

2024年度 卒業論文

タイポグラフィを用いた生成 AI に関する分析

大東文化大学
経営学部経営学科

21161301 山田晴大

21161309 佐藤輝

目次

1. はじめに	3
1.1. 研究概要	3
1.2. 研究意義	3
1.3. 研究背景	3
2. 研究方法	4
2.1. 分析方法	4
2.2. プロンプトの作成	5
3. 事例研究	5
3.1. ChatGPT の特徴	5
3.2. タイポが与える単語修正や文章理解への影響	7
3.3. 単語（名詞）のみのケース	10
3.4. 変換確率「0.9」の文章で検証	13
3.5. その他の検証	15
3.6. まとめ	18
4. オリジナル文章を用いた評価実験	18
4.1. 研究の目的	18
4.2. 実験データの作成	18
4.3. 実験方法	19
4.4. 評価指標	20
5. 実験結果	20
5.1. 用意した文章と問題	20
5.2. 実験結果及び考察	23
5.3. まとめ	24
6. おわりに	25
6.1. 今後の課題	25
謝辞	26
参考文献	26
付録	27

1. はじめに

1.1. 研究概要

本研究は、タイポグリセミア (typoglycemia) が生成 AI (ChatGPT など) の出力にどのような影響を与えるのか調べたものである。

タイポグリセミアとは、人間が文章を読むとき、単語の最初と最後の文字があっければ、単語内の文字の順番を入れ替えても文章の意味を理解できる現象のことをいう。この現象は、「人間が文章の内容を推測し、内容を補完しながら読んでいる」という特徴を利用したものである。そのため、人間は多少の誤字 (typographical error、以下「タイポ」) があっても文章の意図を理解することができる。この人間に起こるタイポグリセミアという現象を、生成 AI の特徴を考慮しながら試し、結果を考察する。

1.2. 研究意義

現在の生成 AI が出力する結果は、人間が短時間で行える領域をはるかに超えている。そのため、タイポグリセミアのような文字の順番が少し変わっただけのものは簡単に修正できるのではないかと考えた。また、文章の校正に関して、もし人間ができるのに生成 AI にはできないことがあるとわかったら、興味深い事例だと思ったため研究した。

1.3. 研究背景

近年、目まぐるしい成長を遂げた AI が社会に大きな影響を与えている。特に、2022 年 11 月 30 日に OpenAI が発表した ChatGPT は、AI 業界を大きく変化させた。ChatGPT は、「文章の添削・校正、アイデアの提案、リサーチ、論点の洗い出し、壁打ち、ブレインストーミング、文章や概念の要約、コードの生成・提案※1」など、多様な物事に使用することができる。ChatGPT が無料で一般公開されたことや性能の高さが話題となり、生成 AI が広く知られるようになった。

しかし、生成 AI を利用することは良いことだけではない。生成 AI には著作権侵害や嘘の情報を出力してしまう (ハルシネーション) という問題を抱えている。そのため、安易に生成 AI を使用してしまうと、予期しない不都合に巻き込まれるかもしれない。

これからも生成 AI の利用者は増加していくことになると考えられる。仕事で利用することになれば、学校などの教育機関にも生成 AI が導入され、これからの児童・生徒においては生成 AI の知識を持つことが必須になる可能性もある。生成 AI で使用されることの多い文章の生成や校正においては、入力するスクリプトの中に誤字 (タイポ) が含まれる可能性も高い。誤字 (タイポ) と生成 AI の関連性を明らかにすることが、生成 AI の理解につながると考え、本研究を行った。

2. 研究方法

以下、確率的なタイポグラフィカルエラーを含む例文の生成においては、中央大学国際情報学部の飯尾淳教授によるプログラムを使用させていただいた。

2.1. 分析方法

(1) タイポグラフィカルエラーの作成。

- ・変換確率を使用して、タイポ文を作成。
- ・変換確率とは、文節の中の文字が一定の確率で入れ替わる度合いを表すもの。
- ・変換確率の幅は、0~1 である。
- ・数字が大きくなるほど変化する。
- ・文節の最初と最後は入れ替えない。
- ・比較基準となる変換確率が0の文を用意する。

例題：(下線部が入れ替わっている。[] 内は変換確率を表す)

[0.0]

「こんにちは みなさん おげんき ですか わたしは げんき です。この ぶんしょう は いぎりす の ケンブリッジ だいがく の けんきゅう の けっか にんげん は もじを に んしき する とき その さいしょ と さいご の もじさえ あつていれば じゅんばん は めちやくちや でも ちゃんと よめる という けんきゅう に もとづいて わざと もじの じゅんばん を いれかえて あります。 どうです ? よめちゃう でしょ ?」

