唆された。

1 はじめに

情報収集の形態が多様化する中で、音声読み上げや機械翻訳など、新聞・雑誌などの記事を対象とした情報技術の応用が普及しつつある。音声読み上げは別の作業をしながら手軽に記事に触れる機会を提供し、機械翻訳は情報のやり取りが国際化する中で需要が高まっている。株式会社日本経済新聞社でも、様々なサービスで記事の音声読み上げ¹⁾や英語・中国語翻訳²⁾の機能を提供している。

記事を対象とした情報処理の難しさの一つに、見出し特有の文体の存在が挙げられる。見出しは短い分量でより多くの情報を伝える目的から、助詞の省略や体言止めといった独特の文体を持つ[1]。これらの文体は視認を前提としているため、文字列を単純に音声で読み上げると違和感に繋がる恐れがある[2, 3]。機械翻訳を適用する際も、一般的な文章には現れない表現や単語の省略が原因で、本文に忠実ではない訳が生成される可能性がある[4, 5]。

本研究では、独特の文体を持つ見出しの内容を

表1 本研究で題材とする記事の見出しと本文の抜粋、および人手による見出し意味具体化の例。「三井不」という略語を「三井不動産」に置換し、本文の内容に合致するように助詞「が」と文末の動詞「を開業する」を補った上で語順を入れ替えている。

見出し	郊外型商業施設、三井不、国内4ヵ所で。
(英訳)	Suburban commercial facilities, Mitsui & Co. in 4 locations in Japan.
本文	三井不動産は28日、2014～15年に大阪府吹田市の遊園地「エキスポランド」跡地など国内4ヵ所で郊外型商業施設を開業すると発表した。…
意味具体化	三井不動産が郊外型商業施設を開業する、国内4ヵ所で。
(英訳)	Mitsui Fudosan opens suburban commercial facilities in four locations in Japan.
理由	略語、助詞、文末、語順

補完し一般的な文体に変換する**見出し意味具体化**[6]を題材とする。日本語の新聞記事データセット『日本経済新聞記事オープンコーパス』[7]から抜粋した記事³⁾を用いて、見出し意味具体化の具体例を表1に示す。見出し「郊外型商業施設、三井不、国内4ヵ所で。」では名詞が列挙されているが、本文と照合し助詞「が」や文末の動詞「を開業する」を補った上で語順を入れ替えると、分量が増える分だけ適切に文意を伝えられる可能性がある。意味具体化前後の見出しを英訳すると、後者の方がより本文に即した内容を出力できた⁴⁾。略語の「三井不」を「三井不動産」に置換することで、企業名も適切に英訳されている。

本研究では(1)見出し意味具体化で必要な処理は何か(2)ルールベースと比較した大規模言語モデルの性能はどの程度か—について調査する。音声読み上げや機械翻訳の需要が高まり、データセット公開や大規模言語モデル[8]の台頭が進む一方で、見出し意味具体化は十分に研究されていない(2節)。我々は日本語の新聞記事にアノテーションを実施

1) <https://www.nikkei.com/article/DGXZQOFM211WK0R21C21A000000/>

2) https://www.nikkei.com/promotion/onboarding/guide/guide_05/

3) 紙幅による文字数の制約が少ない「日経電子版」では異なる見出しが付与されている場合もある(Appendix A)。

4) 英訳には「DeepL」を2023年12月25日に用いた。<https://www.deepl.com/translator>

表 2 日本経済新聞記事オープンコーパスの 96 記事へのアノテーション. 主要な理由として, 助詞を補完する「助詞」, 略語を展開する「略語」, 単語を並び替える「語順」, 末尾の動詞などを補完する「文末」が存在した.

理由	見出し	意味具体化	件数
助詞	日経平均 292 円上昇、1 年 10 ヶ月ぶり高値、円一時 88 円台。	日経平均が 292 円上昇、1 年 10 ヶ月ぶりの高値、円一時 88 円台。	48
略語	国会論戦、8 日の衆院予算委から——維新東国原英夫氏。	国会論戦、8 日の衆院予算委員会から——維新東国原英夫氏。	30
語順	デイリーヤマザキ吸収、山崎パン、仕入れ・物流を一本化。	山崎製パンがデイリーヤマザキを吸収、仕入れ・物流を一本化。	5
文末	郊外型商業施設、三井不、国内 4 ヶ所で。	三井不動産が郊外型商業施設を開業する、国内 4 ヶ所で。	2

し, 特に助詞の省略と略語の観点で見出しと本文に文体の違いがあると明らかにした (3 節). 大規模言語モデルの性能を計測したところ, 助詞の補完では人手評価で誤りがほとんど確認されず, 略語の展開に関しては改善の余地があった (4 節).

