

2022年度(令和4年度) 卒業論文

小説・エッセイの自動生成時に生じる矛盾  
の調査

指導教員

村田真樹  
村上仁一

鳥取大学工学部 電気情報系学科

自然言語処理研究室

B19T2129X 渡邊 武

## 概要

近年, GPT[1] を使用した自動文章生成の技術が, 様々な用途で利用されている. ここでは, その中でも小説・エッセイの自動作成に注目する.

GPT-3(Generative Pre-trained Transformer 3) とは, 2020 年 7 月に OpenAI が発表した高性能な言語モデルである. OpenAI は, Transformer と呼ばれる深層学習の手法を用いた言語モデル GPT(2018 年), GPT-2(2019 年) を発表しており, GPT-3 はその後継の言語モデルとなる. また, GPT-3 をモデルとした「AI のべりすと [2]」という日本語の自然言語処理プロセッサで, AI による文章・小説作成アプリケーションソフトウェアがある. 2021 年 7 月に web で公開されたアプリケーションソフトウェアの「AI のべりすと」は, 初めの数文や, 主人公などの重要な情報を入力するだけでもっともらしい続きの文章を生成する. しかし, 長く次々に文生成を続けると, 前方と矛盾が生じたり, 全然関係のない話題に発展する場合がある.

本論文では, 「AI のべりすと」を同一の入力文, 長い出力文, 様々な入力文の 3 つの条件で動かし文章を自動生成し, それらの文章内で出現した矛盾に注目して, それらの傾向や頻度の調査を行った.

実験結果は同一の入力文を用いた実験 (実験 1) で生成した 100 個の文章中 18 個に矛盾または不自然な文章が生成された. 実験 1 で矛盾を生成しなかった文章の更に後続の文を生成した, 長い出力文を用いた実験 (実験 2) では, 生成した 82 個の文章中 29 個に矛盾または不自然な文章が生成された. 様々な入力文を用いた実験 (実験 3) では, 生成した 60 個の文章中 10 個に矛盾または不自然な文章が生成された. 各実験での矛盾の比率は, 実験 1 は  $\frac{18}{100} = 0.18$ , 実験 2 は  $\frac{29}{82} = 0.35$ , 実験 3 で  $\frac{10}{60} = 0.17$  となった. また, 実験 2 において実験 1 で生成された矛盾の文章も含めた 計 100 個中での割合は  $\frac{47}{100} = 0.47$  と長い出力文では高い割合で矛盾または不自然な文章が生成された.

また, 本研究で得られた矛盾または不自然な文章を分析した結果, 大きく 3 つのタイプに分類することができた. 「お知らせ, 後書きなどの学習文のノイズ」, 「突然の時間, 場面, 状況の変化」, 「会話文がかみ合っていない」の 3 種類である.

さらに, 今回の実験で得られた矛盾の一部について, 共起頻度による半自動の検出が

可能かどうかの検証を行った. 結果は, 本実験で得られた矛盾データ 48 個中, 7 個は検出が可能であった.

# 目次

第1章	はじめに	1
第2章	従来の研究	3
2.1	GPT-3	3
2.2	AIのべりすと	3
第3章	問題点	4
第4章	提案手法	5
第5章	実験1: 同一の入力文で生成した文章中の矛盾調査	7
5.1	実験結果	7
5.2	出力例	8
第6章	実験2: 実験1で使用した文章の後続の文を更に生成し, 矛盾調査	11
6.1	実験結果	11
6.2	出力例	12
第7章	実験3: 様々な入力文で生成した文章中の矛盾調査	16
7.1	実験結果	16
7.2	出力例	18
第8章	考察	26
8.1	矛盾の出現傾向	26
8.2	共起頻度に基づく半自動の矛盾検出	27
8.2.1	概要	27
8.2.2	共起頻度	27
8.2.3	共起頻度の算出方法	27
8.2.4	二項検定	28

8.2.5	結果 . . . . .	29
8.2.6	出力例 . . . . .	30
<b>第9章</b>	<b>今後の課題</b>	<b>31</b>
9.1	小説の自動生成時に生じる矛盾の調査に関して . . . . .	31
9.2	共起頻度に基づく半自動の矛盾検出に関して . . . . .	31
<b>第10章</b>	<b>おわりに</b>	<b>32</b>

# 表 目 次

3.1 矛盾が生じた例 (AI のべりすと) . . . . .	4
5.1 矛盾の分類 . . . . .	7
5.2 矛盾のタイプ：学習文のノイズ（後書き） . . . . .	8
5.3 矛盾のタイプ：突然の時間、場面、状況の変化 . . . . .	9
5.4 矛盾のタイプ：会話文がかみ合っていない . . . . .	10
6.1 矛盾の分類 . . . . .	11
6.2 矛盾のタイプ：学習文のノイズ（お知らせ） . . . . .	12
6.3 矛盾のタイプ：突然の時間、場面、状況の変化 . . . . .	13
6.4 矛盾のタイプ：会話文がかみ合っていない . . . . .	14
6.5 矛盾のタイプ：その他 . . . . .	15
7.1 入力文 1 . . . . .	16
7.2 入力文 2 . . . . .	16
7.3 入力文 3 . . . . .	17
7.4 矛盾のタイプ：学習文のノイズ (入力文 1) . . . . .	18
7.5 矛盾のタイプ：突然の時間、場面、状況の変化 (入力文 1) . . . . .	19
7.6 矛盾のタイプ：学習文のノイズ (入力文 2) . . . . .	20
7.7 矛盾のタイプ：矛盾のタイプ：突然の時間、場面、状況の変化 (入力文 2)	21
7.8 矛盾のタイプ：会話文がかみ合っていない (入力文 2) . . . . .	22
7.9 矛盾のタイプ：学習文のノイズ (入力文 3) . . . . .	23
7.10 矛盾のタイプ：矛盾のタイプ：突然の時間、場面、状況の変化 (入力文 3)	24
7.11 矛盾のタイプ：会話文がかみ合っていない (入力文 3) . . . . .	25
8.1 検索結果と期待値 . . . . .	30
8.2 検索結果と期待値 . . . . .	30
8.3 検索結果と期待値 . . . . .	30

# 第1章 はじめに

近年, GPT を使用した自動文章生成の技術が, 様々な用途で利用されている. ここでは, その中でも小説・エッセイの自動作成に注目する. 2021年7月にwebで公開されたアプリケーションソフトウェアの「AIのべりすと」などは, 初めの数文や, 主人公などの重要な情報を入力するだけでもっともらしい続きの文章を生成する. しかし, 長く次々に文生成を続けると, 前方と矛盾が生じたり, 全然関係のない話題に発展する場合がある. そこで, これらの矛盾にどのような特徴や傾向があるかの調査を行う. また, 一部の矛盾に対しては共起頻度を用いて検出が可能かの検証も行う. 本研究の主な主張点を以下に示す.

