修辞構造と語彙難易度を制御可能なテキスト生成手法に向けて

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

概要

言語は，それが用いられる社会的な状況に応じてその形態が変化する．例えば，同じ内容を説明する際であっても，幼い子供を対象にした発話では，理解が容易な文章構造の下，使用される語彙の難易度が低くなるといった変化が生じる．本研究では，特に発話相手の読解能力レベルに応じた，テキストの修辞構造と語彙の難易度の変化に着目し，それらを考慮した制御可能なテキスト生成手法を開発することを目的とする．テキスト生成の実験においては，レベルに応じた生成を行うために，対象者のレベルが異なるニュース記事で構成されるコーパスである Newsela コーパスを用い，異なるレベルに対しても読解が容易なテキストの生成を行なった．

1 はじめに

発話相手の読解能力レベルによって，適切なテキストは異なる [1]．同じ内容を伝える際であっても，発話相手のレベルに応じ，テキスト全体の構造や使用する語彙の難易度を変化させる必要が生じる．本研究では，レベルに応じた変化のうち，テキストの構造を捉えるものである修辞構造 [2] と，使用される語彙の難易度を，テキスト生成において制御する手法を提案する．レベルが異なるニュース記事で構成されるコーパス [3] を用い，記事の内容を端的に表す文であるニュース記事の表題と，記事本文から抜き出した重要なフレーズであるキーフレーズを入力として，表題に基づいた内容を持ち，キーフレーズを使用したテキストの生成に取り組む．その生成に，レベルに応じた修辞構造と語彙の難易度に基づく制御を加える．テキスト生成において，修辞構造をモデルの訓練において補助として用い，テキスト全体の一貫性を高める手法 [4] があるが，本研究では修辞構造に基づいた制御を生成時にも追加する．また，テキスト平易化における語彙の難易度に着目した手法 [5] を訓練と生成時に取り入れる．生成時には，言語モデルの生成にトピックや感情を導入する Controlled Text Generation における手法 [6] を，修辞構造と語彙の難易度に基づく制御に適用する．

2 関連研究

Controlled Text Generation において，Plug and Play Language Models (PPLM) [6] は，言語モデルに追加の訓練を必要としないことが特徴である．また，テキスト生成における制御のうち，テキスト全体が内容についての一貫性を持つように生成する Content Planning の手法として PAIR [7] が挙げられる．一方で，テキスト全体の構造を捉えるものとして，修辞構造がある．修辞構造を学習において利用することで，テキスト全体が自然な構造を持つようにする手法として FlowNet [4] がある．また，テキスト平易化 [8] において生成されるテキストが，時に難易度の高い単語を含むことが知られており，語彙の難易度を考慮して平易化を行う手法 [5] が提案されている．

本研究においては，Newsela コーパス [3] をデータセットとして用い，テキスト生成において修辞構造と語彙の難易度の制御を行う．生成の基盤として PAIR を用い，そこに PPLM の手法を用いた制御を取り入れる．

3 提案手法

図 1 に提案手法の概要を示す．データセットとして，主にテキスト平易化において広く使用される Newsela コーパス [3] を用いる．Newsela コーパスは，ニュースの元記事と，それを人手により平易な英語に書き換えた 4 つのバージョンで構成される．コーパス内の記事には，英語文章の難易度から読者の読解能力を測るために使用される指標である

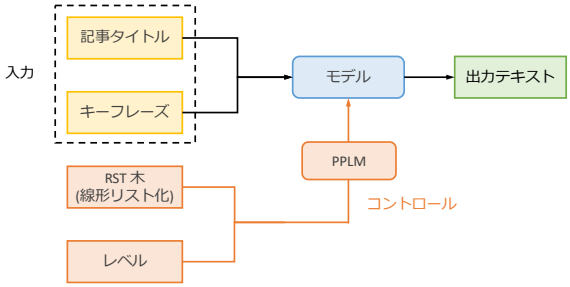

図 1　提案手法

Lexile readability score [1] に基づく 2 から 12 のレベルが付与されている．本研究では，問題を簡単にするため Newsela コーパスに含まれるニュース記事内の段落の内，3 文で構成されるものをデータとして用いる．ニュース記事の表題と，本文中から抜き出した重要なフレーズであるキーフレーズを入力として与え，表題に基づいた内容を持ち，キーフレーズを使用したテキストを生成する．その生成に，レベルに応じた修辞構造と語彙の難易度に基づく制御を加える．

重要なフレーズであるキーフレーズとして，データセットに対して TopicRank アルゴリズム [9] を用い，キーフレーズを事前に抽出する．抽出したキーフレーズは平均で 9.57 個であった．