[0.9]

「こちんには みさなん おんげき ですか わたしたは げんき です。この ぶしょう は いりぎす の ケブソツリジ だがいく の けきゅんう の けっか にげん は もじを に しんき する とき その さしいよ と さいご の もさじえ あれいてっば じばゆん は めくやちちや でも ちんやと よめる という けゅんきう に もづいとて わざと もじの じんばゆん を いかえれて あまりす。 どうです ? よやめちう でしょ ?」

(2) 手順

1. 文章を作成する。
2. 平仮名に変換する。
3. 半角スペースを挿入し、文節単位にする。
4. タイポグリセミアのある文章に変換する。
5. タイポグリセミアの文を生成 AI に送る。
6. 付録に結果をまとめる。
7. 付録にまとめた結果を考察する

(3) 分析対象

- ・ ChatGPT 4o mini、 Gemini

2.2. プロンプトの作成

生成 AI が適切な出力をするために、プロンプトに適切なガイドラインを含める必要がある。以下は、「深津式プロンプト※2」を参考に作成したプロンプトである。

#命令：

あなたはプロの編集者です。

今から、タイポグラフィカルエラーのある文章を送ります。

文章を読んで、タイポグラフィカルエラーを修正してください。

以下の条件をもとに、最高の結果を出力してください。

#条件：

- ・ タイポグラフィカルエラーの単語のみ修正してください。
- ・ 文章はひらがなで送られます。
- ・ 文字の順番を入れ替えて、タイポグラフィカルエラー修正してください。
- ・ 意味がわかる単語にしてください。
- ・ 意味がわかる文章にしてください。
- ・ 修正するときに、文章から単語の意味を推測できる場合は、意味を予測してください。
- ・ 予測した意味も教えてください。
- ・ 解答の文章は、送られた文章と同じ形で答えてください。
- ・ 深呼吸をして、落ち着いて回答してください。

3. 事例研究

3.1. ChatGPT の特徴

(1) GPT とは

GPT で主にやっていることは、「次に現れる単語を予測させ、学習させること※3」である。これは、文章から得られる情報（文法や単語など）を利用して、次の単語を予測することをいう。これを繰り返すことで、生成 AI は文章を生成しているのである。

(2) GPT の「T」である Transformer について

生成 AI の思考例として Transformer の特徴を説明する。

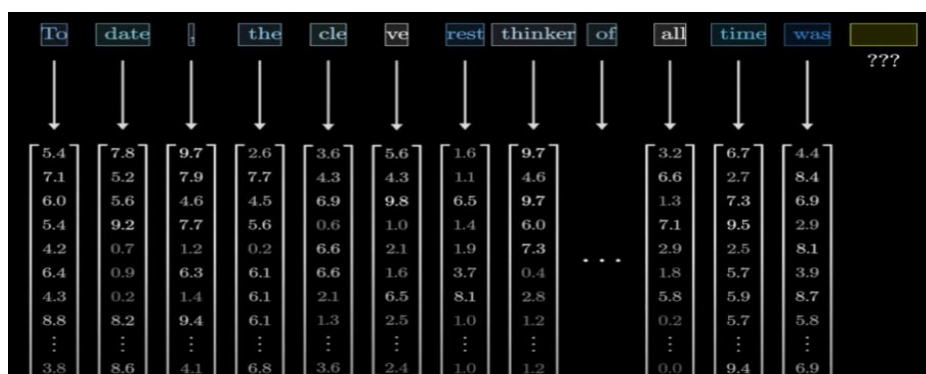
- ① 「Transformer はデータを Tokens に分けて、分類する※3。」人間のよう、文字を

単語や文節などで分けることもあれば、つながりが多い文字同士で分類している。このため、人間から見ると、よくわからない文字の集まりになる場合もある。



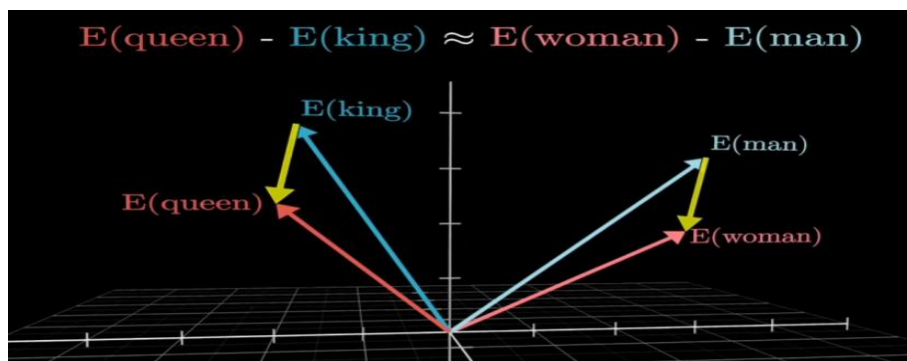
(写真1 ※4 <https://www.youtube.com/watch?v=KIZ-QmPteqM>)