- 本研究の目的は, AIが生成してしまう矛盾の原因, 特徴, 頻度の調査である. これらの調査は, 矛盾を生成しにくい学習の方法や, 矛盾の検知の方法を新たに開発するうえで有用性があると考えられる.
- 本研究で得られた矛盾または不自然な文章を分析した結果, 大きく3つのタイプに分類することができた. 「お知らせ, 後書きなどの学習文のノイズ」, 「突然の時間, 場面, 状況の変化」, 「会話文がかみ合っていない」の3種類である.
- 本研究で得られた矛盾または不自然な文章と出力文の長さから, 出力文の長さを長くすると矛盾の割合も高くなるという関係も確認できた. 矛盾が生成された具体的な割合は, 同一の入力文を用いて100個の文章を生成する実験(実験1)では  $\frac{18}{100} = 0.18$ , 実験1で矛盾を生成しなかった文章の更に後続の文を生成した, 長い出力文を用いた実験(実験2)では  $\frac{29}{82} = 0.35$  となった. また, 実験2において実験1で生成された矛盾の文章も含めた計100個中での割合は  $\frac{47}{100} = 0.47$  と長い出力文では高い割合で矛盾または不自然な文章が生成された.
- 今回の実験で得られた矛盾の一部について, 共起頻度による半自動の検出が可能かどうかの検証を行った. 結果は, 本実験で得られた矛盾データ48個中, 7個は検出が可能であった.

本論文の構成は以下のとおりである。第2章では、本研究に関連する研究としてどのような研究が行われてきたか記述する。第3章では、従来の研究での問題点を記述する。第4章では、本研究での提案手法を記述する。第5章～7章では、本研究の3種類の実験についての実験結果を記述する。第8章では、本研究の考察を記述する。第9章では、今後の課題を記述する。第10章では、本研究を通してのまとめを記述する。

## 第2章 従来の研究

### 2.1 GPT-3

GPT-3(Generative Pre-trained Transformer 3)[1] は, 2020年7月に OpenAI が発表した高性能な言語モデルである. OpenAI は, Transformer と呼ばれる深層学習の手法を用いた言語モデル GPT(2018年), GPT-2(2019年) を発表しており, GPT-3 はその後継の言語モデルとなる. GPT-3 は, 文章の生成, 文章の要約, 質問への回答, 翻訳などに活用することができる.

### 2.2 AIのべりすと

「AIのべりすと [2]」は, AIによる文章・小説作成アプリケーションソフトウェアである. GPT-3 をモデルとした, 日本語の自然言語処理プロセッサで, 日本のクリエイター Sta が個人で開発し, 2021年7月に公開した. Web に設定されたアプリにおいて数行の文章を書き込むと, その後の文章を AI が自動的に書き続けてくれる. 小説 AI とうたっているが, ユーザーの発想でさまざまな文章に対応している.

## 第3章 問題点

GPT は後続の文章生成を得意とし始めの数文は違和感が少ない文を生成するが、GPT で長く次々に文生成を続けると、前方と矛盾が生じたり、全然関係のない話題に発展する場合がある。表 3.1 に「AI のべりすと」を使用して文生成した際、矛盾が生じた例を示す。表 3.1 の下線部分に注目してほしい。「梅雨」と「入学式」という 2 単語が出てきてしまっていて、梅雨の時期の 6・7 月、入学式がある 4 月のように同時に起こりえない事象が起きてしまっている。つまり矛盾している文章である。そこで、本研究ではこのような矛盾の傾向や特徴の調査を行う。

表 3.1: 矛盾が生じた例 (AI のべりすと)

入力	出力
長く続いた梅雨は明け、今日の天気は快晴だ。	俺は今、朝早くからある場所へと向かっている。「ふう……ようやく着いた」バスに揺られること約 1 時間と 30 分。俺の目の前には立派な建物が建っている。ここが今日の目的地である『私立桜ヶ丘高等学校』である。今日から俺も晴れて高校生になるわけだが、あまり実感がない。まあそれも仕方ないと言えるだろう。何せ今までずっと勉強などしたことがなかったのだ。一応入試ではそこそこいい点数を取れたと思うのだが、果たしてどこまで通用するか……。そんなことを考えているうちに入学式が始まってしまった。

## 第4章 提案手法

本章では、提案手法について記述する。本研究の目的は、AIが生成してしまう矛盾の原因、特徴、頻度の調査である。異なる条件下で「AIのべりすと」を用いて文章を生成することで、先ほど述べた矛盾に関する情報が得られると考えた。具体的な実験手順を以下に示す。

**手順1** 「AIのべりすと」を用いて以下の3つの異なる条件を設けて実験を行い、文章を自動生成する。

[実験1] 同一の入力文を用いて生成した後続の10~20文程度の文章を生成する。

- ・入力文は全て「長く続いた梅雨は明け、今日の天気は快晴だ。」とする。
- ・合計100個の文章データを作成する。

[実験2] 1で生成した文章のうち、矛盾が生じなかったものの後続の文章を更に10~20文程度生成する。

[実験3] 入力文を変更し、実験1と同様に後続の文章を生成する。

- ・入力文1「聞こえるのは草花が風に揺られる音と小さな鳥のさえずり。私は森に来ている。」
- ・入力文2「「ハッピーニューイヤー！」 現在時刻は1月1日午前0時00分。」
- ・入力文3「あの日からちょうど1年の今日、」
- ・各20個の文章データを作成する。