テキスト全体の構造を捉えるため，修辞構造理論（Rhetorical Structure Theory, RST）を用いる．RST では，テキストを意味の最小単位である Elementary Discourse Unit (EDU) で区切り，EDU 同士の関係性を木構造で考える．関係は Nuclearity と Relation Label の 2 つの要素で表現される．Nuclearity は Nucleus と Satellite の 2 つの値を取る．木の中で子となる 2 つの EDU の内，Nucleus である EDU は Satellite である EDU より重要であることを表す．Relation Label は [2] による 18 種類のものを採用する．木の中で子となる 2 つの EDU が，例えば原因と結果のように，どのような関係になっているかを表す．データセットに対し RST の学習済みパーザである DPLP [10] を用い，修辞構造を得る．問題を簡単にするため，先行研究 [4] に従い，テキスト中で隣り合う EDU 同士の関係のみ考慮することによって，図 2 に示すように修辞構造の木をリスト化する．リストに変換する際，関係が 2 つ連続して現れた場合は，木の中でより深い位置にある関係を優先する．また，木の中央にあるノードが子となるノードより後に現れた場合は，そのノードに対応する関係は無視される．

1）　https://lexile.com/educators/understanding-lexile-measures/about-lexile-measures-for-reading/

3.1　修辞構造と語彙難易度を考慮した訓練

テキスト生成を行う上で基盤となる手法が PAIR [7] である．PAIR では，正解文からキーフレーズの出現する位置以外をマスクしたテンプレートを作成する．そのテンプレートを BART [11] を用いて穴埋めする．穴埋めした結果のトークン列で確率の低いトークンをマスクしたものを次の生成におけるテンプレートとし，再び穴埋めすることを繰り返して最終的なテキストを生成する．本研究では，トークン列を EDU で区切り，その区切りに特殊トークン [EDU] を挿入する．テンプレートを作成する際に，キーフレーズに加えて [EDU] もマスクしないトークンとする．

修辞構造を訓練において考慮するため，テキスト生成モデルの訓練時には，修辞構造に基づくラベルを作成する．正解文のトークン列から，[EDU] に続くトークンである，各 EDU の先頭トークンに RST における関係を対応させ，それ以外のトークンには null を対応させる．ラベル作成の例を図 3 に示す．RST における関係が対応する，各 EDU の先頭トークンの隠れ状態に対して分類器を適用し，RST の関係について分類する．その分類結果とラベルに対して損失を計算し，訓練時の損失に加える．

語彙の難易度については，先行研究 [5] に従い，語彙の難易度を単語とレベルの正の自己相互情報量（Positive Pointwise Mutual Information, PPMI）に基づき考え，各レベルに特徴的な単語に対して重みを加える．単語 w と，Newsela コーパスで設定されたレベル l から，PPMI を求める：

$$\mathrm{PMI}(w,l) = \log \frac{P(w|l)}{P(w)}, \quad (1)$$

$$\mathrm{PPMI}(w,l) = \max(\mathrm{PMI}(w,l),0). \quad (2)$$

$P(w|l)$ はレベルが l の文書に単語 w が出現する確率であり，$P(w)$ は文書集合全体で単語 w が出現する確率である．式 (2) において値を 0 以上としているのは，PMI が負となる単語はレベルによらず広く出現していると考えられ，そのレベル特有の単語とはいえないためである．PPMI を用い，次のように損失を計算する：

$$f(w,l) = \mathrm{PPMI}(w,l) + 1, \quad (3)$$

$$L(\boldsymbol{o},\boldsymbol{y},w,l) = -f(w,l)\cdot \log o_c. \quad (4)$$

ここで，$\boldsymbol{o}$ はロジットベクトルである．最終的な損

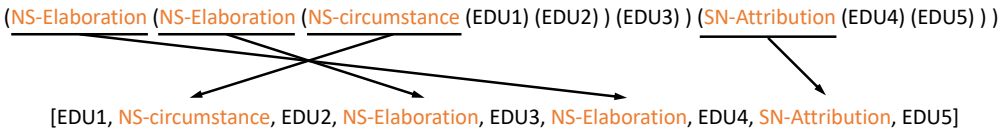

図 2　修辞構造のリスト化

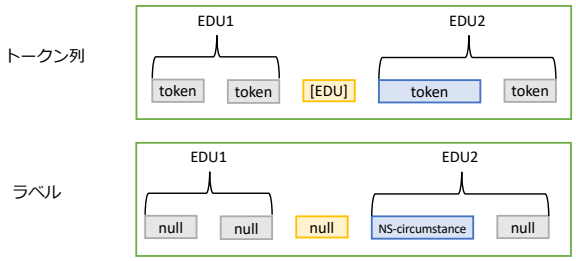

図 3　修辞構造に基づくラベル

失関数は，上記の RST に関する損失とこの語彙難易度に関する損失を足したものになる．

3.2　PPLM を用いた生成の制御

PPLM [6] の手法を用いて修辞構造についての制御を追加する．テキストにトピックや感情といった特定の属性 a を持たせるとき，テキストを x とすると，分布 $p(x|a)$ をモデリングすることが目標となる．このとき PPLM では，ベイズの定理により