2 関連研究

見出し意味具体化は, 見出しを入力として本文に近い文章を出力する点で, 見出し生成の逆の操作と言える. 見出し生成は伝統的に数多くの研究対象となってきた [9, 10]. 著名なベンチマーク [11, 12, 13, 14] も存在し, 自然言語処理の発展に伴い飛躍的な性能向上が確認されている. 一方で見出し意味具体化に関連する研究は多くない. 最も類似する 2015 年の先行研究 [6] は, 見出しと本文を照合し動詞や格を追加・置換するルールベース処理を適用した. その他の研究 [1, 2, 3, 4, 15, 16, 17, 18, 19] も文末表現を中心に変換規則を記述する手法を検討している. これらは適用する見出し意味具体化の処理を定性的に検討している場合が多く, データセットへのアノテーションや統計情報から導出している研究は多くない. 我々の研究は, 公開データセットを用いて見出し意味具体化に必要な処理を明らかにする点や, ルールベースと比較した大規模言語モデルによる性能を評価する点で新規性がある.

3 見出しと本文の文体の違い

本節では, 日本経済新聞記事オープンコーパスを用いて見出しと本文の文体の違いを分析することで, 見出し意味具体化に必要な処理を明らかにする. このコーパスでは「日本経済新聞」の朝夕刊 (2013 年 1~2 月) に掲載された 96 記事の見出しと本文に対し, 国立国語研究所が形態論情報と文節係り受け情報を人手でアノテーションした. 本研究で題材とする見出し意味具体化に関する情報は含まれ

ていないため, 著者 2 人によるアノテーションを実施した. アノテーションの妥当性を調べるため, 定量分析も実施した.

3.1 人手によるアノテーション

著者 2 人が日本経済新聞記事オープンコーパスの 96 記事に対し, それぞれ見出し意味具体化のアノテーションを実施し, その後に合議制で最終的なアノテーション結果を作成した. 理由の具体例を表 2 に示す. 理由として最も多かったのは, 省略されている助詞を補完する「助詞」で 48 記事が該当した. 次いで略語を補う「略語」が 30 記事, 列挙されている主述の単語を並び替える「語順」が 5 記事, 末尾の動詞などを補完する「文末」が 2 記事存在した. 35 記事は, 見出し意味具体化が不要と判断した. 助詞や略語の具体例は Appendix B に示す.

3.2 定量分析

表 3 見出し・見出し意味具体化した見出し・本文に対する助詞・動詞・名詞の割合の平均値.

指標	見出し	意味具体化	本文
助詞の割合	9.56	13.52	22.32
動詞の割合	1.86	1.98	8.33
名詞の割合	55.87	53.93	42.13

人手によるアノテーションの妥当性を調べるため, 見出し・見出し意味具体化した見出し・本文のそれぞれについて, 助詞・動詞・名詞の割合の平均値を算出した (表 3). 品詞推定には MeCab [20] で ipadic 辞書を用いた. まず, 全ての指標で, 見出しと本文で大きく割合が異なっている. 助詞の割合は本文では 22.32% あるが, 見出しでは 9.56% に低下する. 動詞の割合も本文では 8.33% あるが, 見出しでは 1.86% となる. 一方で名詞の割合は本文では 42.13% だが, 見出しでは 55.87% に上昇する. これらは, 助詞・文末を理由とする見出し意味具体化が必要なことと繋がる. 見出し意味具体化した結果に

対する指標は、それぞれの割合が本文の値に寄っており、アノテーションの妥当性が確認された。

4 見出し意味具体化の実験

本節では、大規模言語モデルの性能を計測した実験結果を報告する。比較として、ルールベースと、与えられた見出しを全く変更しない設定も用いた。

4.1 ルールベースによる変換

表 2 で挙げた 4 処理のうち、Imono ら [6] を参考に略語の展開、文末の補完、助詞の補完の 3 つを実装する。品詞推定や構文・格・照応解析には Juman++ [21] と KNP [22]、類似度計算には 2010~2021 年の日経電子版の記事で学習した word2vec [23] を用いた。語順の変更については本研究の対象外とした。