**手順2** 手順1で作成した文章データの矛盾の有無を人手で評価する。

また、今回の矛盾の定義は「前に言ったこととあとに言ったことが一致しないこと。一般に、理屈として二つの事柄のつじつまが合わないこと。」とし、著者が生成された文章を読み判断した。

**手順 3** 手順 2 で得た矛盾をタイプごとに分類し表に整理して集計を行い, そこから AI が生成してしまう矛盾にどのような特徴や傾向があるかを調査する.

# 第5章 実験1:同一の入力文で生成した文章中の矛盾調査

## 5.1 実験結果

入力文は全て「長く続いた梅雨は明け、今日の天気は快晴だ。」とし、続きの文章を10～20文程度出力させる。これを100回行い評価した。結果は、100回中18回が矛盾または不自然な文章が生成された。表5.1に得られた矛盾を分類した。

表 5.1: 矛盾の分類

矛盾のタイプ	数
「お知らせ」、「後書き」などの学習文のノイズ	3
突然の時間、場面、状況の変化	8
会話文がかみ合っていない	7

## 5.2 出力例

実験1で矛盾または不自然な文章が生成された例を表5.2～表5.4に示す。表5.2は矛盾のタイプを「お知らせ」, 「後書き」などの学習文のノイズとした文章である。表5.3は矛盾のタイプを突然の時間、場面、状況の変化とした文章である。表5.4は矛盾のタイプを会話文がかみ合っていないとした文章である。

表 5.2: 矛盾のタイプ : 学習文のノイズ (後書き)

出力
このところずっと雨続きで憂鬱な気分だったから、 今日のような日差しが気持ちいい。 「お兄ちゃん！ 早く行こうよ！」 「ああ」 俺は妹に急かされながら玄関を出た。 ——◆ 後書き ◆—— いつもお読みいただきありがとうございます。 気がつけばPV3万を超えていました。 本当にうれしいです。 皆様の応援のおかげです。 これからもよろしくお願いします。 また、★評価やフォローをぜひお願いします。 モチベーションアップにつながるので、 感想も書いていただけると幸いです。

表5.2の文章を不自然な文章と判断した理由は、10行目でいきなり学習文のノイズと考えられる後書きが生成されてしまっているからである。

表 5.3: 矛盾のタイプ：突然の時間、場面、状況の変化

出力
空を見上げると雲一つない青空が広がっており、 今日も暑くなりそうだなと思いながら学校へ向かう。 「おはよう」 「あ、お兄ちゃん！ おはよう！」 いつもの通学路を歩いていると、後ろから声をかけられた。 振り返って見るとそこには妹の結衣がいた。 結衣とは <u>中学三年生</u> で受験生である妹だ。 俺は <u>高校二年</u> なので、歳的には俺の方が上だが、 <u>学年は同じ</u> なので敬語を使わずタメ口で話すようにしている。 まあ、たまに忘れてしまう時もあるが……

表 5.3 の文章を矛盾を含む文章と判断した理由は、前方の下線部で中学三年生と高校二年生という関係を示していて、その関係を後方の下線部で学年は同じとしてしまっているからである。

表 5.4: 矛盾のタイプ : 会話文がかみ合っていない

出力
「んー……いい天気……」 俺はカーテンを開けると背伸びをして体を解す。 そしてベッドから降りて部屋を出た。 俺の部屋は二階にあるので階段を下りる。 リビングに入ると既に父さんと母さんが居た。 「おはよう」 「あら？今日は早いわね？」 「いつもより 30 分くらい早く起きたんだ」 「珍しいな。何かあったのか？」 「まあちよつとね……」 そう言って俺は自分の席に座った。 すると台所の方から <u>姉さん</u> が出てくる。 「あ！悠真 <u>お兄ちゃん</u> おはよう！」 「おはよう美羽。今朝も元気だね」 「うんっ！私はいつでも元気だよ！」 彼女は美羽。俺の姉である。

表 5.4 の文章を矛盾を含む文章と判断した理由は、登場人物の美羽お姉ちゃんが下線部で悠真という弟に対してお兄ちゃんと言っているからである。

## 第6章 実験2: 実験1で使用した文章の後続の文を更に生成し, 矛盾調査

### 6.1 実験結果

実験1で生成した文章のうち, 矛盾が生じなかった82個について, 後続の文章を更に10~20文程度出力させる. これを評価した. 結果は, 82回中29回が矛盾または不自然な文章が生成された. 表6.1に得られた矛盾を分類した.

表 6.1: 矛盾の分類

矛盾のタイプ	数
「お知らせ」, 「後書き」などの学習文のノイズ	3
突然の時間、場面、状況の変化	11
会話文がかみ合っていない	14
その他	1

実験1と2を比較すると, 実験2の方が多くの矛盾または不自然な文章を生成している. この結果から, やはり文章が長くなればなるほど「AIのべりすと」は矛盾を含む文章を生成してしまう割合が増えてしまうと考えられる.

## 6.2 出力例

実験2で矛盾または不自然な文章が生成された例を表6.2～表6.5に示す。表6.2は矛盾のタイプを「お知らせ」, 「後書き」などの学習文のノイズとした文章である。表6.3は矛盾のタイプを突然の時間、場面、状況の変化とした文章である。表6.4は矛盾のタイプを会話文がかみ合っていないとした文章である。表6.5は矛盾のタイプをその他とした文章である。

表 6.2: 矛盾のタイプ : 学習文のノイズ (お知らせ)

出力
そんな天気にもまれた今日、俺と聖ちゃんは二人で出かけていた。 「おー、さすがに人多いな」 「そうだな……休日だからか？」 今俺達がいるのは遊園地であり、今日はデートをするためにここに来ていた。 この遊園地は最近できたばかりで、結構人気もあるらしい。 ・ ・ ・ まだ開園してそれほど時間は経っていないのだが、すでになんかの人が並んでいる。 「なんかもうちょっと時間かかりそうだね」 「そうだな。まあこういうのもいいんじゃないか？」 「確かにそれもそうかもね」 そんな会話をしながら待っていると、あっという間に俺達の番が来た。 「よし、行くか」 「楽しみだな」 楽しげな様子で聖ちゃんと一緒に、ジェットコースターに乗り込んだ。 ☆☆☆ 執筆の励みになりますので 面白かったら評価・フォローしていただくと嬉しいです!! 皆様の応援がモチベーションとなっております。 よろしくお祈りします! 「いや～、面白かったな！」 「うむ、久しぶりに乗ったがやはり楽しいものだな」

