

# 修辞構造に基づき言語モデルを制御する テキスト生成手法

横川悠香<sup>1</sup> 石垣達也<sup>2</sup> 宮尾祐介<sup>2,3</sup> 高村大也<sup>2</sup> 小林一郎<sup>1,2</sup>  
<sup>1</sup>お茶の水女子大学 <sup>2</sup>産業技術総合研究所 <sup>3</sup>東京大学  
 {g1820542,koba}@is.ocha.ac.jp yusuke@is.s.u-tokyo.ac.jp  
 {ishigaki.tatsuya, takamura.hiroya}@aist.go.jp

## 概要

本研究では、言語モデルによるテキスト生成に、修辞構造に基づく制御を加える手法の提案を行う。分類器を用いて言語モデルを制御する手法を適用し、言語モデルに修辞構造を取り入れた既存の手法と比較する。評価の対象として幅広い対象者を持つテキストを用い、精度に加えて文を接続する単語の観点から評価を行い、精度向上を確認した。

## 1 はじめに

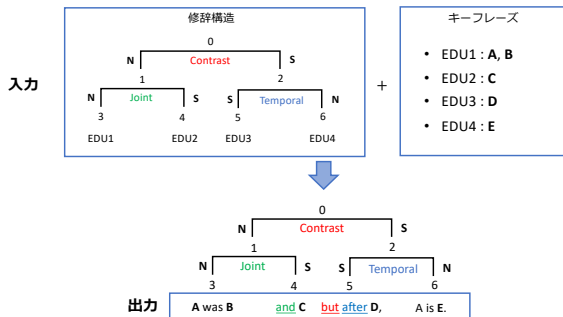


図1 モデルの入力となる修辞構造とキーフレーズの例

修辞構造は、テキスト内の一部分(文や節)の間において論理的な接続関係を考えることを再帰的にを行い、テキスト全体にわたる構造を木構造形式で表現するものである。テキスト生成に活用することで、生成文の一貫性の向上が期待できる [1]。一方で、近年のテキスト生成の性能を高めた言語モデルは単語を左から順に生成する自己回帰型で生成を行うため、テキスト全体の構造を捉える修辞構造のような木構造を生成文に取り入れるには工夫が必要となる。

言語モデルに対して修辞構造を取り入れた手法として、修辞構造全体を埋込みの形に変換し、言語モデルの入力に追加する手法 ([2]) が提案されている

が、生成時において、入力として与えた修辞構造全体と生成しているテキストが持つ構造が対応しているかどうかの確認は行われていない。

これに対し、本研究では、分類器を用いて言語モデルを制御する手法 [3] の適用により、テキスト内に表現される潜在的な関係性を識別しながら生成と同時に制御を行う手法を提案する。これにより、言語モデルの性能を活用することに加え、修辞構造が生成に反映されているかを確認しながら生成することができる。精度に加えて文を接続する単語の観点から評価し、その向上を確かめた。

## 2 背景

### 2.1 修辞構造

修辞構造理論 (Rhetorical Structure Theory;RST) [4] では、意味上の最小単位である Elementary Discourse Unit:EDU (日本語では節に相当) で区切り、全体の構造を EDU を葉とする木の形で表現する。木の中における関係性は、関係ラベルと Nuclearity とで表現される。関係ラベルは原因と結果のような2つのノードの意味的な関係性を表す。Nuclearity は Nucleus と Satellite の2つの値をとり、Nucleus であるノードは Satellite であるノードより意味的に重要であることを表す。

### 2.2 関連研究

RST においては、修辞構造はテキスト自体が実現される前に存在するものと考えるが、木構造に含まれるテキストの単位が、実際にテキストにどのような順番で現れるのかについての原則は示されていない [5]。修辞構造をテキスト生成に利用した例として、木を探索した順番で各ノードのテキストを生成する手法 [1] が提案されているが、テキストに現れ

順番と生成の順番が異なるため、言語モデルによる自己回帰型の生成にそのまま取り入れるのは困難である。

言語モデルによる生成に修辞構造を取り入れた手法として RSTGen [2] が提案されている。RSTGenでは、修辞構造を、木の中で左右両方の子を持つノードに対応する RST の情報を並べた列として表現する。修辞構造を埋め込むための追加の層を用いて、列の表現から埋込み表現へと変換したものを言語モデルの入力となるトークンの埋込みに追加することでテキスト生成に修辞構造を反映させている。