- 略語の展開: 見出しの固有名詞のうち、本文に類似度が 0.7 以上の異なる固有名詞があれば展開する。名詞の特定には、固有名詞の収録が多い辞書を持つ Sudachi [24] を用いた。
- 文末の補完: 見出しの文末の名詞と合致する本文の名詞を探す。名詞の係り先が動詞だった場合、その動詞で補完する。
- 助詞の補完: 見出しの「名詞+動詞」の係り受け関係と合致する本文の「名詞+助詞+動詞」の組み合わせがあれば、該当する助詞を補完する。

4.2 大規模言語モデルによる変換

日本語を扱える大規模言語モデルとして、次の 2 モデルを用いた。

- LLM-jp (13B)⁵⁾: 国立情報学研究所が公開したパラメタ数 130 億のモデル。日本語約 1450 億トークンを含む約 3000 億トークンのコーパスで事前学習されている。
- ChatGPT (gpt-35-turbo)⁶⁾: OpenAI が開発した多言語対応のモデル。API 経由で生成結果が得られる。バージョンは 2023-07-01-preview。

人手でアノテーションした 96 記事から、評価セットとして 50 記事を取り出し、残りの 46 記事のうち見出し意味具体化が必要と判断された 29 記事を Few-shot Learning [25] の学習データの候補とす

5) <https://huggingface.co/llm-jp/llm-jp-13b-instruct-full-jaster-dolly-oasst-v1.0>

6) <https://openai.com/blog/chatgpt>

る (Appendix C)。具体的には、候補からランダムに 0~10 記事を選び、事例としてプロンプトに追加した。見出し意味具体化のために本文の情報が必要な場合もあるため、本文を加えたプロンプトも用意した。本文を加える際は、入力長の制限のためそれぞれ冒頭 200 文字のみを利用した。最終的にはモデルの種類 (2 通り)、本文の有無 (2 通り)、事例数 (11 通り) の計 44 通りの生成結果を得た。生成時の具体的な設定は、Appendix D に示す。

本文を含めたプロンプト

ニュース記事の見出しを、一般的な文章に書き換えてください。できる限り元の見出しに忠実にして変更しないでください。いくつかの例を示すので参考にしてください。必要に応じて本文の情報も参照してください。

見出し: { 見出し 1 }

本文: { 本文 1 }

書き換え後: { 意味具体化 1 }

.....

見出し:

本文:

書き換え後:

4.3 評価指標：編集距離と人手評価

評価指標として、正規化された編集距離の平均値を用いた。具体的には評価セットの 50 記事に対して、生成結果と人手によるアノテーションの正規化された編集距離を RapidFuzz ライブラリ⁷⁾で計算し、その平均値を算出する。値が大きいほど、生成結果が人手によるアノテーションと一致している。自動の評価指標を用いる上で、人手評価との相関を確認するのが望ましい [26]。そこで「変更なし」とルールベース、および事例数が最大 (10 記事) の大規模言語モデルの生成結果に対しては、著者による人手評価を実施した。人手評価では、必ずしも人手によるアノテーションと一致していない場合も、日本語として許容される場合には正解と見なした。

4.4 結果

実験結果を表 4 に示す。事例数が 0 から 1 になることで、それぞれのモデルで大きく性能が改善している。一方で、その後の事例数の増加と性能は明確には相関しなかった。LLM-jp は ChatGPT に比べて

7) <https://github.com/rapidfuzz/RapidFuzz>

表 4 見出し意味具体化の実験結果. 評価指標は正規化された編集距離の平均値. 大きいほど生成が人手によるアノテーションと一致していることを示す.

モデル	本文	事例数										
		0	1	2	3	4	5	6	7	8	9	10
変更なし	-	94.19	-	-	-	-	-	-	-	-	-	-
ルールベース	-	92.35	-	-	-	-	-	-	-	-	-	-
LLM-jp (13B)		82.38	87.75	88.88	90.50	88.64	88.08	86.89	85.05	86.21	84.72	87.87
LLM-jp (13B)	✓	62.06	84.76	86.95	81.65	85.72	77.64	78.30	82.48	82.15	71.13	77.94
ChatGPT (gpt-35-turbo)		55.99	73.96	71.79	73.15	81.12	79.09	78.39	80.71	80.72	83.01	83.51
ChatGPT (gpt-35-turbo)	✓	38.06	66.31	66.14	77.19	68.74	71.80	76.30	75.81	70.02	78.04	76.84

表 5 モデルの生成例.