表6.2の文章を不自然な文章と判断した理由は、下から6行目でいきなり学習文のノイズと考えられるお知らせが生成されてしまっているからである。

表 6.3: 矛盾のタイプ：突然の時間、場面、状況の変化

出力
<p>「うわあ……！ 綺麗！」</p> <p>雲一つない青空を見上げながら、瑞希が感嘆の声を漏らす。</p> <p>「絶好の海水浴日和になったな」</p> <p>「そうね」</p> <p>俺と美智子が二人並んで浜辺に立ち、海を眺める。</p> <p>今日から待ちに待った夏休みだ。</p> <p>俺たち四人は揃って海水浴場へとやって来ていた。</p> <p>ちなみに、俺は水着を持っていない。</p> <p>去年も今年も、プールには行かなかったしな。</p> <p>・</p> <p>・</p> <p>・</p> <p>二人は目を合わせ、ニヤリと笑う。</p> <p>「ん？ どうしたんだ？」</p> <p>遅れてやってきた健一が尋ねる。</p> <p>「あ、あのさっ！ 私たちと一緒に入ろうよ！」</p> <p>「うん！ それが良いと思う！」</p> <p>美咲の提案に、すかさず乗っかり由香里が同意した。</p> <p>「え？ どうしてだよ？」</p> <p>困惑気味の健一に、</p> <p>「だって、男一人で女子三人の中に放り込まれるなんて可哀想だし」</p> <p>と由香里がもっともらしい理由を述べる。</p> <p>「いや、でも俺が女湯に入ったら犯罪になるんじゃないのか？」</p> <p>「大丈夫！ その辺はうまくやるから！」</p> <p>「そ、そうなのか？」</p> <p>「もちろん！」</p>

表 6.3 の文章を矛盾を含む文章と判断した理由は、最初は海水浴場に来ていた登場人物達が、下線部の「女湯」から、突然温泉に場所が変わっているからである。

表 6.4: 矛盾のタイプ：会話文がかみ合っていない

出力
<p>そんな爽やかな空の下を、俺は一人で歩く。  「……なんでこんなことに」  そう呟きながら、俺の足は自然と早くなる。  ・  ・  ・  「ん？ どうした？」  「あのさ……もし良かったら一緒に出ない？」  「一緒になって……どの競技に出るつもりなんだ？」  「その……二人三脚とか……」  頬を赤らめながら言う彼女を見て、思わず苦笑してしまう。  「おいおい……二人三脚なんて恥ずかしいだろ。絶対無理だって」  「どうしてよお～」  「どうしてもこうしてもない。そもそも俺達付き合っていないし」  「むう……」  すると、今度は拗ねるように唇を尖らせる。忙しい奴だな……。  「それにしても、どうして急に二人三脚なんか出たいと思ったんだ？」  <u>「実はさっき女子同士で話してた時に、誰と一緒に出るのか</u>  <u>決めてないっていう子がいたから誘ったの。そしたらその子が、</u>  <u>『それいいかも！』って言ってくれたんだよね</u>  「へえ……意外といいところあるじゃん」  「えへへ♪」  素直に褒めると、彼女は照れくさそうにはにかんだ。  そして、再び口を開く。  「だからさ、二人で一緒に頑張ろうよ！絶対に楽しいと思うし！」  そこまで言われてしまうと断りづらい。</p>

表 6.4 の文章を矛盾を含む文章と判断した理由は、下線部は他の人物と二人三脚に出る約束をしたことを主張していて、どうして二人三脚に出たいと思ったかという疑問に対する返答として不適切だからである。

表 6.5: 矛盾のタイプ：その他

出力
<p>「さてと……今日も頑張りますか」 俺は自分自身に気合いを入れると、仕事着に着替え始めるのだった。</p> <p>◆</p> <p>『——おはようございます』 開店して一時間ほどが経った頃だろうか。 俺がレジカウンターで品出しをしていると、 自動ドアから来店を告げるチャイムが鳴ると共にそんな声が聞こえてきた。 「いらっしやいませー！」 反射的に挨拶を返すと、入口の方へと視線を向ける。 するとそこにはスーツ姿の女性の姿があった。 その女性は店内を見回すようにして歩いてくると、 真っ直ぐに菓子コーナーへと向かう。</p> <p>・ ・ ・</p> <p>男性は意外にも落ち着いた声で話しかけてくる。 「君がこの店の店長かい？」 「え？あ、はい……」 戸惑いながらも返事をする。すると彼はポケットから一枚の名刺を取り出した。 「私はこういう者なんだが」 差し出された名刺には彼の名前と思われるものが書かれていた。 しかしそこには肩書きなどはなく、『<u>弁護士</u>』の<u>二文字</u>だけが印刷されている。</p>

表 6.5 の文章を矛盾を含む文章と判断した理由は、弁護士は三文字であるのに下線部では二文字としてしまっているからである。

# 第7章 実験3:様々な入力文で生成した文章中の矛盾調査

## 7.1 実験結果

入力文1は「聞こえるのは草花が風に揺られる音と小さな鳥のさえずり。私は森に来ている。」, 入力文2は「「ハッピーニューイヤー！」 現在時刻は1月1日午前0時00分。」, 入力文3は「あの日からちょうど1年の今日、」とし, 続きの文章を10~20文程度出力させる。これを各20回行い評価した。結果は, 入力文1では20回中3回, 入力文2では20回中3回が, 入力文3では20回中4回矛盾または不自然な文章が生成された。以下にそれぞれ得られた矛盾を分類した。

表 7.1: 入力文1

矛盾のタイプ	数
「お知らせ」, 「後書き」などの学習文のノイズ	1
突然の時間、場面、状況の変化	2
会話文がかみ合っていない	0

表 7.2: 入力文2

矛盾のタイプ	数
「お知らせ」, 「後書き」などの学習文のノイズ	1
突然の時間、場面、状況の変化	1
会話文がかみ合っていない	1

表 7.3: 入力文 3

矛盾のタイプ	数
「お知らせ」、「後書き」などの学習文のノイズ	1
突然の時間、場面、状況の変化	1
会話文がかみ合っていない	2

それぞれの入力文での矛盾または不自然な文章が生成された回数は入力文 1 で 3 回, 入力文 2 で 3 回, 入力文 3 で 4 回となり, 大きな差は見られなかった. また, 実験 1 と実験 3 での矛盾の比率を比較しても, 実験 1 で 0.18, 実験 3 で 0.17 となり, ここでも大きな差は見られなかった. この結果から, 極端に特殊なものでなく, 基本的な小説の冒頭文を使用する場合, それらの違いによる矛盾への影響は小さいと考えられる.