RSTGen が修辞構造全体を埋め込んでモデルに与えるのに対して、本研究では、分類器を用い、生成されるテキスト内の EDU が持つ関係性が、指定した修辞構造に従っているかの識別を生成と同時に行う手法を通し、修辞構造に基づくテキスト生成の実現を目指す。

### 3 提案手法

#### 3.1 概要

**入出力の定義** RSTGen に従い、入力は以下の 2 つとする。

##### 1. 修辞構造

本研究では、考慮する関係の数を制限するため、木の深さは 2 以下に制限する。

Nuclearity は既知のものとし、ノード 0, 1, 2 に対応する関係ラベルを並べた列  $rel = (rel_0, rel_1, rel_2)$  として表す。

##### 2. 重要なフレーズ (キーフレーズ)

RSTGen に従い、キーフレーズとその位置を与える。本来キーフレーズの位置は学習の対象とすべきであるが、本研究では修辞構造に基づく生成に限って手法の提案を行うために、既知のものとした。

正解文のトークン列から、キーフレーズと、EDU の区切りを示す特殊なトークンの位置以外を全てマスクに置換したトークン列  $\hat{x}$  として表す。

出力はトークン列  $x = (x_0, \dots, x_T)$  であり、修辞構造を反映し、キーフレーズを使用したテキストである。

修辞構造に基づくテキスト生成を行うために、まず、EDU 単位で与えられたキーフレーズを元にテ

キストを生成するモデルを導入する。続けて、EDU 間の関係ラベルを識別する分類器を用いてテキスト生成を行うモデルを制御する手法を提案する (3.2 節)。そして、3.2 節で提案した手法をトークン列の各トークンに対して逐次的に適用することで、修辞構造に基づいたテキスト生成を行う。修辞構造の階層性を考慮し、2 つの階層についてそれぞれ生成を行う (3.3.1 節)。得られた 2 通りの出力に対して、トークン列全体が指定した関係ラベル列を持つ確率を分類器により求め、比較することで最終的な出力を得る (3.3.2 節)。

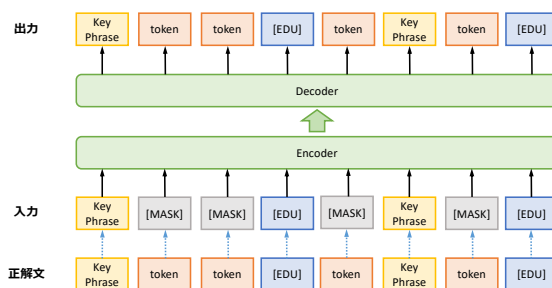


図 2 キーフレーズを元に生成するモデル

**生成の基盤として用いるモデル** 生成の基盤として、EDU 単位で与えられたキーフレーズを元にテキストを生成するよう学習したモデルを用いる。図 2 に示すように、キーフレーズの情報  $\hat{x}$  はエンコードされ、トークン列  $x$  を自己回帰的に生成する。このモデルは以下の分布を表現する言語モデルと言える。

$$p(x_{t+1} | x_{:t}, \hat{x}) \quad (1)$$

#### 3.2 分類器を用いたモデルの制御

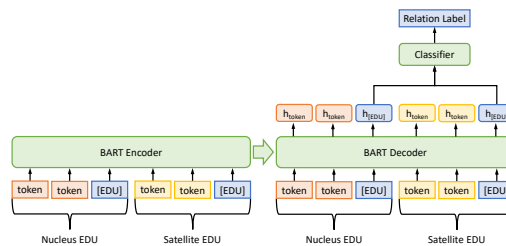


図 3 関係ラベルを識別する分類器

**関係ラベルを識別する分類器** 図 3 に示す関係ラベルを識別する分類器を導入する。入力は EDU の組であり、出力は関係ラベルである。分類器は以下の分布をモデリングする。

$$p(rel | x_{t+1}, x_{:t}, \hat{x}) \quad (2)$$

分類器の実験は付録 A で示す。

本研究では、「ある EDU に対し、生成対象の EDU が指定した関係ラベルを持つように生成対象の EDU に含まれるトークンを出力する」ことを以下のように表現する。