モデル	本文	事例数	生成
見出し	-	-	日航、ビジネス客好調——「787」運航停止が影、長期化なら路線展開制約。
ChatGPT (gpt-35-turbo)		0	「787」の運航停止が影響し、日航は路線展開に制約が生じる可能性があるが、ビジネス客は好調である。
ChatGPT (gpt-35-turbo)		10	日本航空、ビジネス客が好調——「787」運航停止の影響で長期化なら路線展開に制約が生じる。
見出し	-	-	仏、現金決済の上限下げ、課税逃れ取引防止、税収増・公平性を確保。
本文	-	-	…フランスはこれまで現金決済を認めていた上限を引き下げ、今後は1000ユーロ（約12万5000円）以上の現金決済は認めない方針を決定。…
ルールベース	-	-	フランス、現金決済の上限下げ、課税逃れ取引防止、税収増・公平性を確保。
LLM-jp (13B)		10	仏、現金決済の上限を下げ、課税逃れ取引防止、税収増・公平性を確保。
LLM-jp (13B)	✓	10	フランス、現金決済の上限を下げ、課税逃れ取引を防止、税収増や公平性を確保。

表 6 人手評価による正解数.

モデル	本文	全体	助詞	略語
変更なし	-	18/50	0/20	0/18
ルールベース	✓	36/50	17/20	5/18
LLM-jp (13B)		31/50	19/20	2/18
LLM-jp (13B)	✓	28/50	17/20	5/18
ChatGPT (gpt-35-turbo)		32/50	19/20	2/18
ChatGPT (gpt-35-turbo)	✓	38/50	19/20	10/18

総じて良好な結果を示したが、比較対象の「変更なし」の 94.19 やルールベースの 92.35 は上回らない性能となった。プロンプトに本文を含めた場合、性能が劣化している。表 5 の例では、事例数を 0 から 10 にすることで人手によるアノテーションに従って編集箇所が限定的になり（上）、プロンプトに本文を含めることで「仏」が「フランス」に変換された（下）。表 6 に示した人手評価による正解数は、プロンプトに本文を含めた ChatGPT が 38 と最も多かった。特に略語に関する正答数が大幅に改善している。LLM-jp でもプロンプトに本文を含めると略語に関する正答数が 3 増えたが、本文の情報に基づく生成誤りが発生したため全体の正答数は悪化した。助詞の補完に関してはほとんど誤りがなかつ

たが、略語の展開については改善の余地があると分かった。編集距離と人手評価は十分に相関しておらず、評価方法の再考の必要性が示唆された。

5 結論と今後の展望

本研究では、見出し意味具体化に焦点を当て、日本語の新聞記事 96 本を用いてルールベースや大規模言語モデルの性能を評価した。助詞の補完では大規模言語モデルが十分に機能したが、略語の展開に関しては比較的苦手であると分かった。プロンプトの試行錯誤は今後の展望の一つである。得られた知見を活かして、略語に関しては外部データベースを参照するなど全体として高性能なシステムを構築できる可能性がある。実験結果を踏まえたアノテーションの精緻化も検討している。今回の実験では生成結果を人手評価したが、迅速な試行錯誤を含むベンチマークとしての役割を鑑みると、自動評価の枠組みの構築が望ましい。人手評価と相関のある自動の評価指標を模索する他、高性能な大規模言語モデルの活用や、人手評価の結果を用いた回帰モデルの構築などが選択肢として挙げられる。