## 7.2 出力例

実験3で矛盾または不自然な文章が生成された例を表7.4～表7.11に示す。表7.4は入力文1で矛盾のタイプを「お知らせ」, 「後書き」などの学習文のノイズとした文章である。表7.5は入力文1で矛盾のタイプを突然の時間、場面、状況の変化とした文章である。表7.6は入力文2で矛盾のタイプを「お知らせ」, 「後書き」などの学習文のノイズとした文章である。表7.7は入力文2で矛盾のタイプを突然の時間、場面、状況の変化とした文章である。表7.8は入力文2で矛盾のタイプを会話文がかみ合っていないとした文章である。表7.9は入力文3で矛盾のタイプを「お知らせ」, 「後書き」などの学習文のノイズとした文章である。表7.10は入力文3で矛盾のタイプを突然の時間、場面、状況の変化とした文章である。表7.11は入力文3で矛盾のタイプを会話文がかみ合っていないとした文章である。

表 7.4: 矛盾のタイプ : 学習文のノイズ (入力文1)

入力	出力
<p>聞こえるのは草花が風に揺られる音と小さな鳥のさえずり。私は森に来ている。</p>	<p>「んー、今日もいい天気だ」 腕を伸ばして深呼吸をする。空気がおいしい。こんなにいい天気なのに家に閉じこもってるなんて勿体ないよね。 「よしっ！ 今日頑張ろう！」 頬を軽く叩いて気合いを入れる。 そうして私は森の中に入っていく。</p> <hr/> <p>&lt;コメント : 2件&gt; 『ハルちゃんおはよう！』 『ハル様今日も可愛い』 『頑張ってくださいね！』</p> <hr/> <p>「ふう……」 木漏れ日を浴びながら泉のほとりで休憩する。ここはとても落ち着く場所だ。澄んだ水がとても綺麗で見ただけで心が落ち着く。</p>

表7.4の文章を不自然な文章と判断した理由は、8行目でいきなり学習文のノイズと考えられるコメントが生成されてしまっているからである。

表 7.5: 矛盾のタイプ：突然の時間、場面、状況の変化(入力文1)

入力	出力
<p>聞こえるのは草花が風に揺られる音と小さな鳥のさえずり。私は森に来ている。</p>	<p>「んー……」寝ぼけ眼をこすりながら起き上がる。          まだ少し眠いけど、二度寝するほどではない。          今日も今日とて森で薬草探した。          昨日のうちに下見しておいたから、そのへんはぬかりない。          まずは顔を洗おう。川まで行って  <u>水浴びでもしようかと思ったけれど、やめた。</u>  <u>今は夏だけど、この季節の水浴びはまだちょっと寒い。</u>          風邪を引いてしまうかもしれないしね。          それに、こんなに気持ちの良い朝なんだから。          そんなわけで、小川のすぐ側にある大きな木の下へ。  <u>ここなら水浴びしても大丈夫そうかな？</u>          ああ、でも着替えがないなあ……まあいいや、          とりあえずやってみよう！          バシヤッ「つめた!?」          勢いよく水を被ったせいでびっくりしてしまった。          もう、冷たいよ……。          幸いなことに、服はそれほど濡れなかったみたい。</p>

表 7.5 の文章を矛盾を含む文章と判断した理由は、下線部で登場人物の水浴びに対する考え方が唐突に変化し、拒んでいた水浴びをしているからである。

表 7.6: 矛盾のタイプ：学習文のノイズ(入力文2)

入力	出力
<p>「Happyニュー イヤー！」現在時 刻は1月1日午前 0時00分。</p>	<p>—————</p> <p><b>【あしがき】</b>            新年、あけましておめでとうございます！            今年もよろしくお願いいたします！            というわけで、あけましておめでとうございます！            そして……今年の投稿はこれが最後になります……。            来年からまた毎日更新再開予定ですので、            応援していただけると嬉しいです。            それでは良いお年を～！            『あけましておめでとうございます』            「あけましておめでとうございます」            「あけおめー！」            「明けましておめでとうございます……」            『今年もよろしくお願いいたします！』            というわけで、元旦。            皆で初詣に行ってきました。</p>

表 7.6 の文章を不自然な文章と判断した理由は、冒頭からいきなり学習文のノイズと考えられるあしがきが生成されてしまっているからである。

表 7.7: 矛盾のタイプ：矛盾のタイプ：突然の時間、場面、状況の変化 (入力文 2)

入力	出力
<p>「ハッピーニューイヤー！」現在時刻は<u>1月1日</u>午前0時00分。</p>	<p>俺は、新年を迎えると同時に、そう叫んだ。  <u>今日は12月31日大晦日だ。</u>          そしてここは俺の家のリビングであり、          こたつに入りながらテレビを見ている状態だ。          こたつの上にあるみかんを食べながら、俺は言う。          「いやー、今年ももう終わりかあ……」          ちなみに今は冬休みである。          去年までは家でゴロゴロしながら、          年末の歌番組を見て過ごしていたのだが、          今年からは違う。なぜなら——          「なに言ってるのよ。まだ始まったばかりじゃない」          俺の言葉に対し、隣で一緒に座っている少女がそう言った。</p>

