

令和 4 年度茨城大学大学院理工学研究科情報工学専攻
修士学位論文
T5 と GPT2 を利用した特定キャラクタ風の言語モデルの
構築

所属 情報工学専攻
著者 岸野望叶 (21NM722Y)
指導教員 新納浩幸教授

令和 5 年 2 月 5 日 (水)

T5 と GPT2 を利用した特定キャラクタ風の言語モデルの構築

著者

岸野望叶 (21NM722Y)

指導教員

新納浩幸教授

論文要旨

本論文では特定の発話者に対する言語モデルの構築手法を提案する。現在、Siri などの対話エージェントが盛んに利用されていたり、RPG などのゲームで大量のセリフが必要になったりする。それらの発話はキャラクタらしさを含んでいることが求められる。しかし、特定のキャラクタに特化した言語モデルの構築を行うには学習データが限られており、性能の向上は困難である。そのため本論文では対象の発話者と同作品に出てくる別人物の発話を、T5 を用いて対象発話者の発話風に変換し、学習データを増補する。その学習データを「ドメイン」の学習データ、対象の発話者の発話を「タスク」の学習データとし、DAPT (domain adaptive pretraining) + TAPT (task adaptive pretraining) の手法でベースの言語モデルとなる GPT-2 に fine-tuning を行った。比較対象として、T5 で変換していない対象発話者と同作品に出てくる別人物の発話や、CSJ (日本語話し言葉コーパス)、CSJ を T5 を用いて対象の発話者風の発話に変換したものを「ドメイン」の学習データとして同様の手法で言語モデルの作成を行った。また T5 の学習データを選ぶ手法として、TFIDF を用いることを提案する。TFIDF 値の高い発話をそれぞれの発話者らしい発話群として捉え、これらの発話から手作業で発話者の特徴を除いて対応する一般的な発話を作る。これらのペアを T5 の学習データとして用いて、一般-対象変換モデルと同作品-一般変換モデルを作成した。その結果、GPT-2 に対象の発話者の発話のみで学習を行った場合のパープレキシティが 46.23 であったのに対し、この手法で行った場合のパープレキシティは 43.93 となり、性能が向上した。また、提案手法で作成した言語モデルで発話の生成を行った。その結果、提案手法で作成したモデルの方が、対象の発話者の発話のみで fine-tuning を行い作成したモデルより発話者の特徴を表す発話が多く生成された。しかし、提案手法で作成したモデルのほうが不自然な発話を生成してしまう側面も見られた。そのため今後の課題として、より自然な発話を生成させること、そのため T5 モデルの学習方法の見直しがあげられる。

Master's Thesis in Scholastic 2023, Major in Computer and Information Sciences,
Graduate School of Science and Engineering, Ibaraki University

Construction of a specific character-like language model using T5 and GPT2

Author : Mika Kishino (21NM722Y)

Adviser : Prof. Hiroyuki Shinnou

Abstract

This paper proposes a method for building language models for specific speakers. Currently, interactive agents such as Siri are being actively used, and games such as RPGs require large amounts of dialogue. Those utterances must contain characterization. However, the accuracy of constructing a language model specific to a particular character is difficult to improve due to the limited training data. For this reason, this paper uses T5 to transform the speeches and the speech of other characters who appear in the same literature where the target speaker appears into the target speaker's speech style to augment the training data. Using the transformed training data as the "domain" training data and the target speaker's original utterances as the "task" training data, we used the DAPT (domain adaptive pretraining) + TAPT (task adaptive pretraining) method to finetune GPT-2, the base language model. We also propose to use TFIDF to select the training data for T5. The utterances with high TFIDF values are taken as a group of utterances that are typical of each speaker, and from these utterances, the corresponding general utterances are created by manually removing the speaker's characteristics. Then, pairs of extracted utterances and utterances from which the speaker's features were manually removed were used. The experiments showed that the perplexity was 46.23 when the GPT-2 was trained with only the target speaker's utterances, whereas it was 43.93 when this method was used. In addition, we evaluated the language model created by the proposed method. The proposed model produced more utterances that expressed the characteristics of the speaker than the model fine-tuned with the target speaker's utterances alone. However, the proposed method tended to generate more unnatural speeches. Therefore, our future work is to generate more natural speeches and to revise the learning method of the T5 model.

目次

第 1 章	序論	6
第 2 章	関連研究	8
第 3 章	T5 と GPT2	11
3.1	事前学習モデル	11
3.2	T5	11
3.3	GPT2	12
第 4 章	T5 と GPT を用いた特定キャラクタ風発話の言語モデルの構築	14
4.1	DAPT+TAPT の手法を用いた特定キャラクタ風発話の言語モデルの構築	14
4.2	T5 による発話文変換モデルの構築	16
4.3	T5 学習データの抽出方法	17
第 5 章	実験	22
5.1	コーパスの収集	22
5.2	DAPT と TAPT のためのデータ作成	22
5.3	実験の設定	24
第 6 章	結果	26
6.1	学習方法と学習データによるパープレキシティの変化	26
6.2	T5 の学習データの抽出方法によるパープレキシティの変化	28
第 7 章	考察	30
7.1	作成したモデルでの文章生成	30

目次	5
7.2 今後の課題	33
第 8 章 結論	36
参考文献	38

第1章

序論

本論文では、小説、アニメあるいはゲームなどに登場する特定のキャラクターを発話者とした言語モデルの構築手法を提案する。

小説、アニメあるいはゲームなどの登場人物は、仮想的にはあるが、ある特徴を有しており、その人物による自然な発話にはその特徴を有したものにすることが必要である。高度化された対話システムにおいても、ディスプレイに表示される仮想的な発話者あるいは実際のロボットの発話では、その見た目から想起される発話者らしい発話が自然である。このような背景から発話者の特徴を反映した発話文を生成する研究がいくつか行われているが、それらはみな規則ベースの手法である（[1] [2] など）。規則ベースの手法は規則の構築が手作業となるためその構築コストが高く、また対象を別の発話者に変更することが困難という問題もある。

本論文では発話者の特徴を有した発話文を生成するために、その発話者の言語モデルを構築することを試みる。発話者の言語モデルはその発話者の発話文を大量に収集することで自動的に構築できる。ただしその発話者は、通常、仮想的な発話者であるため収集できる発話文の量は限られる。そのため本論文では GPT-2 [3] をベースの言語モデルに設定し、対象発話者の少量の発話文によって、そのベースの言語モデルを fine-tuning することで目的の言語モデルを構築する。更に fine-tuning により構築される言語モデルの性能を上げる（パープレキシティを下げる）ために T5 [4] を利用して別人物の発話文を対象発話者風の発話文に変換するモデルを学習し、対象発話者の発話文を増補する。このとき増補された発話文を対象発話者の発話文を含む「ドメイン」の発話文、対象発話者の発話文を「タスク」の発話文と捉え、DAPT (domain adaptive pretraining)+TAPT (task adaptive pretraining) [5] の手法を利用することで効果的な fine-tuning を目指す。

実験では株式会社スクウェア・エニックスのゲームであるドラゴンクエスト IV 内のセリフをコーパスとし、コーパス内で現れるマリベルを対象発話者に設定した。このコーパスから取り出した対象発話者の発話文を訓練データとテストデータに分け、テストデータに対する言語モデルのパープレキシティから構築した言語モデルの評価を行った。訓練データを用いて GPT-2 を fine-tuning した言語モデルのパープレキシティは 46.23 であったが、本論文の提案手法を用いることで 43.93 まで改善できた。

本論文の貢献は以下である。

- (1) 汎用的な大規模言語モデルの fine-tuning を行うことで、特定のキャラクターの特徴をとらえた発話を自動的に作成する言語モデルを構築した
- (2) 対象のキャラクターの登場する作品の別の登場人物の発話をドメインのデータ、対象のキャラクターの発話をタスクのデータとして利用して、DAPT+TAPT の手法を用いて fine-tuning を行う手法を提案した
- (3) 対象のキャラクターの登場する作品の別の登場人物の発話を大規模言語モデルによる変換器を用いて対象キャラクター風に変換し、ドメインのデータとして利用する手法を提案した
- (4) ドメインのデータとしてさまざまなデータを利用した実験を行い、言語モデルの性能をパープレキシティにより比較した
- (5) (2) と (3) を組み合わせた提案手法により作成した特定のキャラクターの発話の言語モデルの作成した発話を示し、考察を行った
- (6) 本研究は学習データが少ない他のドメインに特化した学習を行う際にも応用できる可能性がある。