$$x_{t+1} \sim p(x_{t+1}|rel, x_t, \hat{x}) \quad (3)$$

本研究では、これを Plug and Play Language Models (PPLM) [3] に基づいて行う。PPLM では、属性  $a$  を持つテキスト  $X$  の生成、すなわち分布  $p(X|a)$  のモデリングを、バイズの定理  $p(X|a) \propto p(X)p(a|X)$  に基づき言語モデルの分布  $p(X)$  に分布  $p(a|X)$  を掛け合わせることで行う。分布  $p(a|X)$  は単層の分類器を用いて定義される。詳細は付録 B に示すが、PPLM では、まず言語モデルを用いてトークンを生成し、続けて分布  $p(a|X)$  の下で属性  $a$  の尤度が高くなるように言語モデルの内部状態を更新し、更新した内部状態を用いて新しくトークンを生成し直すことで制御を行う。

本研究では、以下のように基盤となるモデルの分布 (式 (1)) に対し、関係ラベルについての分類器を用いて定義した分布 (式 (2)) を掛け合わせることで、目的となる式 (3) 右辺の分布をモデリングする。

$$p(x_{t+1}|rel, x_t, \hat{x}) \propto p(rel|x_{t+1}, x_t, \hat{x})p(x_{t+1}|x_t, \hat{x}) \quad (4)$$

### 3.3 修辞構造に基づく生成の制御

例として、図 1 の修辞構造における EDU1 と EDU2 が生成された状態から、EDU3 に含まれるトークン  $x_{t+1}$  を出力することを考える。

まず、基盤となるモデルの分布 (式 (1)) から、制御を加える前の出力として  $x_{t+1}$  を得る。

$$x_{t+1} \sim p(x_{t+1}|x_t, \hat{x}) \quad (5)$$

#### 3.3.1 階層性を考慮したトークンの出力

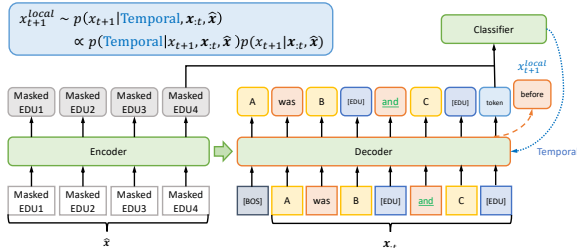


図 4 低い階層における関係を考慮したトークンの出力

続いて、 $x_{t+1}$  が関係ラベルに基づくように制御を行う。

3.1 節において修辞構造の木の深さを 2 以下に制限したことにより、ある EDU について、生成時に考慮する関係は、低い階層における関係と高い階層における関係の 2 通りが考えられる。

例として、図 1 の EDU3 については、低い階層における関係は EDU4 に対する「Temporal (時間に関する)」となり、高い階層における関係は EDU1 と EDU2 の組に対する「Contrast (対比)」となる。

図 4 に示すように、低い階層における関係として、EDU4 に対する「Temporal」の関係を考える。式 (4) における関係ラベル  $rel$  を「Temporal」、分類器の入力は EDU3 と EDU4 の組とし、以下に示す分布から出力  $x_{t+1}^{local}$  を得る。

$$x_{t+1}^{local} \sim p(x_{t+1}^{local}|Temporal, x_t, \hat{x}) \propto p(Temporal|x_{t+1}, x_t, \hat{x}) \cdot p(x_{t+1}^{local}|x_t, \hat{x}) \quad (6)$$

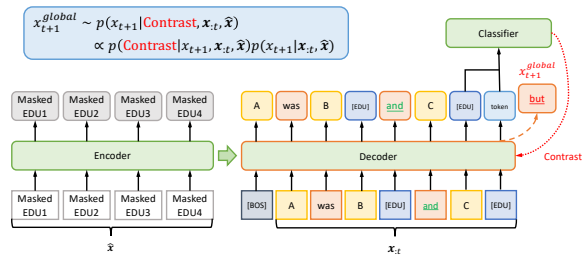


図 5 高い階層における関係を考慮したトークンの出力

同様に、図 5 に示すように、高い階層における関係である EDU1-2 に対する「Contrast」の関係に基づき出力  $x_{t+1}^{global}$  を得る。