表 7.7 の文章を矛盾を含む文章と判断した理由は、入力文で1月1日と記載しているのに下線部で12月31日と記載されているからである。

表 7.8: 矛盾のタイプ：会話文がかみ合っていない(入力文2)

入力	出力
<p>「ハッピーニューイヤー！」現在時刻は1月1日午前0時00分。</p>	<p>「あけましておめでとうございます」  「お、おう……おめでとさん……」  年が明けた。つまり、今日は元旦だ。  この瞬間から、俺の新たな一年が始まる。  「どうしたの？ 元気ないよ？」  「いや、ちょっと寝不足なだけだ」  「昨日は何時に寝たんですか？」  <u>「日付が変わる前にはベッドに入ったぞ」</u>  「なら大丈夫ですね。ほら、早く起きて！ 初詣に行こ！」  俺はまだ眠気まなこを擦りながら体を起こす。  「ふぁ～あ……」  大きな欠伸をしながらリビングに行くと、  既に2人とも準備万端だった。</p>

表 7.8 の文章を矛盾を含む文章と判断した理由は、1月1日午前0時00分時点の会話文として下線部は不自然だからである。

表 7.9: 矛盾のタイプ：学習文のノイズ(入力文3)

入力	出力
<p>あの日からちよ うど1年の今日、</p>	<p>俺は1人の少女に告白する。 「好きです」と。 ————— &lt;後書き&gt; 更新遅れて申し訳ございません！ 次話は必ず明日に投稿しますので お楽しみにしてください m ( _ _ ) m また、いつもこの作品を読んでもらっている 読者様方には本当に感謝しております！！ 今後とも応援よろしく願いいたします m ( _ _ ) m 「好きです」 俺がそう告げると目の前にいる彼女は 顔を真っ赤にして俯いてしまった。 彼女の名前は橘花玲奈。 1年前の春に初めて会った時からずっと好きだった女の子だ。</p>

表 7.9 の文章を不自然な文章と判断した理由は、3 行目でいきなり学習文のノイズと考えられる後書きが生成されてしまっているからである。

表 7.10: 矛盾のタイプ : 矛盾のタイプ : 突然の時間、場面、状況の変化 (入力文 3)

入力	出力
<p>あの日からちょうど1年の今日、</p>	<p>僕は彼女に告白する。                  彼女——黒井真雪は僕と同じ高校に通う同級生で、                  そして僕の大切な<u>幼馴染</u>だ。                  1年前の春、桜の花びらが舞う季節に、  <u>彼女は突然転校してきた。</u>                  彼女の両親の仕事の都合で、                  この町に引っ越してくることになったらしい。                  その頃の僕は、まだ彼女と知り合って間もなかったけれど……  <u>それでも彼女はとても綺麗な人だとすぐにわかった。</u>                  艶のある長い髪に、整った顔立ち。</p>

表 7.10 の文章を矛盾を含む文章と判断した理由は、最初の下線部で彼女と幼馴染とあるが、後続の下線部では彼女は一年前に転校してきて知り合ったと記載されているからである。

表 7.11: 矛盾のタイプ：会話文がかみ合っていない(入力文3)

入力	出力
<p>あの日からちょうど1年の今日、</p>	<p>私は彼に会いに行こうと決めていた。  「……」私は手元のチケットを見つめる。  このチケットは、とある劇団の公演のもので、  彼が主演を務める舞台だ。  彼は私を元気づけようとしてくれているのか、  「今度の舞台も見に来て欲しい」と言ってくれた。  そして、私が彼のことを好き  だということを知っていたからか、  <u>「もしよかったら一緒に観劇に行きませんか？」</u>  と誘ってくれたのだ。  私は少し悩んだものの、結局行くことにした。  彼に会えるかもしれないという  期待はもちろんあったけれど、  それ以上に、このままではいけないと思ったからだ。  この機会に、自分の気持ちを  整理して、きちんと向き合いたい。</p>

表 7.11 の文章を矛盾を含む文章と判断した理由は、下線部は物語の流れも含め、劇団の主演を務めている「彼」という登場人物の発言としては不適當だからである。

## 第8章 考察

### 8.1 矛盾の出現傾向

本節では第5～7章の実験で得られたデータより考えられる、AIが生成してしまう矛盾の特徴や傾向についてまとめる。

まず、実験1と2の矛盾の比率を比較すると、実験1では  $\frac{18}{100} = 0.18$ 、実験2は  $\frac{29}{82} = 0.35$  となり、実験2の方が約2倍の矛盾または不自然な文章を生成している。また、また、実験2において実験1で生成された矛盾の文章も含めた計100個中での割合は  $\frac{47}{100} = 0.47$  と長い出力文では5割に近い割合で矛盾または不自然な文章が生成された。この結果から、やはり文章が長くなればなるほど「AIのべりすと」は矛盾の含む文章を生成してしまう頻度が増えてしまうと考えられる。

次に、実験3ではそれぞれの入力文での矛盾または不自然な文章が生成された回数は入力文1で3回、入力文2で3回、入力文3で4回となり、大きな差は見られなかった。また、実験1と実験3での矛盾の比率を比較しても、実験1で0.18、実験3で0.17となり、ここでも大きな差は見られなかった。この結果から、極端に特殊なものでなく、基本的な小説の冒頭文を使用する場合、それらの違いによる矛盾への影響は小さいと考えられる。

## 8.2 共起頻度に基づく半自動の矛盾検出

本節では第5～7章の実験で得られた一部の矛盾の文章に対して、Google 検索ヒット件数を用いて共起頻度による矛盾の検出が可能かどうかの検証を行った。

### 8.2.1 概要

村田らの論文、頻度に基づく正の例からの負の例の予測 [3][4] を参考に共起頻度による矛盾検出を行う。第5～7章の実験で得られた一部の矛盾の文章には、一般的に共起されにくいと考えられる2単語または2フレーズが確認できた。そこで、Google 検索で対象の2単語または2フレーズを含む文章がどの程度存在しているのかを調べ、その数値を出現数の期待値と比較して、期待値より十分に小さい値である場合、その2単語または2フレーズは共起頻度が低いといえる。このようにして共起頻度を算出し、共起頻度が低い2単語または2フレーズが存在している場合、その文章は矛盾を含む文として検出できるとし、その精度の検証を行った。簡潔に説明すると、本来一緒に出現することが少ない2単語または2フレーズが出現していた場合、矛盾を含む文章と判断するという方法である。