第 2 章

関連研究

Suchin Gururangan ら [5] は、言語モデルを構築する際にドメインに適応させる事前学習である DAPT (domain adaptive pretraining) のみ、タスクに適応させる事前学習である TAPT (task adaptive pretraining) のみ、DAPT+TAPT を利用した場合をそれぞれ比較し、DAPT+TAPT が最もよい性能であることを示している。DAPT のみの実験では、生物医学の論文、コンピュータサイエンスの論文、ニュース、レビューの 4 つをドメインとし、それぞれのドメインのデータで RoBERTa に追加学習を行い分類実験を行った。その結果、ニュース以外のドメインで性能の向上が見られた。ニュースドメインで改善が見られなかったのは、事前学習のデータにニュースデータが含まれているからと述べている。また、対象のタスクのドメイン以外のデータを使って DAPT を行った場合、性能が悪くなることも示した。TAPT のみの実験では、DAPT の実験と同様のドメインのタスクのデータで追加学習、分類実験を行った。その結果、少量のデータであるにも関わらずニュースドメインのタスクである AGNews やレビュードメインのタスクである Helpfulness など、一部のタスクで DAPT の性能を上回っていた。また勝又ら [6] は話し言葉 BERT の作成に DAPT+TAPT の手法を利用し、DAPT+TAPT の 2 種類の追加学習を組み合わせることで、文境界推定、重要文抽出などの様々なタスクで安定して高い性能のモデルを作成した。また同時に話し言葉においてもこの手法は有効であることを示した。彼らは DAPT を行うことで BERT のパラメータが TAPT を行うのよよい初期値となり、少ない学習ステップで高い性能になったのではないかと述べている。

また、キャラクター性についての研究も盛んにおこなわれている。Mairesse と Walker [7] によって開発された PERSONAGE (personality generator) は最初の高度なパラメータ化された会話生成器である。彼らは PERSONAGE が外向性の度合いに沿ってバリエー

ションのある発話を作成できることを示した。Walker ら [8] は映画の対話コーパスを収集し言語構造やキャラクターの原型のアノテーションを行った。さらに彼らはジャンル、性別、ディレクタ、映画の年代のようなグループによって言語のスタイルを分類するためにキャラクターの言語モデルの実験を行った。宮崎ら [9] は日本語の会話を特徴付ける言語表現を基礎的に分析し、部分的に言い換えることで会話を特徴付ける技術を開発した。また、宮崎ら [10] は、日本のフィクションの登場人物の言語的特徴のカテゴリーを報告した。さらに会話の特徴づけやバリエーションを豊かにする方法として、各文の機能部分を対象人物に適した確率で言い換えて、読者が対話者の特徴づけを理解できるかどうかを確認する実験を行った [11]。また、日本の架空の人物の発話を特徴付けるために、日本語の音変化表現に着目し、これらの表現を収集し、分類した研究もある [1]。さらに、奥井・中辻 [12] は、日本語の対話システムにおいて、ポインタ生成機構を用いて、複数の異なるキャラクターの応答を参照し、様々な応答を生成した。彼らは、少ないデータ量で応答の特徴づけを学習した。また岸野ら [2] [13] [14] の研究では、SentencePiece を用いて単語分割を行い、TFIDF を用いてキャラクターの特徴と考えられる単語を抽出した。

また、意味を保持したまま、テキストのスタイルを変換する「テキストスタイル変換」の研究も本研究と深い関係がある。キャラクター性を変換する論文 [9] をはじめとして、シェークスピアの時代の文への変換 [15] や、散文調べの変換 [16]、さらにテキスト音声合成を目的とした、流暢性や方言の変換を行った論文 [17] がある。本研究は、人手により意味を保持したままキャラクター性の削除の処理を行い、そのデータを利用して T5 を用いて自動的なテキストスタイル変換を利用しているが、本研究は最終的に特定のキャラクターの発話の言語モデル作成を目的としている点で異なっている。

また、応答生成ではキャラクター性を反映した応答を行う研究も行われている。Jiwei Li ら [18] は、話者の一貫性の問題に対応するため、人物の背景情報や話し方などの個人の特徴を捉えた embedding として表現することで人物像ベースのモデルを作成した。これらは 2 人の発話者のモデルから、発話者同士の関係の特性を捉えることができる。このモデルはベースラインである sequence-to-sequence モデルに対してパープレキシティと BLEU スコアの両方で性能を向上させ、さらに人による話者の判定でも同様に向上させた。Stephen Roller ら [19] は、人間らしい魅力的な応答生成に必要なものの一つとして人物像の一貫性をあげており、キャラクターのプロフィールを与えて応答文を生成する方法を提案した。これらの研究もキャラクター性を反映させた発話の生成ではあるが、本研

究は特定の人物像のみでなく特定のキャラクタ（個人）に限定した発話の生成を目的としていること、またそのキャラクタが実在しないため、存在する学習データに限りがあるという点で異なっている。

第3章

T5 と GPT2

本章では本論文で利用する T5 と GPT2 について説明する。

3.1 事前学習モデル

事前学習モデルは、特定タスクに特化しない大量のデータを使用して学習したモデルのことである。その後、その事前学習モデルに fine-tuning を行うことで、タスク特化のモデルを高精度に構築することができる。自然言語処理タスクにおいて、事前学習モデルとして大量のテキストデータを使用して学習した言語モデルを使用し、その後タスクに応じて、そのモデルに fine-tuning を行い、質問応答タスクや機械翻訳タスクなどに特化させることができる。これにより、事前学習モデルは、自然言語処理タスクにおいて高精度なモデル構築に貢献することができる。

3.2 T5

T5 (Text-To-Text Transfer Transformer) とは、入出力をテキストに統一し、fine-tuning を行うことで様々なタスクを解くことができるモデルである。T5 は、Transformer をベースに構築されている。Transformer は、自然言語処理タスクにおいて高精度なモデルを構築するために広く使用されているアーキテクチャである。T5 は、Encoder-Decoder を使用している。Encoder は、入力テキストを理解するために使用され、Decoder は、出力テキストを生成するために使用される。Encoder は、多層の Transformer ブロックから構成され、各ブロックは、Self-Attention と Position-wise

表 3.1: GPT2 モデルの種類

	small	mediam	large	XL
パラメータ数	117M	345M	762M	1542M
レイヤー数	12	24	36	48
次元数	768	1024	1280	1600

Feed-Forward Network を組み合わせたものである。これにより、入力テキスト中の各位置の関係性を学習することができる。Decoder は、Encoder と同様に、多層の Transformer ブロックから構成され、Self-Attention と Position-wise Feed-Forward Network を組み合わせている。T5 モデルは、大量のパラメータを持ち、これらのパラメータを使用して、入力テキストから出力テキストを生成するためのマッピングを学習する。また、T5 は、文脈タスクに対応するために、タスクスペシフィック文字列を入力テキストの先頭に付与する。これにより、モデルは入力テキストのタスクを理解し、適切な出力を生成することができるようになる。

3.3 GPT2

GPT2 とは、OpenAI が開発した文章生成を行うことができる事前学習モデルである。大規模なコーパスを使うことで zero-shot や few-shot の追加学習でも高精度で汎用的に使用できる。事前学習に用いられたデータセットは、Reddit というソーシャルメディアからリンクされているページのデータを収集し、文章の質や内容の重要性を考慮するため、その中でも一定以上の評価を得ている投稿に絞った。その結果収集された 800 万文書を事前学習に用いた。GPT2 のモデルは 4 段階公開されており、それぞれの違いを表 3.1 に示す。GPT2 の構造は transformer の decoder 部分を何層にも重ねたものである。transformer の masked multi-head attention を使用することで GPT2 は与えられた単語を元に逐次的に次の次に出てくる単語を予想することで文章を生成することができる。本研究では GPT2 の small モデルを使用した。