参考文献

- [1] 山本和英, 池田諭史, 大橋一輝. 「新幹線要約」のための文末の整形. 自然言語処理, Vol. 12, No. 6, pp. 85–111, 2005.
- [2] Yukiko Hayashi and Shigeki Matsubara. Sentence-style conversion of japanese news article for text-to-speech application. In **Proceedings of 7th International Symposium on Natural Language Processing**, pp. 257–262, 2007.
- [3] 松原茂樹, 村田匡輝. 自然なニュース音声の自動生成に向けた体言止めの文への文末表現の補完. 電子情報通信学会論文誌 D, Vol. 97, No. 1, pp. 181–185, 2014.
- [4] 吉見毅彦, 佐田いち子. 英字新聞記事見出し翻訳の自動前編集による改良. 自然言語処理, Vol. 7, No. 2, pp. 27–43, 2000.
- [5] 李正政. ウェブニュース見出しの文末表現における日中翻訳. 通訳翻訳研究, Vol. 16, pp. 21–41, 2016.
- [6] Misako Imono, Eriko Yoshimura, Seiji Tsuchiya, et al. **Method of Embodying the Newspaper Headlines by Using Words and Phrases in the Article**, pp. 51–63. 2015.
- [7] 浅原正幸, 高松純子, 若狭絢ほか. 日本経済新聞記事オープンコーパス: 新聞記事コーパスと形態・統語情報アノテーション. 言語処理学会第 29 回年次大会併設ワークショップ「日本語言語資源の構築と利用性の向上」, 2023.
- [8] Wayne Xin Zhao, Kun Zhou, Junyi Li, et al. A survey of large language models. **arXiv preprint arXiv:2303.18223**, 2023.
- [9] Tanya Goyal, Junyi Jessy Li, and Greg Durrett. News summarization and evaluation in the era of GPT-3. **arXiv preprint arXiv:2209.12356**, 2022.
- [10] Pengshan Cai, Kaiqiang Song, Sangwoo Cho, et al. Generating user-engaging news headlines. In **Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 3265–3280, 2023.
- [11] Xiang Ao, Xiting Wang, Ling Luo, et al. PENS: A dataset and generic framework for personalized news headline generation. In **Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)**, pp. 82–92, 2021.
- [12] Max Grusky, Mor Naaman, and Yoav Artzi. Newsroom: A dataset of 1.3 million summaries with diverse extractive strategies. In **Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)**, pp. 708–719, 2018.
- [13] Shashi Narayan, Shay B. Cohen, and Mirella Lapata. Don’t give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization. In **Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing**, pp. 1797–1807, 2018.
- [14] Tahmid Hasan, Abhik Bhattacharjee, Md. Saiful Islam, et al. XL-sum: Large-scale multilingual abstractive summarization for 44 languages. In **Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021**, pp. 4693–4703, 2021.
- [15] 岩越守孝, 増田英孝, 中川裕志. Web と携帯端末向けの新聞記事の対応コーパスからの文末言い換え抽出. 自然言語処理, Vol. 12, No. 5, pp. 157–183, 2005.
- [16] Hiroshi Nakagawa and Hidetaka Masuda. Extracting paraphrases of japanese action word of sentence ending part from web and mobile news articles. In **Proceedings of the 2004 International Conference on Asian Information Retrieval Technology**, AIRS’04, p. 94–105, 2004.
- [17] Hiroshi Nakagawa, Hidetaka Masuda, and Dai Sato. Terminal device oriented comparable corpora and its alignment- towards extracting paraphrasing patterns. In **Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC’04)**, 2004.
- [18] 清水嶺, 酒井哲也. 新聞記事からのラジオ読み上げ原稿の自動生成. 第 13 回データ工学と情報マネジメントに関するフォーラム, 2021.
- [19] 林由紀子, 松原茂樹. 自然な読み上げ音声出力のための書き言葉から話し言葉へのテキスト変換. 情報処理学会研究報告, 2007.
- [20] Takumitsu Kudo. Mecab: Yet another part-of-speech and morphological analyzer. 2005.
- [21] Arseny Tolmachev, Daisuke Kawahara, and Sadao Kurohashi. Juman++: A morphological analysis toolkit for scriptio continua. In **Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations**, pp. 54–59, 2018.
- [22] Daisuke Kawahara and Sadao Kurohashi. A fully-lexicalized probabilistic model for Japanese syntactic and case structure analysis. In **Proceedings of the Human Language Technology Conference of the NAACL, Main Conference**, pp. 176–183, 2006.
- [23] Tomáš Mikolov, Kai Chen, Greg Corrado, et al. Efficient estimation of word representations in vector space. In **Workshop Track Proceedings of 1st International Conference on Learning Representations**, Scottsdale, Arizona, USA, 2013.
- [24] Kazuma Takaoka, Sorami Hisamoto, Noriko Kawahara, and Others. Sudachi: a Japanese tokenizer for business. In **Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)**, 2018.
- [25] Tom Brown, Benjamin Mann, Nick Ryder, et al. Language models are few-shot learners. In **Advances in Neural Information Processing Systems**, Vol. 33, pp. 1877–1901, 2020.
- [26] Satyanjeev Banerjee and Alon Lavie. METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In **Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization**, pp. 65–72, 2005.