#### 3.3.2 全体の関係を考慮したトークンの選択

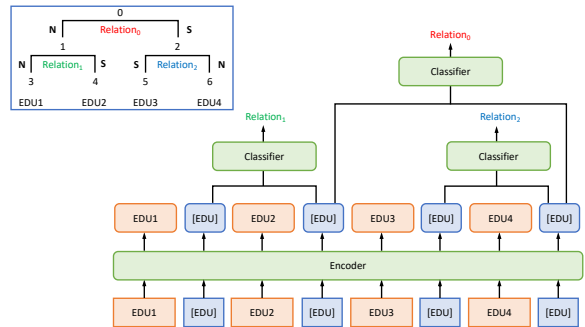


図 6 分類器の拡張

ここで、図 3 で示した分類器を、2 つの EDU の組同士に対しても適用することにより拡張する。入力はトークン列であり、出力は関係ラベル列  $rel$  である。拡張した分類器は以下の分布をモデリング

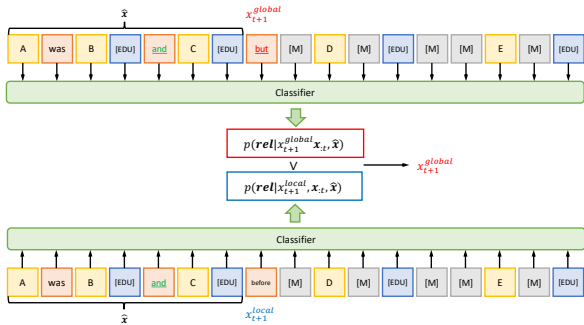


図7 拡張した分類器を用いたトークンの選択

する。

$$p(\text{rel}|x_{t+1}, x_t, \hat{x}) \quad (7)$$

時刻  $t$  までの出力列  $x_{:t}$  と、3.3.1 節で得られた  $x_{t+1}^{local}$  と  $x_{t+1}^{global}$  それぞれを  $\hat{x}$  に挿入する。2通りのトークン列に対し、拡張した分類器を用いて式(7)に示す関係ラベル列  $\text{rel}$  を持つ確率分布を求める。確率が高い方のトークンを  $x_{t+1}$  として選択する。

## 4 実験設定

**データセット** 幅広い対象者にとっての適切なテキストの生成のため、テキスト対象者の英語読解力レベルの異なるニュース記事で構成される Newsela コーパス [6] を用いる。テキストから段落を抜き出し、学習済 RST パーザ [7] によってパーズングし、学習済キーフレーズ抽出器 TopicRank [8] を用いてキーフレーズを抽出する。データセットの大きさは学習/検証/評価で 25,173/3,108/3,131 である。

表1 実験設定

学習済モデル	facebook/bart-base
学習エポック数	20
勾配法	AdamW
バッチサイズ	学習:10, 検証:5, 評価:4
損失関数	負の対数尤度
学習率	$5 \times 10^{-5}$

表1に示す設定で提案モデルの実験を行い、RSTGen に従う修辭構造の埋込みを入力に追加したモデルと比較を行う。

## 5 実験結果

表2 実験結果 (修辭構造埋込みは [2] に従う)

モデル	B-4	R-L	MTR	BERT スコア
BART	58.10	73.03	74.05	95.32
+ 修辭構造埋込み	54.42	72.44	74.77	95.41
+ PPLM 制御	59.47	73.93	74.70	95.85

精度の評価指標として BLEU [9], ROUGE [10], METEOR [11] を用い、意味的な精度の指標として

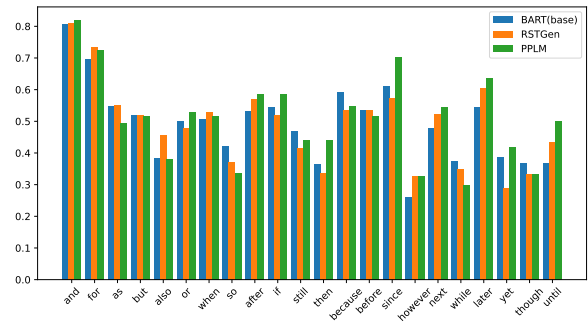


図8 Discourse Marker 別の再現率

BERTScore [12] を用いた実験結果を表2に示す。提案手法のモデルでは BLEU, ROUGE, BERTScore において改善が見られた。修辭構造を取り入れたモデルは共に BERTScore が向上していることから、修辭構造の情報はテキストの意味的な精度向上に寄与するとも考えられる。