GPT2 のインプット時には、BPE (Byte Pair Encoding) が使われている。BPE とは、文書内の低頻度の単語をバイト単位まで分割することで未知語を扱うことを減らすことができるアルゴリズムである。本実験ではゲームのキャラクターの発話を利用した

ため、辞書には載っていない語尾などに使われる単語や「アルス」などの固有名詞が多く使われているが、BPEによってそれらをひとまとまりの単語として扱うことができる。

第 4 章

T5 と GPT を用いた特定キャラクター風発話の言語モデルの構築

4.1 DAPT+TAPT の手法を用いた特定キャラクター風発話の言語モデルの構築

本論文では特定キャラクター風発話の言語モデルを構築するために、DAPT+TAPT の手法を利用する。DAPT とは Domain-Adaptive Pretraining の略であり、ターゲットが属するドメインのデータで学習を行うものである。TAPT とは Task-Adaptive Pretraining の略であり、対象のタスクのデータで学習を行うものである。DAPT+TAPT の学習方法とは、既に一般的なコーパスで事前学習された言語モデルに対し、対象のドメインのデータで追加の事前学習を行い、さらに対象のタスクのデータで追加の事前学習を行う方法である [5]。

本研究では DAPT+TAPT の基となる言語モデルに GPT-2 の small サイズのモデル^{*1}を使用し、TAPT と DAPT の実験を行う。このモデルはウェブサイトから集められた 800 万文書のデータセットを利用し、学習させたものである。

本研究の提案手法を図 4.1 に示す。本研究では TAPT と DAPT の手法を用いて特定キャラクター風発話の言語モデルの構築を行う。この際、特定キャラクター風発話を行う言語モデルを構築することを目的としているため、TAPT に用いる「対象のタスクのデータ」を、特定キャラクターの発話とする。DAPT に用いる「対象の属するドメインのデー

^{*1} <https://github.com/tanreinama/gpt2-japanese>

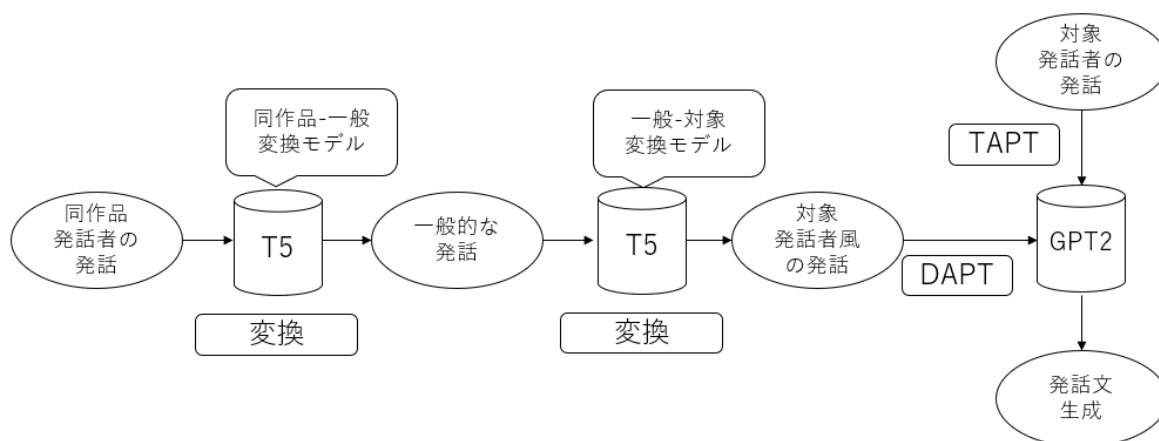


図 4.1: 提案手法の流れ

タ」には、そのキャラクタが登場する作品（以降、同一作品と表記する）の別の登場人物の発話を利用することが考えられる。そのまま利用することも可能であるが、本研究ではさらに、これらの発話を対象の発話者風の発話に変換して、対象の発話者の発話の疑似データを作成する。この疑似データの作成により、対象の発話者の発話を増補したとみなすことができる。本研究では、このデータを DAPT 用の学習データとすることで、対象の発話者の発話の言語モデルの性能を上げることを試みる。ここで対象の発話者風の発話への変換には T5 モデル^{*2}を利用する。本研究では、以下の手順を用いて対象発話者の発話を増補し、GPT2 の fine-tuning を行う（図 4.1）。

- Step-1 T5 モデルを用いて同一作品の別の登場人物の発話を一般的な発話に変換する
- Step-2 Step-1 で作成した一般的な発話を T5 モデルを用いて対象発話者風の発話に変換する
- Step-3 Step-2 で作成した対象発話者風の発話を DAPT の疑似的な学習データとして扱い、GPT-2 の fine-tuning を行う
- Step-4 （変換で得たものではない）実際の対象発話者の発話を TAPT の学習データとして扱い、さらに GPT-2 の fine-tuning を行う

^{*2} <https://huggingface.co/sonoisa/t5-base-japanese/discussions>

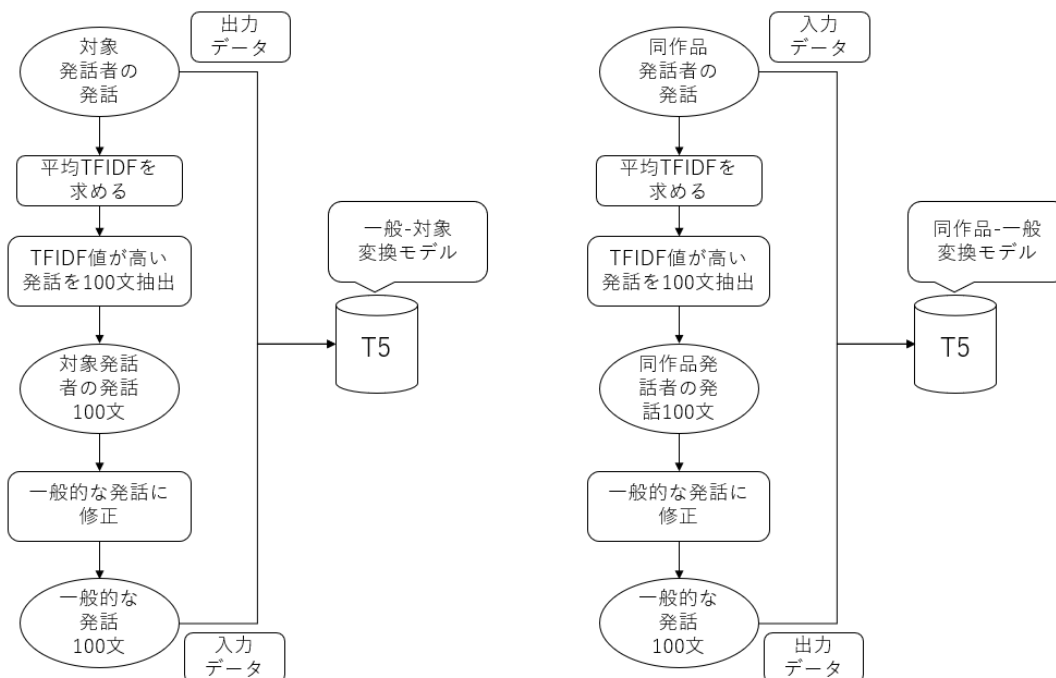


図 4.2: 一般-対象変換モデル作成フロー（左）と、同作品-一般変換モデル作成フロー（右）

4.2 T5 による発話文変換モデルの構築

学習データを増補するため、以下の2種類の T5 モデルを作成した（図 4.1 参照）。

1. 同作品-一般変換モデル：同一作品の別の登場人物の発話を一般的な発話に変換する。同一作品の別の登場人物の発話を入力、一般的な発話に修正したものを出力として学習した。
2. 一般-対象変換モデル：一般的な発話を対象の発話者風の発話に変換する。対象の発話者の発話を一般の発話に修正したものを入力、修正前の対象発話者の発話を出力として学習した。

それぞれのモデルの作成のフローを図 4.2 に示す。

一般-対象変換モデルは、まず対象の発話者の発話から TFIDF を求め、それを用いてより対象キャラクタの特徴を捉えていると考えられる発話を 100 発話抽出した。詳しい抽出方法は 3.3 節で述べる。抽出した 100 発話を、そのキャラクタの特徴をなくした一般的な発話に手作業で修正した。この修正により得られた一般的な発話 100 発話を T5