$$p(x|a) \propto p(x)p(a|x) \tag{5}$$

となることから，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 $p(x|a)$ が得られると考える．分布 $p(a|x)$ は，トピックの導入の場合はトピックに固有の単語を集めた Bag-of-Words $= \{w_1, \cdots, w_k\}$ を用いて以下のように定義される：

$$p(a|x) = \sum_{i}^{k} p_{t+1}(w_i). \tag{6}$$

感情の導入の場合は単層の分類器 f を用いて以下のように定義される：

$$p(a|x) = f(o_{t+1}). \tag{7}$$

ここで，o_{t+1} はロジットである．PPLM では，これらの分布 $p(a|x)$ を用いて，以下の式 (8) に従って言語モデルの内部状態 H_t（具体的には，Transformer 各層の Key および Value の値）を更新することで，目的の属性を持ったトークンが出力されやすくなるよう調整を行う：

$$\Delta H_t \leftarrow \Delta H_t + \alpha \frac{\nabla_{\Delta H_t} \log p(a|H_t + \Delta H_t)}{||\nabla_{\Delta H_t} \log p(a|H_t + \Delta H_t)||^{\gamma}}. \tag{8}$$

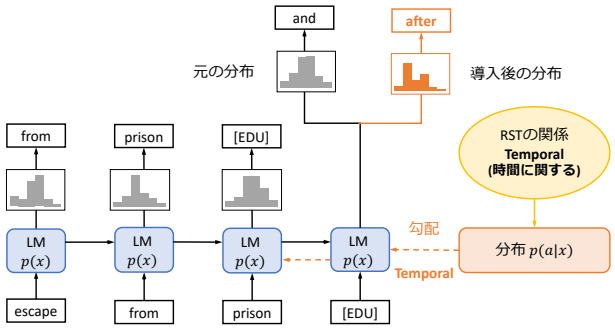

図 4　修辞構造の導入

更新された $\tilde{H}_t = H_t + \Delta H_t$ 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：

$$\tilde{o}_{t+1}, H_{t+1} = \mathrm{LM}(x_t, \tilde{H}_t), \tag{9}$$

$$\tilde{x}_{t+1} \sim \tilde{p}_{t+1} = \mathrm{Softmax}(W\tilde{o}_{t+1}). \tag{10}$$

修辞構造に基づいた制御では，式 (7) において RST の関係に基づく分類器を用いる．この分類器は，3.1 節において修辞構造に基づく損失を付与するために追加で訓練したものを用いる．これにより言語モデルは，図 4 に示すように，EDU の区切りを示す特殊トークン [EDU] に続けて各 EDU の先頭トークンを生成する際に，RST の関係に基づく分類器を用いた PPLM の制御を反映することができる．

また，語彙の難易度については，各難易度レベルに応じた Bag-of-Words を用いた PPLM による制御を行う．先行研究 [5] に基づき定義される語彙の難易度から，Newsela コーパスにおける各レベルに対応する Bag-of-Words を作成する．まず，Newsela コーパスでは，1 つの記事に対して 2 から 12 のレベルの内いずれかのレベルが設定されているが，同じ記事に含まれる文は全て同じレベルに属すると仮定し，同じレベルの文を集めたものを 1 文書とする．語彙の難易度を調べるため，各レベルに対応する文書の集合に対して TF-IDF を計算する：

$$\mathrm{TFIDF}(w, l) = P(w|l) \cdot \log \frac{D}{\mathrm{DF}(w)}. \tag{11}$$

ここで，D は全レベル数であり，Newsela コーパスにおいては $D = 11$ である．$\mathrm{DF}(w)$ は単語 w の出現す

表 1　実験設定

使用モデル	facebook/bart-large
訓練エポック数	20
勾配法	AdamW
バッチサイズ	訓練:10，開発:10，評価:1
損失関数	負の対数尤度
学習率	5×10^{-5}
生成を繰り返す回数	5

るレベルの総数である．TF-IDF を計算することにより，各レベルに特徴的な単語を求めることができる．本研究においては，各レベルに対応する文書に対して TF-IDF が高い上位 200 単語をそのレベルの Bag-of-Words とする．こうして求めた Bag-of-Words から式 (6) のように分布 $p(a|x)$ を求めて式 (8) の計算に用いる．このようにして，指定したレベルの Bag-of-Words に含まれる単語が生成される確率が高くなるように制御する．

4　実験

4.1　実験設定

データとして Newsela コーパスに含まれるニュース記事内の段落の内，3 文で構成されるものを使用する．データのサイズは，訓練が 35,502 事例，開発が 4,438 事例，評価が 4,438 事例である．また，実験のパラメータは PAIR に従った (表 1)．