- ② 「Tokens はベクトルと結び付けられている※3。」これは、分類されたものを表現するために必要になる。定義済みの語彙や単語は、埋め込み行列として個々に列を持っている。



(写真2 ※4 <https://www.youtube.com/watch?v=KIZ-QmPteqM>)

- ③ Attention ブロックでベクトルが互いに情報を渡しあい、情報を更新する。
 例： あめがふる。「雨」 あめをなめる。「飴」
 上記の文字は一緒でも意味が異なる場合などが例となる。周囲の情報（降る、舐める）を考えて、文脈や単語の意味の関連性を把握することで、情報の更新（雨と飴の違い）を助ける。
- ④ 空間座標の方向が言葉の意味と関連を示す。その関係性を利用して単語を推測することができる。例えば、いいこと（善）、わるいこと（悪）の位置関係を理解した場合、困った人を助けること（善）と犯罪行為（悪）の関係を推測できるなどがある。



(写真4 <https://www.youtube.com/watch?v=KlZ-QmPteqM>)

これらは、生成 AI のテキストに関する特徴といえる。人間とは思考のプロセスが異なるが、出力される結果は人間が思考したものと類似する結果を出すことができる。この特徴を考慮しながら、研究を進めた。

3.2. タイポが与える単語修正や文章理解への影響

この研究では、「生成 AI がタイポグラフィカルエラーを修正し、文章を読むことができるか」ということを検証する。先ほど述べた Transformer の特徴をもとに仮説を立てた。それは、「文脈と関連性の高い単語は理解でき、関連性の低い単語は理解ができない」というものである。

この仮説を立証するために 37 個のオリジナル文章を作り、目的となるような名詞や動詞 1 つをタイポグラフィカルエラーに変換した。そして、作成した文章を ChatGPT に送り、修正できるか検証した。分類として、①「対象の単語が文脈と関連性が高いもの」・②「対象の単語が文脈と関連性が低いもの」の 2 つに分けた。

「関連性が高い単語」とは、一般的な常識や経験で作られた文章に含まれる 1 つの単語のことである。「関連性が低い単語」とは、常識や経験からは予測ができないような単語や対義語などの、文章とは関係のない単語のことである。

データ収集のルールを簡単に決める。①の文章から検証し、①を修正できた場合、②を実行。①と比較して修正できたかどうかのデータを集める。①を修正できなかった場合、①と②を比較できないため、②は無視をする。

(※今回の研究で、意味は理解できるが文字を修正できないという事例が多いことがわかった。単語の意味を理解できており、漢字などで修正し、説明したものに限り修正できたということにする。)