学習時の入力データ、対象の発話者の発話を出力データとして T5 モデルの学習を行った。得られた T5 モデルが一般-対象変換モデルである。

同作品-一般変換モデルも同様に、同作品の 5 名の発話者の発話から 100 発話ずつ、合計 500 発話抽出し、一般的な発話に修正した。そして同作品の発話者の発話を入力データ、一般的な発話を出力データとして学習を行った。

ここでの一般的な発話とは、性別、年代、人物像を想像させないような発話である。例えば、対象発話者であるキャラクタの一人称である「あたし」から「私」、発話の末尾につけられる口癖である「だわ」や「よね」から「です」に修正することで作成した。これは「私」や「です」の方が使用者として考えられる人物や使われるシチュエーションが広いと考えられるからである。ここでの修正作業はルールに基づいたものではなく、筆者の主観で行われた。また修正に際して、砕けた雰囲気からフォーマルな雰囲気になる、元の発話は語気の強いものだが、それがなくなるなど、若干のニュアンスが変わることは容認した。これは発話の持つ雰囲気が発話者の雰囲気でもあると考えたためである。「あら」や「まあ！」のような感嘆詞はキャラクタの特徴を表していると考えたが、代わりの一般的な表現が思いつかなかったこと、無くても発話の意味が変わらないことから削除した。ここでの手作業での対象発話者の発話の修正例を表 4.1 に、対象発話者と同作品に出てくる人物の発話の修正例を表 4.2 示す。

この 2 つのモデルを使って変換した文章例を表 4.3 に示す。変換した際に同じ単語が何度も繰り返されている文章は GPT-2 の学習データからは排除した。不自然な文章であるもの、変換前の発話と若干意味が異なるものはそのまま利用した。

4.3 T5 学習データの抽出方法

T5 の学習に用いる発話データは、よりキャラクタの特徴を含んだものがよいと考えた。そのため、T5 の学習に用いた発話データの抽出を以下の方法で行った。

まず、対象の発話をそれぞれ T5 で用いられているトークナイザーを用いて単語に分割した。そして単語ごとに以下の式で TFIDF 値を求めた。

$$tf(t, d) = \frac{n(t, d)}{\sum_{s \in d} n(s, d)} \quad (4.1)$$

ここで、 $tf(t, d)$ は文書 d のサブワード t における tf 値、 $n(t, d)$ は文書 d におけるサブワード t の出現回数、 $\sum_{s \in d} n(s, d)$ は文書 d における全サブワードの総出現回数を示す。

表 4.1: マリベル (対象発話者) の発話の手作業での修正例

元の発話	修正後の発話
あたしのがまん強さに感謝しなさいよね。	私の我慢強さに感謝してください。
……あたしたちがだからやるしかないのよね。	私たちだからやるしかありません。
今回はあたしたちもいい経験しちゃったよね。	今回は私たちもいい経験してしまいました。
ゼボットさんのとこなんてあたしあんまり行きたくないけどしかたないわよね。	ゼボットさんのところにはあまり行きたくありませんが仕方ありません。
ここ村のそばよね？あたしたち生きてるのよね？……助かった！	ここは村のそばですよ？私たちは生きていますよね？助かりました。
アイツってば相変わらず情けないヤツよねー。	彼は相変わらず情けない人です。
あたしといっしょに旅ができるなんて光栄に思いなさいよね。	私と一緒に旅が出来ることを光栄に思ってください。
ここであたしに化けた敵を見てるとよく分かるわ。あたしって美人よねえ〜。	ここで私に化けた敵を見てるとよく分かります。私は美人です。
あたしもいっしょにいくわ！いいわよね？	私も一緒に行きます。いいですね？
あたしの状態がどうなっているのかちゃんと確認してからやりなさいよね！	私の状態がどうなっているのか確認してからやってください。

$$idf(t) = \log \frac{N}{df(t)} \quad (4.2)$$

ここで、 $idf(t)$ は文書 d のサブワード t における idf 値、 N は総文書数、 $df(t)$ はサブワード t が出現した文書数を示す。

$$TFIDF = tf(t, d) * idf(t) \quad (4.3)$$

TFIDF 値は tf 値と idf 値を掛け合わせることで求められる。ここで TFIDF 値が高い

表 4.2: 対象発話者と同作品に出てくる人物の発話の手作業での修正例

発話者名	元の発話	修正後の発話
アイラ	さあアルスわたしたちはわたしたちの旅を続けましょう。	さあアルス私たちは私たちの旅を続けましょう。
アイラ	わたしたちも自分自身にしかできない戦いを始めましょう。	私たちも自分自身にしかできない戦いを始めましょう。
ガボ	えーオイラ食いたいぞー！	私は食べたいです。
ガボ	そうだぞ。オイラたちは強えーんだぞ。	そうです。私たちは強いのです。
キーファ	でもオレはオレさ！いつかきっとオレにしかできないことを見つけてやるぞ！	でも私は私です。いつかきっと私にしかできないことを見つけます。
キーファ	どうかオレたちを受け入れてくれ！！オレたちに新たな道を開いてくれ！！	どうか私たちを受け入れてください。私たちに新しい道を開いてください。
メルビン	老人でござるか？わわしではないでござるよ！	老人ですか。私は老人ではありません。
メルビン	アルスどのメルビンでござる。聞こえるでござるか？	アルスメルビンです。聞こえますか。
リーサ姫	まあっお兄さまお帰りなさい！	兄さんお帰りなさい。
リーサ姫	いつまでもお兄さまっ子じゃお兄さまに笑われてしまうもんね。	いつまでも兄さんっ子では兄さんに笑われてしまいますから。

単語はそのキャラクタの特徴を表していると考え [2]。対象発話者の TFIDF 上位 10 発話、下位 10 発話を表 4.4 に示す。

続いて、これらの単語の TFIDF 値を用いて発話ごとの平均 TFIDF 値を求める。これは発話内に出てくる TFIDF 値の合計を出現単語数で割ることで得た。また単語数が少ない発話だと、「あ、アルス。」のような名前を呼んでいるだけの発話など、TFIDF 値は高くなるが発話者の特徴を含まない発話が多く含まれる可能性があるため、10 単語以内

表 4.3: T5 での発話文変換例

別人物の発話	同作品-一般変換モデルで変換した一般的な発話	一般-対象変換モデルで変換した対象発話者風の発話
あんな～んだ。さっき私たちの村をたずねてきた旅の人ね。	さっき私たちの村をたずねてきた旅の人ですね。	さっきあたしたちの村をたずねてきたのよね。
ありがとうございます。あなたがたのおかげでこの牢ごくから出られました。	ありがとうございます。あなた達のおかげでこの牢獄から出られました。	ありがとう!あんな達のおかげでこの牢獄から出獄したのよね。
こらーっ!オイラこのままのカッコでいいんだってばー!	私はこのままの格好でいいと言ってるでしょう。	いやあたしはこのままでいいっていうのよね。
おっアルスやっと来たか!	アルスやっと来ました。	アルスようやく来たわよね。
最後の決戦をまえにして神がわしを封じたのでござった。のちの世のために……。	最後の決戦をまえにして神が私を封じたのです。のちの世のために。	最後の決戦をまえにして神があたしたちを封じたのね。
お父さまはお兄さまといちばん親しいあなたからお話を聞いたかったのねきっと。	父さんは兄さんと一番親しいあなたから話を聞いたかったのですきっと。	あんなといちばん親しいひとから話を聞いたかったのよね。

表 4.4: マリベル (対象発話者) の TFIDF 値上位・下位 10 発話

上位順位	上位発話	下位順位	下位発話
1	あたしのがまん強さに感謝しなさいよね。	1	な…なんなのよ！ここは！
2	……あたしたちがだからやるしかないのよね。	2	……たいくつ……なのよう。すう……すう……。
3	今回はあたしたちもいい経験しちゃったよね。	3	……。な…なによアルス。文句でもあるわけ！？
4	ゼボットさんのとこなんてあたしあんまり行きたくないけどしかたないわよね。	4	うう……やだなあこういうの……。
5	ここ村のそばよね？あたしたち生きてるのよね？……助かった！	5	……さっそく出たわね。いくわよアルス！
6	アイツってば相変わらず情けないヤツよねー。	6	クモやトカゲ……。うげげだわね！
7	あたしといっしょに旅ができるなんて光栄に思いなさいよね。	7	なによあの人アルスたちの知り合いなの？
8	ここであたしに化けた敵を見てるとよく分かるわ。あたしって美人よねえ～。	8	すぐなんにでも名前をつける人だわね……。
9	あたしもいっしょにいくわ！いいわよね？	9	なにがって？いろいろよ。いろいろ。
10	あたしの状態がどうなっているのかちゃんと確認してからやりなさいよね！	10	なに？王さまどうかしたのアルス？