4.2　評価指標

PAIR [7] に従い，まずは通常のテキスト生成の評価指標である BLEU [12]，ROUGE [13]，METEOR [14] を用いて評価する．

また，修辞構造の評価指標として，各 EDU の先頭トークンに対する正解率を用いる．加えて，語彙難易度の評価指標として，各レベルの Bag-of-Words に対する再現率を用いる．再現率は，生成されたテキスト中に正しく出現した Bag-of-Words に含まれる単語の数を，正解文に出現した Bag-of-Words に含まれる単語の数で割ったものとして定義する．

4.3　実験結果

4.1 節におけるデータセットを対象に，PAIR の手法と，提案手法である PAIR に修辞構造と語彙の難易度に基づく制御を加えたモデルで実験し，評価をした．提案手法では，修辞構造に関する制御のみ加えたモデルと，修辞構造と語彙の難易度の両方の制御を加えたモデルで比較を行なった．

表 2　評価結果

モデル	B-4	R-L	MTR
PAIR-full	43.20	57.21	55.41
+ 修辞構造	44.60	58.01	56.06
+ 修辞構造 + 語彙の難易度	43.17	56.99	54.91

表 3　EDU の先頭に対する正解率・BoW に対する再現率

モデル	正解率	再現率
PAIR-full	31.99	63.90
+ 修辞構造	40.25	65.49
+ 修辞構造 + 語彙の難易度	33.13	67.85

4.4　考察

PAIR を用いて生成した結果と比較して，PPLM による修辞構造の制御を加えた提案手法のモデルでは精度が少し向上した．精度における大きな上昇とはならなかったが，精度を落とすことなく表 3 の正解率において見られるように，正しく修辞構造を反映した生成が行われた例を見ることができた．一方で，修辞構造の制御に加え，レベルに基づいた語彙による制御を加えた生成においては，PAIR と比較して少し精度が落ちる結果となった．この原因として，生成において PPLM を利用すると同じ単語を繰り返し生成しやすくなることが観察されており，それが提案手法においても生じたことが挙げられる．精度の向上には繋がらなかったが，表 3 の再現率において見られるように，指定したレベルの Bag-of-Words 内の単語を用いた生成が確認できた．

5　まとめ

発話相手のレベルに応じたテキストの変化の内，特に修辞構造と語彙の難易度に着目してテキスト生成の制御を行い，発話相手のレベルを考慮したテキストの生成を目指した．結果として，先行研究のモデルより少し精度を改善しながら同時に修辞構造を考慮して生成を行えた例も確認できた．その一方で，同じ単語が繰り返し生成されるという問題が生じた例もあり，手法の改善の余地を確認した．

今後の課題としては，まず評価が挙げられる．特に修辞構造の制御の評価は，各 EDU の先頭トークンを用いた簡易的なものになっているのでより厳密な評価が必要である．また，本稿では修辞構造の一部についてのみ考慮して生成を行なったが，さらに複雑な修辞構造を対象とした実験も必要である．

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] Chao Jiang, Mounica Maddela, Wuwei Lan, Yang Zhong, and Wei Xu. Neural crf model for sentence alignment in text simplification. In **Proceedings of the Association for Computational Linguistics (ACL)**, 2020.

[2] WILLIAM C. MANN and SANDRA A. THOMPSON. Rhetorical structure theory: Toward a functional theory of text organization. **Text - Interdisciplinary Journal for the Study of Discourse**, Vol. 8, No. 3, pp. 243–281, 1988.

[3] 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, , 2015.

[4] Dongyeop Kang and Eduard Hovy. Linguistic versus latent relations for modeling coherent flow in paragraphs. In **Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)**, Hong Kong, China, November 2019.

[5] Daiki Nishihara, Tomoyuki Kajiwara, and Yuki Arase. Controllable text simplification with lexical constraint loss. In **Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop**, Florence, Italy, July 2019.

[6] 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.

[7] Xinyu Hua and Lu Wang. PAIR: Planning and iterative refinement in pre-trained transformers for long text generation. In **Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)**, Online, November 2020.

[8] Sanqiang Zhao, Rui Meng, Daqing He, Andi Saptono, and Bambang Parmanto. Integrating transformer and paraphrase rules for sentence simplification. In **Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing**, Brussels, Belgium, October-November 2018.

[9] 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**, Nagoya, Japan, October 2013.

[10] Yangfeng Ji and Jacob Eisenstein. Representation learning for text-level discourse parsing. In **Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, Baltimore, Maryland, June 2014.

[11] 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**, Online, July 2020.

[12] 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**, Philadelphia, Pennsylvania, USA, July 2002.

[13] 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.

[14] 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**, Ann Arbor, Michigan, June 2005.