A 編集者による見出しの編集

紙幅による文字数の制約が少ない「日経電子版」では、編集者が見出し意味具体化に相当する処理を実施している場合がある。例えば、表 1 の記事の日経電子版での見出しは「三井不動産が 4 商業施設相次ぎ開業 郊外型、14～15 年に」だった⁸⁾。しかし、全ての記事に対する見出し意味具体化は編集者の労働負荷に繋がるため、可能な限り自動で処理できることが望ましい。編集者による見出しの編集を用いたアノテーションの構築は、今後の展望の一つである。

B アノテーションでの助詞・略語

3 節での日本経済新聞記事オープンコーパスへの人手によるアノテーションでの助詞・略語の具体例を示す。補完した助詞の中では「の」が 25 記事で最も多く、次いで「を」が 18 記事、「が」が 15 記事、「で」と「は」が 3 記事、「と」が 1 記事だった。登場した略語の例を表 7 に示す。「伊 (イタリア)」「英 (イギリス)」「米 (アメリカ)」などの国名、「三井不 (三井不動産)」「メトロ (東京メトロ)」「GE (ゼネラル・エレクトリック)」「横河電 (横河電機)」「T & G (テイクアンドギヴ・ニーズ)」「山崎パン (山崎製パン)」などの企業名といった例が存在した。その他、「T P P (環太平洋経済連携協定)」「E E Z (排他的経済水域内)」といった略語も含まれている。「明大 (明治大学)」は、音声読み上げの応用を想定すると「名大 (名古屋大学)」との曖昧性を回避する上で必要となる。

C 実験でのデータセット分割

実験では、日本経済新聞記事オープンコーパスの 96 記事を article_id を軸に昇順に並び替えた後、末尾の 50 記事を評価セットとして用いた。人手によるアノテーションでは、うち 18 記事で意味具体化が不要と判断された。残りの 32 記事のうち 20 記事で助詞、18 記事で略語に関するアノテーションを含んだ。

表 7 人手によるアノテーションで登場した略語の例。

見出しでの略語	意味具体化
W杯	ワールドカップ
放医研	放射線医学総合研究所
T P P	環太平洋経済連携協定
原発	原子力発電所
自民	自民党
尖閣	尖閣諸島
伊	イタリア
10年債	新発10年物国債利回り
メトロ	東京メトロ
国交相	国土交通大臣
GE	ゼネラル・エレクトリック
火力合弁	火力発電設備の合弁会社
三井不	三井不動産
英	イギリス
ウイングドフット	ウイングドフットGC
シン	ビジェイ・シン
EEZ	排他的経済水域内
Sトラック	ショートトラック
女子1000	女子1000メートル
日航	日本航空
「787」	ボーイング787型機
リーマン	リーマン・ショック
横河電	横河電機
衆院予算委	衆院予算委員会
米	アメリカ
T P P	環太平洋経済連携協定
T & G	テイクアンドギヴ・ニーズ
山崎パン	山崎製パン
参院予算委	参院予算委員会
民主	民主党
P B	プライベートブランド
明大	明治大学

D 実験の生成時の設定

実験での大規模言語モデルの生成時の設定を示す。LLM-jp では、Hugging Face のサイト⁹⁾で提示されたハイパーパラメータとして max_new_tokens に 100, top-p に 0.95, temperature に 0.7 を設定した。Transformers ライブラリ¹⁰⁾のバージョンは 4.36.2 とした。ChatGPT では、デフォルトの設定を用いた。

8) <https://www.nikkei.com/article/DGXNASDD280HO- Y3A120C1TJ1000/>

9) <https://huggingface.co/llm-jp/llm-jp-13b-instruct-full-jaster-dolly-oasst-v1.0>

10) <https://github.com/huggingface/transformers>