の発話は除外した。その中で平均 TFIDF 値が高い発話ほどキャラクタの特徴を含んだ発話と考え、平均 TFIDF 値が高い上位 100 発話を T5 の学習に利用した。

第 5 章

実験

5.1 コーパスの収集

本実験の提案手法モデルの学習や評価時に用いられたコーパスは以下の方法で収集、作成した。

- (1) インターネットから発話を収集する。^{*1}
- (2) 「」（鍵括弧）や、発話者を示すキャラクター名表記を削除
- (3) 文末に「！」や「？」などの記号がない場合、文末が分かりやすくなるよう、「。」を入れる。
- (4) 鍵括弧内を1つの発話として区切る。

本研究では対象の発話者としてゲーム、ドラゴンクエスト IV のキャラクターであるマリベルを利用し、同一作品の別の登場人物にはドラゴンクエスト IV のキャラクターであるキーファ、ガボ、メルビン、アイラ、リーサ姫を利用した。マリベルを対象としたのは、セリフを収集したキャラクターの中でもセリフ量が多かったためと、性別や人物像を表す特徴的な話し方をしているためである。

ここで収集した各キャラクターと発話数を表 5.1 に示す。

5.2 DAPT と TAPT のためのデータ作成

本研究では以下の 5 種類のデータを利用した。

^{*1} <http://sutton-kyouwa.com/g/dq7mari.htm>

表 5.1: 収集コーパス

発話者名	発話数
マリベル	4131
キーファ	1614
ガボ	3062
メルビン	2172
アイラ	1815
リーサ姫	148

- 対象の発話者の発話
- 同一作品の発話者の発話：同一作品の別の登場人物の発話
- CSJ：日本語話し言葉コーパス (Corpus of Spontaneous Japanese: CSJ) [20]
- 対象風同作品の発話：「同一作品の発話者の発話」を「対象の発話者の発話」風に T5 を用いて変換したもの
- 対象風 CSJ：「CSJ」を「対象の発話者の発話」風に T5 を用いて変換したもの

本研究では「対象の発話者の発話」を TAPT に使用する対象タスクのデータとする。また、「同一作品の発話者の発話」、「対象風同作品の発話」を DAPT に利用する対象ドメインのデータとする。さらに、比較対象として「CSJ」、「対象風 CSJ」を DAPT に用いる対象ドメインのデータとして利用した。

「対象風同作品の発話」は、まず「同一作品の発話者の発話」を同作品-一般変換モデルを使い一般の発話に変換する。続いて一般-対象変換モデルを使い先ほど変換した一般の発話から「対象風同作品の発話」に変換する。以上の2段階の変換を行い「対象風同作品の発話」を作成した。「対象風 CSJ」は一般-対象変換モデルを使い CSJ を対象発話者風の発話に変換させて作成した。

5.3 実験の設定

5.3.1 TAPT, DAPT を利用した GPT-2 の学習方法

GPT-2 をそれぞれ以下の方法で fine-tuning を行い、モデルごとに平均パープレキシティを求めた。

- (a) TAPT のみ
- (b) DAPT のみ
- (c) DAPT+TAPT

(a)TAPT のみの手法を用い、「対象の発話者の発話」のみで学習したものをベースラインとする。(b)DAPT のみの手法では、「同一作品の発話者の発話」、「対象風同作品の発話」をそれぞれ学習データとして GPT-2 に fine-tuning を行い、パープレキシティを求めた。このうちの「対象風同作品の発話」は T5 による変換によって作成した疑似データである。また、比較対象として「CSJ」、「対象風 CSJ」を DAPT 用の対象のドメインデータとして利用する手法についてもそれぞれ実験した。比較手法の「対象風 CSJ」も、T5 によって変換されたものである。(c)DAPT+TAPT の手法では、(b)DAPT のみの手法で作成したモデルにさらに「対象の発話者の発話」で GPT-2 に fine-tuning を行った。

GPT-2 の学習は 20 エポックずつ 100 エポックまで行い、検証データを用いてパープレキシティを算出し、最も性能がよかった学習回数でテストを行った。この時、最も性能がよかった学習回数が 100 エポックだった場合、追加で 200 エポックまで学習、検証を行った。

「対象の発話者の発話」の 9/10 を学習データとし、GPT-2 と T5 の学習に用いた。残りの 1/10 をパープレキシティの算出に利用した。そのうち 1/5 を検証データとし、4/5 をテストデータとした。

5.3.2 T5 の学習データセット

T5 の学習データセットには以下の 3 種類を利用した。

- A TFIDF 値上位 100 発話
- B TFIDF 値上位 200 発話

C ランダムな 100 発話

データセット A,B は、3.3 節で説明した方法で抽出した。またデータセット B,C は

- 「対象風同作品の発話」をドメインの学習データとし (b)DAPT のみの手法で学習
- 「対象風同作品の発話」をドメインの学習データ、「対象の発話者の発話」をタスクの学習データとし (c)DAPT+TAPT の手法で学習

の 2 パターンの検証を行った。A に関しては、上記 2 パターンを含む 8 パターン行った (表 6.1 の 2~9 行目参照)。

第6章

結果

6.1 学習方法と学習データによるパープレキシティの変化

TFIDF 値上位 100 発話を利用した場合の (a)TAPT のみ、(b)DAPT のみ、(c)DAPT+TAPT の手法の学習データを変えた際のパープレキシティの変化を表 6.1 に示す。ベースライン ((a)TAPT のみ) よりよい評価値のもの (パープレキシティが低いもの) を太字で表す。

まず、表 6.1 において、同一作品の発話者の発話をドメインのデータ、対象の発話者の発話をタスクのデータとして (c)DAPT+TAPT の手法で学習を行った場合と、対象風同作品の発話をドメインの学習データ、対象の発話者の発話をタスクのデータとして (c)DAPT+TAPT の手法で学習を行った場合がベースラインを超える結果となっ

表 6.1: 学習データによるパープレキシティ (PPL) の変化

学習方法	T5 の学習データの抽出方法	ドメインの学習データ	タスクの学習データ	PPL
(a)TAPT のみ			対象の発話者の発話	46.230
(b)DAPT のみ	TFIDF 値上位 100 発話	同一作品の発話者の発話		59.419
(b)DAPT のみ	TFIDF 値上位 100 発話	CSJ		136.737
(b)DAPT のみ	TFIDF 値上位 100 発話	対象風同作品の発話		72.903
(b)DAPT のみ	TFIDF 値上位 100 発話	対象風 CSJ		110.846
(c) DAPT+TAPT	TFIDF 値上位 100 発話	同一作品の発話者の発話	対象の発話者の発話	45.606
(c) DAPT+TAPT	TFIDF 値上位 100 発話	CSJ	対象の発話者の発話	48.245
(c) DAPT+TAPT	TFIDF 値上位 100 発話	対象風同作品の発話	対象の発話者の発話	44.173
(c) DAPT+TAPT	TFIDF 値上位 100 発話	対象風 CSJ	対象の発話者の発話	49.750

ている。このことから、T5 の利用にかかわらず、同一作品の発話者の発話を用いて (c)DAPT+TAPT の手法で fine-tuning することが有効であることが分かる。また、提案手法が最もよい結果となっており、同一作品の発話者の発話の T5 を利用した変換が有効だったことが分かる。