### 8.2.2 共起頻度

共起頻度とは、任意の文書や文において、ある単語とある単語が同時に出現する頻度のことである。例としては、「囲碁」と「打つ」は共起すると言えるが、「将棋」と「打つ」は共起するとは言えない。

### 8.2.3 共起頻度の算出方法

第5～7章の実験で得られた一部の矛盾に対して、以下のような手順で共起頻度を調べた。

**手順1** 実験で生成された矛盾を含む文章から共起頻度が低いと考えられる2フレーズを人手により抜き出す。

**手順2** 手順1で抜き出したそれぞれのフレーズを含むページ数を google で検索する。

**手順 3** 抜き出した 2 フレーズをともに含むページ数を google 検索機能の AROUND(X) を利用して検索する.

google 検索機能の AROUND(X)

- 2 つの語句を含むページを検索する.

- 例: "A" AROUND(1) "B" で検索をすると, A, B を含み

互いの間にある単語数や文字数が 1 以内であるページが表示される.

**手順 4** 手順 2, 3 で得られた検索ヒット数を以下の算出式に利用してフレーズ間の共起頻度を調べる.

共起頻度の算出式を以下に示す.

google 検索において, フレーズ A, B について, それぞれの google 検索ヒット数を  $n_A, n_B$  とする. google 検索全体のヒット数が  $N$  (本実験では「”の”」の検索ヒット数: 約 3,170,000,000 件を  $N$  とする.) の場合, A, B がともに出現する検索ヒット数の期待値は  $\frac{n_A n_B}{N}$  となる. ここで, "A" AROUND(X) "B" で検索を行った際のヒット数  $n$  と比較する.

$$\frac{n_A n_B}{N} \gg n \quad (8.1)$$

式 (8.1) の関係になったとき, A, B の共起頻度は低いと考えられる. また, 式 (8.1) の正確な判断には二項検定を用いる.

## 8.2.4 二項検定

二項検定 [5] は, 勝ちと負けや表と裏といった相反する 2 つのペアの差に符号 + と - を割り当てる統計的検定として用いられる, その符号の出現率が二項確率に従うかどうかを検定する手法である. 以下, 二項検定の流れを示す.

**手順 1** 仮説を設定

帰無仮説 (+ の出現確率が有意に  $x$  以下でない) と対立仮説 (+ の出現確率が有意に  $x$  以下) を仮定し, 帰無仮説は成立すれば, + の符号がでる確率は  $x$  以下とは限らない. 対立仮説は成立すれば, + の出現確率は  $x$  以下となる.

**手順 2** 有意水準の設定

有意水準とは, 検定において帰無仮説が棄却される確率のことで, 5% に設定する.

### 手順 3 p 値を求める

帰無仮説が正しいという前提において、極端に偏った検定量が得られる p 値を求める。

### 手順 4 帰無仮説を棄却

帰無仮説が正しいと仮定して計算した p 値が有意水準以下であれば、帰無仮説は棄却され、対立仮説が証明される。従って、統計的に有意な差が存在すると言える。

## 8.2.5 結果

第 5~7 章の実験で得られた矛盾に対して、共起頻度による検出が可能かどうかの評価を行った。結果は「お知らせ」、「後書き」などの学習文のノイズによる文章を除いた 48 個の矛盾のうち、25 個は共起頻度による検出が可能であると考え、それらの矛盾に対して上記の方法で共起頻度を調査した。結果は 25 個のうち 11 個は共起回数が期待値を下回った。

また、共起回数が期待値を下回った 11 個の事象に対しては式 (8.1) を十分に満たしているのかを検証するため、二項検定を行い評価した。二項検定の結果は、共起回数が期待値を下回った 11 個の事象のうち 7 個は有意差あり、すなわち式 (8.1) を十分に満たしていると判断された。従って、今回の手法で検出可能であったものは  $\frac{7}{48} = 0.146$  で低い精度となった。

本節では、共起頻度による矛盾の検出を試みたが、この他にも矛盾と似通った考え方である含意を用いた言語モデルの性能を向上させる研究 [6][7] などもある。含意を用いた研究は A という文章から B という文章が推論出来るかどうかの関係に注目したものである。このように、様々なアプローチからの研究が存在する。

## 8.2.6 出力例

以下に実際の検索文, 検索ヒット数, 期待値のデータ例を表 8.1, 表 8.2, 表 8.3 に示す. 表 8.1 は表 5.4 の文章から「姉さん」と「お兄ちゃんおはよう」の 2 フレーズを抜き出し, 2 フレーズを含む文章の検索ヒット数が期待値より十分に小さい値となり二項検定の結果も, 有意差があった. つまり, 検出が可能であった例である. 表 8.2 は表 5.4 の文章と同様に姉が弟に対してお兄ちゃんと言っている文章から「姉貴」と「お兄ちゃん! 遅いよ」の 2 フレーズを抜き出し, 2 フレーズを含む文章の検索ヒット数が期待値より小さい値となったが二項検定の結果, 有意差はなかった. つまり, 検出が上手くできなかった例である. 表 8.3 は表 3.1 の文章から「梅雨」と「入学式」の 2 フレーズを抜き出し, 2 フレーズを含む文章の検索ヒット数が期待値より大きな値となった. つまり, 検出が上手くできなかった例である.