また、修辭構造に基づく評価指標として、Discourse Marker (DM) についての再現率を用いる。DM とは、テキストを接続する役割を持つ単語であり、修辭構造がテキスト表面に現れたものと考えることができる。再現率は、正解文において使用された DM の数に対して生成結果に正しく使用された DM の数の比率として計算する。評価に使用した DM を C に示す。図8に示した DM 別の再現率では、全体的に改善傾向が見られた。「after」や「since」といった DM で再現率の向上が大きかったが、「and」や「for」では違いが小さかった。この原因として、前者のような DM は関係ラベルとの結びつきが大きいが、後者のような DM は汎用的で結びつきが小さく、関係ラベルに基づく制御の効果が小さくなったと考えられる。

## 6 まとめ

本研究では、修辭構造に基づきテキスト生成を制御する手法の提案を行った。扱う修辭構造の木の深さに制限はあるものの、対象者のレベルが幅広いテキストにおける精度向上を確認した。また、テキストの意味的な精度の向上から、修辭構造の情報がテキストの意味的な品質の向上に繋がる可能性を考察した。課題として、より階層が深い修辭構造に対する拡張が挙げられる。また、DM という修辭構造がテキストの表面に現れた場合を評価しているため、潜在的なテキストの構造についての評価も検討していきたい。

## 謝辞

この成果は、国立研究開発法人新エネルギー・産業技術総合開発機構 (NEDO) の助成事業 (JPNP20006) の結果得られたものである。

## 参考文献

- [1] Eduard H. Hovy. Planning coherent multisentential text. In **26th Annual Meeting of the Association for Computational Linguistics**, pp. 163–169, Buffalo, New York, USA, June 1988. Association for Computational Linguistics.
- [2] Rilwan Adewoyin, Ritabrata Dutta, and Yulan He. RST-Gen: Imbuing fine-grained interpretable control into long-FormText generators. In **Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**, pp. 1822–1835, Seattle, United States, July 2022. Association for Computational Linguistics.
- [3] Sumanth Dathathri, Andrea Madotto, Janice Lan, Jane Hung, Eric Frank, Piero Molino, Jason Yosinski, and Rosanne Liu. Plug and play language models: A simple approach to controlled text generation. In **International Conference on Learning Representations**, 2019.
- [4] William C. Mann and Sandra A. Thompson. Rhetorical structure theory: Toward a functional theory of text organization. **Text & Talk**, Vol. 8, pp. 243 – 281, 1988.
- [5] Manfred Stede. Rhetorical structure and thematic structure in text generation. **Proceedings of LORID**, Vol. 99, pp. 44–50, 1999.
- [6] Wei Xu, Chris Callison-Burch, and Courtney Napoles. Problems in current text simplification research: New data can help. **Transactions of the Association for Computational Linguistics**, Vol. 3, pp. 283–297, 2015.
- [7] Naoki Kobayashi, Tsutomu Hirao, Hidetaka Kamigaito, Manabu Okumura, and Masaaki Nagata. A simple and strong baseline for end-to-end neural RST-style discourse parsing. In **Findings of the Association for Computational Linguistics: EMNLP 2022**, pp. 6725–6737, Abu Dhabi, United Arab Emirates, December 2022. Association for Computational Linguistics.
- [8] Adrien Bougouin, Florian Boudin, and Béatrice Daille. TopicRank: Graph-based topic ranking for keyphrase extraction. In **Proceedings of the Sixth International Joint Conference on Natural Language Processing**, pp. 543–551, Nagoya, Japan, October 2013. Asian Federation of Natural Language Processing.
- [9] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In **Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics**, pp. 311–318, Philadelphia, Pennsylvania, USA, July 2002. Association for Computational Linguistics.
- [10] Chin-Yew Lin and Eduard Hovy. Automatic evaluation of summaries using n-gram co-occurrence statistics. In **Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics**, pp. 150–157, 2003.
- [11] Satanjeev 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, Ann Arbor, Michigan, June 2005. Association for Computational Linguistics.
- [12] Felix Wu Kilian Q. Weinberger Yoav Artzi Tianyi Zhang, Varsha Kishore. Bertscore: Evaluating text generation with bert. In **International Conference on Learning Representations**, 2020.
- [13] Lynn Carlson and Daniel Marcu. Discourse tagging reference manual. **ISI Technical Report ISI-TR-545**, Vol. 54, No. 2001, p. 56, 2001.
- [14] Daniel Marcu Mary Ellen Okurowski Carlson, Lynn. Rst discourse treebank. Linguistic Data Consortium, 2002.
- [15] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics**, pp. 7871–7880, Online, July 2020. Association for Computational Linguistics.
- [16] 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, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.
- [17] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. Transformers: State-of-the-art natural language processing. In **Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations**, pp. 38–45, Online, October 2020. Association for Computational Linguistics.
- [18] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, L ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, **Advances in Neural Information Processing Systems**, Vol. 30. Curran Associates, Inc., 2017.