さらに、それぞれの手法の有効性を見ていく。まず、ドメインの学習データが「同一作品の発話者の発話」、「CSJ」、「対象風同作品の発話」、「対象風 CSJ」の全ての場合において、(b)DAPT のみの手法で学習するよりも、(a)TAPT のみのベースラインの方が良いことが見て取れる。また、同じく全ての場合において、(b)DAPT のみの手法で学習するよりも、タスクの学習データとして対象の発話者の発話を用いた (c)DAPT+TAPT の手法で学習を行った場合の方がパープレキシティが低くなっていることが確認できる。このことから、特定キャラクタ風発話の言語モデルの作成において、対象の発話者の発話を利用する重要性が分かる。

次に、T5 による疑似データの効果を網羅的に見る。まず、学習手法 (b)DAPT のみで「同一作品の発話者の発話」と「対象風同作品の発話」を比較した場合、T5 を使って対象の発話者風の発話に変換したはずの「対象風同作品の発話」のほうがパープレキシティが高かった。これは、T5 を使って2回変換を行っており、対象の発話者の特徴を掴んだ発話にはなっていない不自然な日本語になってしまったものが多くあったためだと考える。しかし、学習手法 (c)DAPT+TAPT で学習した場合同士で比較すると「対象風同作品の発話」の方が良い結果になった。これは後で「対象の発話者の発話」で学習することで発話の不自然さから受ける影響が少なくなったためではないかと考えられる。

次に、CSJ を対象とした場合の T5 による変換の効果に注目する。学習手法 (b)DAPT のみで「CSJ」と「対象風 CSJ」を比較した場合、T5 による対象の者風に変換したモデル、「対象風 CSJ」の方がパープレキシティが下がっている。しかし、学習手法 (c)DAPT+TAPT で学習した場合同士で比較すると「対象風 CSJ」より「CSJ」のパープレキシティが低い。このため、CSJ を対象とした場合の T5 による変換は (b)DAPT のみの手法を利用した際には有効であるが、(c)DAPT+TAPT の手法で学習する場合には有効ではないことが分かる。これは「同一作品の発話者の発話」を T5 で変換した場合とは逆の結果となった。

さらに見てみると、「CSJ」や「対象風 CSJ」を学習データに利用したモデルは、ベースラインと比較してパープレキシティが改善されなかった。これは「CSJ」には物語に出

表 6.2: T5 用の学習データの抽出方法によるパープレキシティ (PPL) の変化

学習手法	T5 の学習データの抽出方法	ドメインの学習データ	タスクの学習データ	PPL
(b)DAPT のみ	TFIDF 値上位 100 発話	対象風同作品の発話		72.903
(b)DAPT のみ	TFIDF 値上位 200 発話	対象風同作品の発話		69.881
(b)DAPT のみ	ランダムに選んだ 100 発話	対象風同作品の発話		72.622
(c) DAPT+TAPT	TFIDF 値上位 100 発話	対象風同作品の発話	対象の発話者の発話	44.173
(c) DAPT+TAPT	TFIDF 値上位 200 発話	対象風同作品の発話	対象の発話者の発話	43.930
(c) DAPT+TAPT	ランダムに選んだ 100 発話	対象風同作品の発話	対象の発話者の発話	47.892

てこない学術的な単語が多く出てくるため、語尾や一人称が対象の発話者風になっても、その他の部分でパープレキシティが下がらなかったためだと考える。この文書としての性質が、疑似データの有効性についての振る舞いも分けた可能性がある。

6.2 T5 の学習データの抽出方法によるパープレキシティの変化

T5 用の学習データの抽出方法によるパープレキシティの変化を表 6.2 に示す。表 6.1 同様、ベースライン ((a)TAPT のみ) よりよい評価値のもの (パープレキシティが低いもの) を太字で表す。

表 6.2 の学習手法 (b)DAPT のみ、(c)DAPT+TAPT のみ、のそれぞれについて、T5 の学習データの抽出方法でのパープレキシティの変化を見ていく。

(b)DAPT のみの手法の場合、もっともよい結果になったのは TFIDF 値上位 200 発話を T5 の学習データとした場合である。これは単純に学習データ量が多いことが原因の可能性もある。学習データに TFIDF 値上位 100 発話を使った場合とランダムに選んだ 100 発話を使った場合の結果の差は 0.281 と大差ではないが、ランダムに 100 発話選んだ場合のほうがよい結果になった。このことから、学習手法 (b)DAPT のみ の場合は、T5 の学習データを TFIDF を用いて選ぶことの有効性は見られなかった。

次に、学習手法 (c)DAPT+TAPT の手法の場合を見ていく。このとき、もっともよい結果になったのは T5 の学習データに TFIDF 値上位 200 発話を選んだ場合であり、また TFIDF 値上位 100 発話を選んだ場合がそれに僅差で続く結果になった。TFIDF 値上位 100 発話を選んだ場合とランダムに 100 発話選んだ場合を比較すると、TFIDF 値

上位 100 発話を選んだ場合、ランダムの時よりも、パープレキシティが 3 以上下がっている。このことから、学習手法 (c)DAPT+TAPT の手法では、T5 の学習データとして TFIDF 値を利用することは有効であること、特に学習データ数を増やすよりも効果的であることが確認できた。

第7章

考察

7.1 作成したモデルでの文章生成

最も結果がよかったモデルは、T5の学習データの抽出方法としてTFIDF上位200発話を選び、ドメインの学習データとして対象風同作品の発話、タスクの学習データとして対象発話者の発話を用いて作成したものである。このモデルを提案手法モデルと呼ぶ。このモデルでランダムに100発話の生成を行い、生成された文に対して主観的な評価を行った。評価は、以下に分類して行った。

1. 日本語として不自然な文章
2. 違和感がある文章
3. 文章としては問題ないが、対象キャラクターの発話とは思えない文章
4. 対象キャラクターの発話とも見なせる文章（はっきりとした特徴はないが、場面によっては対象キャラクターの発話として考えられる）
5. 明らかに対象キャラクターの特徴を含む文章

また比較対象として、対象の発話者の発話のみで学習を行ったモデルでも同様に100発話生成し、評価した。このモデルをベースラインモデルと呼ぶ。これら評価結果を図7.1に示す。

その結果、提案手法モデルの方が、キャラクターの特徴がはっきり表れている発話を多く生成した。また、提案手法モデルの方が、キャラクターらしくない発話の生成数が少なかった。

生成例の一部を表7.1に示す。表7.1の1~5番の例が提案手法モデルで生成したもの、

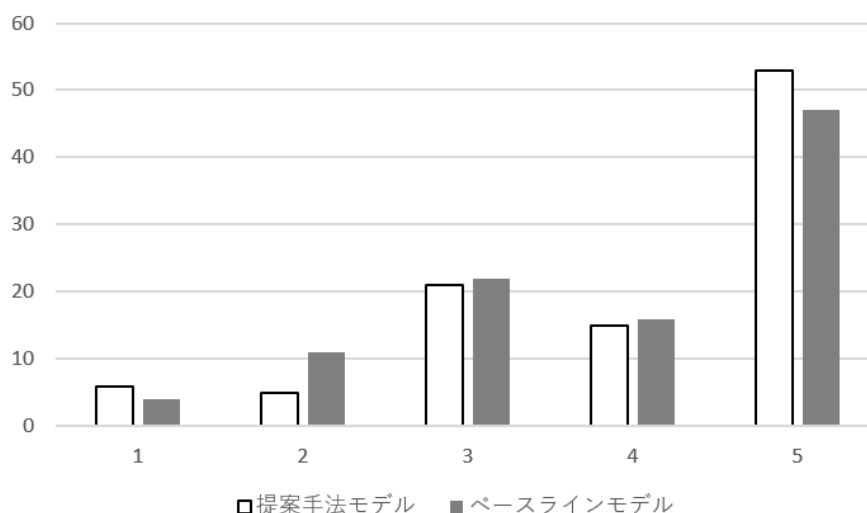


図 7.1: ランダムに生成した発話の評価

表 7.1 の 6~10 番の例がベースラインモデルで生成したものである。3 番と 8 番の例のように、キャラクタらしくない発話に大きな違いがある。提案手法モデルでは、人によってはキャラクタらしいと思うような発話が生成されたが、ベースラインモデルでは明らかに対象キャラクタとは異なる特徴の発話も生成された。