表 8.1: 検索結果と期待値

検索文	検索ヒット数 (件)
“ 姉さん ”	約 155,000,000
“ お兄ちゃんおはよう ”	約 3,270
“ A ” AROUND(1) “ B ”	約 0
期待値	159

表 8.2: 検索結果と期待値

検索文	検索ヒット数 (件)
“ 姉貴 ”	約 7,070,000
“ お兄ちゃん! 遅いよ ”	約 1,180
“ A ” AROUND(1) “ B ”	約 0
期待値	3

表 8.3: 検索結果と期待値

検索文	検索ヒット数 (件)
“ 梅雨 ”	約 29,200,000
“ 入学式 ”	約 18,900,000
“ A ” AROUND(1) “ B ”	約 2,580,000
期待値	174,095

## 第9章 今後の課題

### 9.1 小説の自動生成時に生じる矛盾の調査に関して

本研究で行った小説の自動生成時に生じる矛盾の調査に関して今後の課題を記述する。まず、本研究では最大でも40文の出力文で調査を行ったが、実際の小説はもっと長い文章を扱うため出力文を更に増やした時の矛盾調査もしていく必要がある。また、入力文の種類も4種類と少ないため入力文を増やして調査も行うべきである。

矛盾の分類に関しては、本研究では「お知らせ、後書きなどの学習文のノイズ」、「突如の時間、場面、状況の変化」、「会話文がかみ合っていない」の大きく分けて3種類としたが、細分類することで、新たに矛盾の特徴が獲得できる可能性があると考えている。

### 9.2 共起頻度に基づく半自動の矛盾検出に関して

8.2節の共起頻度に基づく半自動の矛盾検出に関して今後の課題を記述する。今回の結果は $\frac{7}{48} = 0.146$ と精度が低かった。精度が低かった一番の原因は、今回の手法は人手による2フレーズの抽出であり、抽出数に限りがあったからと考えられる。今後は機械による自動化での抽出を行っていく必要がある。しかし、今回の手法は2フレーズ抽出の部分の影響が大きく結果に作用するもので、このまま機械による自動化で全てのフレーズに関して共起頻度を調べる処理を行った場合、人手では見つけられなかった共起頻度の低い2フレーズの抽出による矛盾の検出の可能性もあるが、逆に矛盾を含まない文章中での誤検知などの可能性もあり、必ずしも精度が向上するとはいえない。従って、新たな条件を加えることで誤検知を減らし共起頻度による矛盾の検出の精度向上を図っていく必要がある。

## 第10章 おわりに

本研究ではAIによる小説の自動文章生成時に生じる矛盾に注目して、それらの傾向や頻度の調査、また、一部の矛盾に対しては共起頻度による検出が可能かどうかの検証を行った。

本研究の調査結果は、同一の入力文を用いた実験（実験1）では生成した100個の文章中18個に矛盾または不自然な文章が生成された。長い出力文を用いた実験（実験2）では、生成した82個の文章中29個に矛盾または不自然な文章が生成された。様々な入力文を用いた実験（実験3）では、生成した60個の文章中10個に矛盾または不自然な文章が生成された。各実験での矛盾の比率は、実験1は $\frac{18}{100} = 0.18$ 、実験2は $\frac{29}{82} = 0.35$ 、実験3で $\frac{10}{60} = 0.17$ となった。また、実験2において実験1で生成された矛盾の文章も含めた計100個中での割合は $\frac{47}{100} = 0.47$ と長い出力文では高い割合で矛盾または不自然な文章が生成された。この結果から、AIが生成してしまう矛盾は、文章が長くなればなるほどその頻度は増えるといえる。

その他、様々な入力文を用いた実験（実験3）において、それぞれの入力文での矛盾または不自然な文章が生成された回数は入力文1で3回、入力文2で3回、入力文3で4回となり、大きな差は見られなかった。また、実験1と実験3での矛盾の比率を比較しても、実験1で0.18、実験3で0.17と、ここでも大きな差は見られなかった。この結果から、基本的な小説の冒頭文を入力文として使用する場合、それらの違いによる矛盾への影響は小さいといえる。

また、本研究で得られた矛盾または不自然な文章を分析した結果、大きく3つのタイプに分類することができた。「お知らせ、後書きなどの学習文のノイズ」、「突然の時間、場面、状況の変化」、「会話文がかみ合っていない」の3種類である。

さらに、8.2節の共起頻度に基づく半自動の矛盾検出では、 $\frac{7}{48} = 0.146$ と精度は低いものの一部の矛盾に対しては検出することができた。しかし、今回の手法は人手による2フレーズ抽出の部分の影響が大きく結果に作用するもので、このまま機械による自動化で全てのフレーズに関して共起頻度を調べる処理を行った場合、人手では見つけられなかった共起頻度の低い2フレーズの抽出による矛盾の検出の可能性もあるが、逆に矛盾を含ま

ない文章中での誤検知などの可能性もあり,必ずしも精度が向上するとはいえない.従って,新たな条件を加えるなどして誤検知を減らし共起頻度による矛盾の検出の精度向上を目指していきたい.

# 謝辞

本研究を進めるに当たり,終始に渡り研究の進め方や本論文の書き方など,細部にわたる御指導を頂きました,鳥取大学工学部電気情報系学科自然言語処理研究室の村田真樹教授に心から御礼申し上げます.また,本研究を進めるにあたり,御指導,御助言を頂きました,村上仁一准教授に心から御礼申し上げます.その他様々な場面で御助言を頂いた自然言語処理研究室の皆様へ感謝の意を表します.

## 参考文献

- [1] GPT-3(Generative Pre-trained Transformer 3)—用語解説 [https://www.nri.com/jp/knowledge/glossary/1st/alphabet/gpt\\_3](https://www.nri.com/jp/knowledge/glossary/1st/alphabet/gpt_3)
- [2] AI のべりすと <https://ai-novel.com/>
- [3] 村田真樹, 井佐原均, 頻度に基づく正の例からの負の例の予測—日本語表記の誤り検出と外の関係の文の自動抽出—
- [4] MasakiMurata, HitoshiIsahara, Automatic Detection of Mis-Spelled Japanese Expressions Using a New Method for Automatic Extraction of Negative Examples Based on Positive Examples
- [5] TUNGPUJUI, 機械学習と統計的検定を用いた知見獲得とその評価
- [6] 芝山直希, 古宮嘉那子, 新納浩幸, 日本語 PromptBERT におけるプロンプトベース対照学習の有効性とプロンプトの性質の調査
- [7] 小谷通隆, 柴田知秀, 中田貴之, 黒橋禎夫, 日本語 Textual Entailment のデータ構築と自動獲得した類義表現に基づく推論関係の認識
- [8] 青空文庫 <https://www.aozora.gr.jp/>