## A 生成に用いた分類器

### A.1 概要

入力は Nucleus となる EDU と Satellite となる EDU の組であり，出力は関係ラベルである。

1. Elaboration	7. Background	13. Comparison
2. Attribution	8. Cause	14. Manner-Means
3. Joint	9. Enablement	15. Topic-Change
4. Same-unit	10. Evaluation	16. Summary
5. Contrast	11. Temporal	17. Textual-organization
6. Explanation	12. Condition	18. Topic-Comment

図 9 使用した関係ラベル ([13] より作成)

関係ラベルは図 9 に示す Discourse Tagging Manual [13] における 18 種類のものを用いる。

### A.2 実験

**データセット** ニュース記事に RST に基づくアノテーションをつけたデータセットである RST-DT データセット [14] から，Nucleus と Satellite となる EDU の組を抜き出し，データセットとして用いる。データセットの大きさは学習/検証/評価で 64,098/8,025/8,021 である。

表 3 実験設定

学習済モデル	facebook/bart-base
学習エポック数	20
勾配法	AdamW
バッチサイズ	学習:10, 検証:5, 評価:4
損失関数	負の対数尤度
学習率	$5 \times 10^{-5}$

**実験設定** 表 3 に実験設定を示す。言語モデルには BART [15] を用い，比較のために BERT [16] を用いた設定でも実験を行う。

表 4 実験結果

モデル	正解率	F1 スコア
BERT	55.17	37.59
BART	54.53	37.70

**実験結果** 表 4 の実験結果より，BART を用いた分類器は正解率では BERT に劣るものの，F1 スコアにおいては上回った。

## B PPLM による生成の制御

Transformer の効率的な実装 [17] において，出力トークン列  $x_t$  に基づいて時刻  $t+1$  のトークン  $x_{t+1}$  を出力するとき， $x_t$  と言語モデルの内部状態  $H_t$  を

入力とする。

$$o_{t+1}, H_{t+1} = \text{LM}(x_t, H_t) \quad (8)$$

$$x_{t+1} \sim p_{t+1} = \text{Softmax}(W o_{t+1}) \quad (9)$$

ここで，内部状態とは，具体的には Transformer [18] のアテンション計算における Key-Value を保存した行列である。

PPLM では，分布  $p(a|X)$  からの勾配を用い，属性  $a$  を反映するように言語モデルの内部状態を更新する。

$$\Delta H_t \leftarrow \Delta H_t + \alpha \frac{\nabla_{\Delta H_t} \log p(a|H_t + \Delta H_t)}{\|\nabla_{\Delta H_t} \log p(a|H_t + \Delta H_t)\|^\gamma} \quad (10)$$

更新された内部状態  $\tilde{H}_t = H_t + \Delta H_t$  を用いて，言語モデルにより現在のトークン  $x_t$  が与えられたときの  $\tilde{x}_{t+1}$  を求める。

$$\tilde{o}_{t+1}, H_{t+1} = \text{LM}(x_t, \tilde{H}_t) \quad (11)$$

$$\tilde{x}_{t+1} \sim \tilde{p}_{t+1} = \text{Softmax}(W \tilde{o}_{t+1}) \quad (12)$$

## C 評価に用いた Discourse Marker

1. and	6. when	11. then	16. next
2. for	7. so	12. because	17. while
3. as	8. after	13. before	18. yet
4. but	9. if	14. since	19. though
5. or	10. still	15. however	20. until

図 10 評価に用いた Discourse Marker の一覧

本研究の評価において，Discourse Marker は PDTB Annotation Manual<sup>1)</sup> Appendix A に示されたもののうち，図 10 に示す正解文において 30 回以上使用された Discourse Marker を用いた。

1) <https://www.cis.upenn.edu/elenimi/pdtb-manual.pdf>