4 番、9 番のような、敬語など、対象の発話者のセリフとしてありえるものの、対象発話者の特徴を含まない発話はどちらのモデルからも同程度に生成された。これは、対象の発話者も場面や話している相手によっては敬語で話すことがあり、TAPT の学習データである対象発話者の発話にそのようなものが含まれていたためだと考える。

1 番、6 番のような日本語として明らかに不自然な発話は提案手法モデルの方が若干ではあるが増加していた。これは、T5 で変換された疑似的な学習データの中に不自然な日本語のものが含まれていたことが原因だと考える。

次に、テストデータから 100 発話抽出し、その前半部分を使ってそれぞれのモデルで文章生成を行った。生成に使用した部分が丁度発話の区切りにある場合、新たに文章が生成されないことがあったため、区切りの悪い部分までを生成に使用した。明らかにおかしい文法がないか、意味が通じる文章であるか、キャラクタらしさを含んでいるか、等の観点で、生成された文章を比較した。比較した結果を以下に分類した。

1. どちらのモデルもよい文章が生成されなかった
2. どちらのモデルもよい文章が生成された

表 7.1: ランダムに生成された文章

	モデル	生成された発話	評価
1	提案手法	なに？あんた今までのどの辺りにあんた一人も残ってなかったとでもしないかね。	(1)
2	提案手法	こんなところで眠ってちゃあ心配しないでね。	(2)
3	提案手法	ふしぎみたいだね。これ以上は見られないの。	(3)
4	提案手法	……え。そういうことですか？	(4)
5	提案手法	……ちょっと早すぎない？さすがにちょっと心配になっちゃうわ。	(5)
6	ベースライン	今でもあんたとあたしとの関係を気にしている？あんたは？あたしとあたしとっても大切な幼馴染が大切なのね。	(1)
7	ベースライン	……ねえ。その人たちってどれだけあの人たちのせいなんだろうね。	(2)
8	ベースライン	アルス。こんなとこまで来てもらってどうせ魔物に追いつかれるんじゃないかねえの？	(3)
9	ベースライン	これが伝説の精霊神さまと言われている方ですか？	(4)
10	ベースライン	ちょっと待ちなさいよ！これほどの者がこの国にいると思うわけないじゃない！	(5)

3. ベースモデルの方がよい文章を生成した

4. 提案手法モデルのほうがよい文章を生成した

結果を図 7.2 に示す。その結果、提案手法の方がよい文章を生成することが多かった。生成例を表 7.2 に示す。表 7.2 の 1 番の例のように、生成に使用した発話の中に学習データにはない表現があった場合特に不自然な発話を生成することが多かった。また 5 番のように、ベースラインモデルを使用した方はキャラクターのセリフ調ではなく、小説調の文章が生成されることもあった。このような例はベースの言語モデルである GPT-2 で生成を行った場合にも見られるため、GPT-2 を作成する際の学習データに含まれる小説調の文章が影響していると考えられる。

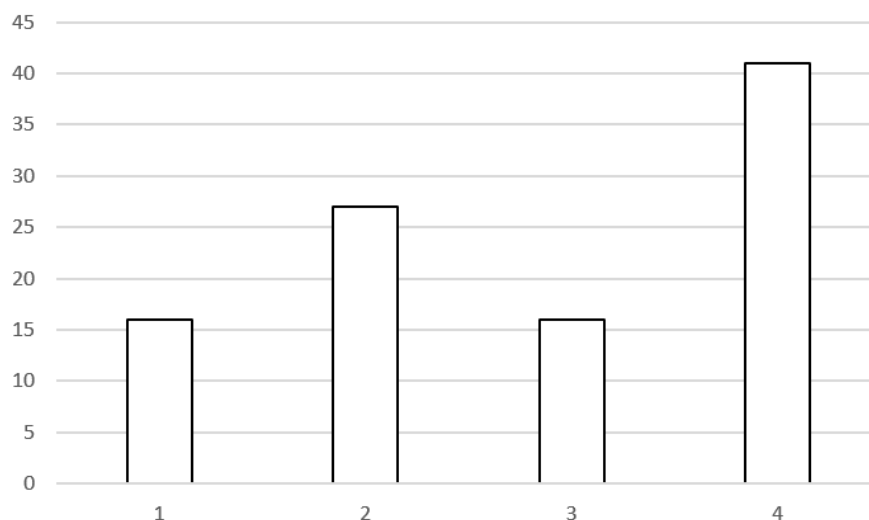


図 7.2: 発話の一部を利用して生成した発話の評価

7.2 今後の課題

本研究の提案手法により構築できた言語モデルを用いて、キャラクターらしい発話を生成することができたが、不自然な発話もいくつか生成されている。これは T5 モデルで発話文の変換を行った際に不自然な発話が生成了れたことが原因として考えられる。このため今後は、より自然な発話に変換するように T5 モデルを改善する必要があると考えている。そのための 1 つの方法としては、T5 モデルの学習データを増やすことがある。本研究で利用する T5 の学習データはキャラクターの発話を一般の発話へ修正することで作成されるが、この修正作業は筆者のみの主観で行われた。そのため量も十分ではなく、修正方法もある程度パターン化されており、学習データに現れないような一般の発話には対応できていない可能性がある。複数人で学習データを構築することで、これらの問題に対処できる可能性がある。

また現在のシステムでは、長い発話に不自然なものが多い。この点を考察するために、生成する最低文字数を指定し、テストデータの一部を利用して長い発話の生成を行った。その生成例を表 7.3 に示す。1 番～3 番は提案手法モデルを用いて生成した発話、4 番～6 番は 1 番～3 番と同じ発話の一部からベースラインモデルで生成した発話である。

表 7.3 の 1 番の例は、発話内の 1 文ずつを見れば自然だが、生成された発話全体を読むとまとまりのない内容になってしまっている。これは学習データに長い発話が少なく、

表 7.2: テストデータの一部を使用して生成した発話例

	モデル	生成に使用した 発話の一部	生成された発話	評価
1	提案手法	かたやあんと と	かたやあんと 同じ気持ちよ。	(1)
	ベースライン		かたやあんと違ってあたしはおぼえても ないし！	
2	提案手法	アルスと違っ てあたし	アルスと違ってあたし 魔力がないの。みんな こんな状態だし早く回復していきましょう よ！	(2)
	ベースライン		アルスと違ってあたし 体力ないみたいね。	
3	提案手法	まったくなん であたしたち	まったくなんであたしたち は今こんなに大 変なんだってのっ！	(3)
	ベースライン		まったくなんであたしたち の味方してんの よ。ねえアルス！あたしはさっきアイリーンの 魔力を感じたけど？なんで戦いに参加して ないのよ。	
4	提案手法	うちの船より	うちの船より も大きいわね。	(4)
	ベースライン		うちの船より もデカいんだ。	
5	提案手法	ガボを回復さ せた	ガボを回復させた 姿に神の復活は間違い無 いのね？	(4)
	ベースライン		ガボを回復させた 。そして魔力を使い果た したのか消えつつある大鷲の体を見つめる。 大鷲の傷が治っていったが……。	

文末の後に更に出力する文の予測が難しかったためであると考えられる。また T5 モデルを用いて変換を行う時に発話の意味が変わってしまったことで、不自然な流れの発話も生成されていたため、その影響を受けていることも考えられる。そのため、学習データに長い発話を含ませること、T5 の変換時に意味が変わってしまうことを防ぐことで改善することができる可能性があると考えている。ここで、同じ発話の一部からベースラインモデルで生成した 4 番と比較すると、1 番と同様に、4 番にも意味的なまとまりはない。また、4 番は小説の地の文のような文や、叫び声のような表現が生成されている。

表 7.3: 長い発話の生成例

	モデル	生成された発話
1	提案手法	あたしとしてはさっきのようなことは言わない方がいいかと思ってさ！今のはたしかにおどると悪いこと考えちゃうもんね。さあ行きましょアルス。ここは良いけどさ。
2	提案手法	そう！あたしこそ 気を付けるようがんばり屋さんなのね。みんなに会ったらきっとお礼を言ってくれるよ！！アルス！！！！！！！！！！あたしが生き残ってよかったよね！
3	提案手法	まあたしには 関係ないんじゃないかな！？……というかさっさと帰ってアイラに言いたい事でもあるのだろうなあ。アルス！アルスに頼まれてる物なんとかしてくれる？
4	ベースライン	あたしとしてはさっきまで感じていた暗さはもうなくなっていた。きつとずっと先になっていたんだわ。こんなに暗かったのにちゃんとここにいるのね。あーあーっ。あの人たちは何を考えてる人なの？
5	ベースライン	そう！あたしこそ 役に立てなくてごめんね！！やっぱりあたしの人生は決まった道だわ。あたしがここで倒れても関係ないよ！！！！！！！！！！
6	ベースライン	まあたしには ちょうどよかった。さっさとこの船に乗り込みましょ。アルス。早くアルスと一緒に船に乗って逃げましょ！アルス！！ わかり合えるってもんよ！！

また、表 7.3 の 2 番、3 番では、発話内で文が途切れた後には「アルス」という単語と「。」や「！」などの文末記号のみの文が多く生成された。これは、学習データ内に「アルス」という単語が多く含まれ、特に文の開始時、文末に出現する確率が他の単語より高くなってしまったことが原因であると考えられる。5.3 節でランダムに発話を生成したときも、提案手法モデルでは 100 発話中 10 発話が「アルス」ではじまっていた。この問題への対処としては、学習データに極端に多く出現する単語を他の単語に変換することを考えている。これによって長い発話でも自然になると予想している。また 2 番、3 番にみられる問題は 5 番、6 番にも同様に見られることから、どちらのモデルの学習に使用された対象発話者の発話の中の頻出単語を一部変えることが有効である可能性がある。

第 8 章

結論

本研究では、DAPT+TAPT の手法を利用して GPT-2 の fine-tuning を行うことで、特定の話者の発話の言語モデルを構築する方法を提案した。この際、対象の話者の発話をタスクのデータとみなし、対象の話者の登場する作品の他の登場人物の発話をドメインのデータに利用した。また、対象の話者の登場する作品の他の登場人物の発話はそのまま利用するだけでも有用であるが、これを T5 を用いて対象の話者風に変換することで対象の話者の発話の疑似データを作成し、DAPT のデータとして利用する方法を提案した。ベースの言語モデルには GPT-2 を利用した。その結果、T5 を用いて対象の話者と同作品のキャラクターの発話を対象の話者風に変換したものをドメインのデータセットとして学習し、その後対象の話者の発話をタスクのデータセットとして学習したものが最もよい結果となった。この結果から、提案手法が有効であることを示した。また、T5 の学習データとして TFIDF を用いてキャラクターの特徴を表していると考えられる発話を選ぶことで、より結果を改良することができた。

また構築した言語モデルを使って発話の生成を行い、生成される発話の自然さやキャラクターらしさについて考察した。その結果、提案手法により構築した言語モデルでは、不自然な発話を生成することも確認できた。これは構築した T5 モデルの性能に由来とすると考えられる。今後は T5 モデルの学習データを増強することで、T5 モデルの性能を高めたい。

謝辞

本研究に関わり手助けしてくださった方々に深くお礼申し上げます。新納研究室、古宮研究室の方々をはじめ、多くの方にお力添えいただき今日まで研究を進めることができました。新納先生、古宮先生には、研究のテーマ決めを手伝っていただき、論文の添削、研究会などの発表前に発表スライドの添削や発表練習に付き合っただけなど、非常にお世話になりました。また新納先生には実験を手伝っていただき、古宮先生にはジャーナルの書き方をご指導いただきました。心より感謝申し上げます。新納研究室の方々には、私が修士課程から研究室に所属したにも関わらず覚えてくださり、また気さくに話しかけてくださり大変嬉しかったです。ありがとうございました。古宮研究室の東京農工大組の方々には、ゼミや zoom で関わられたり、大勉強会の後にご意見をいただいたり、農工大の研究室に案内していただいたりしました。ありがとうございました。古宮研究室同期の方々には、学部時代と変わらず仲良くしていただきありがとうございました。

参考文献

- [1] 宮崎千明, 佐藤理史. 発話テキストへのキャラクター性付与のための音変化表現の分類. *自然言語処理*, Vol. 26, No. 2, pp. 407–440, 2019.
- [2] Mika Kishino and Kanako Komiya. Extracting linguistic speech patterns of japanese fictional characters using subword units. *International Journal on Natural Language Computing*, Vol. 11, No. 3, pp. 1–14, 2022.
- [3] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. *OpenAI blog*, Vol. 1, No. 8, p. 9, 2019.
- [4] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, Peter J Liu, et al. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, Vol. 21, No. 140, pp. 1–67, 2020.
- [5] Suchin Gururangan, Ana Marasović, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey, and Noah A Smith. Don't stop pretraining: Adapt language models to domains and tasks. *arXiv preprint arXiv:2004.10964*, 2020.
- [6] 勝又智, 坂田大直. CSJ を用いた日本語話し言葉 BERT の作成. *言語処理学会 第 27 回年次大会 発表論文集*, pp. 805–810, 2021.
- [7] François Mairesse and Marilyn Walker. *Personage: Personality generation for dialogue*, 2007.
- [8] Marilyn A. Walker, Grace I. Lin, and Jennifer E. Sawyer. An annotated corpus of film dialogue for learning and characterizing character style. In *the Proceedings of LREC 2012*, pp. 1373—1378, 2012.
- [9] 宮崎千明, 平野徹, 東中竜一郎, 牧野俊朗, 松尾義博, 佐藤理史. 話者のキャラクター性

- に寄与する言語表現の基礎的分析. 言語処理学会 第 20 回年次大会 発表論文集, pp. 232–235, 2014.
- [10] Chiaki Miyazaki, Toru Hirano, Ryuichiro Higashinaka, and Yoshihiro Matsuo. Towards an entertaining natural language generation system: Linguistic peculiarities of japanese fictional characters. In *the Proceedings of SIGDIAL 2016*, pp. 319–328, 2016.
- [11] 宮崎千明, 平野徹, 東中竜一郎, 牧野俊朗, 松尾義博, 佐藤理史. 文節機能部の確率的書き換えによる言語表現のキャラクター性変換. 人工知能学会論文誌, Vol. 31, No. 1, pp. DSF–515, 2016.
- [12] 奥井颯平, 中辻真. ポインタ生成機構を用いたキャラクター応答生成の検証. 第 34 回人工知能学会全国大会論文集, pp. 1I4–GS–2–01, 2020.
- [13] 岸野望叶, 古宮嘉那子. Sentencepiece を用いたキャラクターの特徴語抽出. 言語処理学会年次大会発表論文集, Vol. 2021, No. 27, pp. 8–12, 2021.
- [14] 岸野望叶, 古宮嘉那子, 新納浩幸ほか. T5 による特定キャラクター風発話への変換とその言語モデルの構築. 研究報告自然言語処理 (NL), Vol. 2022, No. 13, pp. 1–6, 2022.
- [15] Harsh Jhamtani, Varun Gangal, Eduard Hovy, and Eric Nyberg. Shakespearizing modern language using copy-enriched sequence to sequence models. In *Proceedings of the Workshop on Stylistic Variation*, pp. 10–19, Copenhagen, Denmark, September 2017. Association for Computational Linguistics.
- [16] Keith Carlson, Allen Riddell, and Daniel Rockmore. Evaluating prose style transfer with the bible. *Royal Society Open Science*, Vol. 5, No. 10, pp. 1–13, 2018.
- [17] 吉岡大貴, 安田裕介, 松永悟行, 大谷大和, 戸田智基. 内容語保存機構を備えた変分自己符号化器に基づくテキスト発話スタイル変換. 情報処理学会研究報告, Vol. 2022-NL-254, No. 8, pp. 1–6, 2022.
- [18] Jiwei Li, Michel Galley, Chris Brockett, Georgios P Spithourakis, Jianfeng Gao, and Bill Dolan. A persona-based neural conversation model. *arXiv preprint arXiv:1603.06155*, 2016.
- [19] Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan

- Liu, Jing Xu, Myle Ott, Kurt Shuster, Eric M Smith, et al. Recipes for building an open-domain chatbot. *arXiv preprint arXiv:2004.13637*, 2020.
- [20] Kikuo Maekawa. Corpus of spontaneous japanese: Its design and evaluation. In *the Proceedings of ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition 2003*, pp. 7–12, 2003.