以下はこの研究の例である。

- ・文脈から意味を想像できる単語がひとつだけタイポグラフィカルエラーを起こしている

文章。

本文：「きょうは しごと だから、かいしゃ へ むかう。」

Typo：「きょうは しごと だから、かしいゃ へ むかう。」

ChatGPT 4o mini の回答

修正後：「きょうは しごと だから、かいしゃ へ むかう。」

「かしいゃ」→「かいしゃ（会社）」に修正しました。

・文脈とは関連性が薄い単語がひとつだけタイポグラフィカルエラーをしている文章。

変更「かいしゃ」→「ぱちんこ」

Typo：「きょうは しごと だから、ぱんちこ へ むかう。」

ChatGPT 4o mini の回答

修正後の文章：「きょうは しごと だから、ぱんちこ へ むかう。」

修正した単語：「ぱんちこ」→「パンチコ」→「パンチカフェ」

予測した意味：「ぱんちこ」という単語は明確に意味を推測できませんが、「パンチカフェ」というカフェの名前である可能性があります。

・文脈から意味を想像できる単語がひとつだけタイポグラフィカルエラーしている文章。

本文：「ぎゃんぶる いぞんしょう の ひとは、ぱちんこ に いくのを やめられない。」

Typo：「ぎゃんぶる いぞんしょう の ひとは、ぱんちこ に いくのを やめられない。」

ChatGPT 4o mini の回答

修正後の文章：ギャンブル 依存症 の 人は、パンチコ に 行くのを やめられない。

ぱんちこ → パチンコ：「ぱんちこ」は「パチンコ」の誤記だと予測しました。

・文脈とは関連性が薄い単語がひとつだけタイポグラフィカルエラーをしている文章。

変更：「ぱちんこ」→「かいしゃ」

Typo：ぎゃんぶる いぞんしょう の ひとは、かしいゃ に いくのを やめられない。

ChatGPT 4o mini

修正後の文章：ぎゃんぶる いぞんしょう の ひとは、かしおや に いくのを やめられない。

「かしいゃ」は「かしおや」のタイポだと考え、「文脈的には「賭け屋」に関連していると思われるため、「かけや」と予測しました。

結果と考察：

自作した37の文章のうち「①関連性の高そうな単語」を修正できた数は、37個中27単語、できなかった数は、37個中10単語。「②関連性の低そうな単語」を修正できなかった数は、27個中25単語。①・②どちらも修正できた数は、2単語であった。

Transformerの特徴からわかる仮説通り、「関連性」が高ければ修正率は高くなり、「関連性」が低ければ、修正率はかなり低かった。

また、文脈から意味を予測する能力は高いが、文字の順番を修正して正しい単語に修正する能力は低いことがわかった。人間も文字単体を認識しながら意味を理解しているわけではない。文章の文脈から文字の集合（単語）に意味づけをして文章を読んでいる。人間に似た特徴を生成AIは、データと確率によって行っている。そのため、ChatGPTは文章と「関連性の低い文字の集合」よりも「関連性の高い文字の集合」を優先して考え、「関連性の高い文字の集合」に関わるデータを使用して確率的に文字の変換を行ったと考えられる。

しかし、タイポグラフィカルエラーがあるという情報があれば、人間は文字の順番が入れ替わっていることを理解できるはずである。生成AIは「文字の順番が入れ替わっている」と認識する柔軟性がない可能性がある。

文字の入れ替わりを認識しない点についてもう少し考えてみた。なぜChatGPTは、単語の意味や正しい文字の順番を出力しているのに、間違っただけの文字の順番で出力してしまうことがあるのか。以下はその例である。

・間違っただけの文字の順番を出力した例

ChatGPT 4o mini の回答

修正後の文章： ギャンブル依存症の人は、パンチコに行くのをやめられない。

ばんちこ → パチンコ：「ばんちこ」は「パチンコ」の誤記だと予測しました。

上記は、研究例の一部を抜粋したものである。一見修正できているようにみえるかもしれないが、修正後の文章はパンチコと真ん中の文字が入れ替わっていない。しかし、正しい順番のパチンコだということも理解しているのである。

この理由の仮説は、ChatGPTにとって「パチンコ」も「パンチコ」も同じ確率的に選ばれた「関連性の高い文字の集まり」であり、どちらも「ギャンブル依存症の人がよく行う行為」の「パチンコ」と考えている。文字が1文字変わっていても、関連性の高いデータから選ばれた文字であるため、そのミスに気づけなかったと考えられる。

これらのことから、たとえ普通の文章で起こったタイポグラフィカルエラーだとしても、一定の確率で文字の順番は修正されない可能性がある。そのため、ChatGPTに文章を校正してもらうより、ChatGPTが作成した文章を人間が校正し、使用する方が実用的であると考えられる。

※付録

タイポが与える単語修正への影響①

タイポが与える単語修正への影響②

タイポが与える単語修正への影響③

3.3. 単語（名詞）のみのケース

「情報を交換し合い、文字の集合を修正する」という Transformer の特徴がある。この特徴から、1つのジャンルにおける名詞（例えば：スポーツや国名など）を並べて、タイポグラフィカルエラーの単語にしても、関連性が高いので修正できるのではないかと考え検証した。

「国名・県名・スポーツ・野菜など」10ほどのジャンルを10個の単語で分け、修正できるか試した。基本的に難しくなく、有名なものに限定して検証を行った。

修正できた文章の例

(1) すべての単語が関連性の高いもの テーマは「野菜」

とまと にんじん きゃべつ きゅうり はくさい ぶろっこりー じゃがいも ほうれんそう
だいこん かぼちゃ

タイポグラフィカルエラーの単語

とまと にじんん きべやつ きゅうり はさくい ぶこっろりー じいやがも ほれうそんう
だこいん かちぼや

ChatGPT 4o mini

修正後の文章は以下のようになります：

とまと にんじん きゃべつ きゅうり はくさい ぶろっこりー じゃがも ほうれんそう だ
いこん かぼちゃ

予測した意味：

- にじんん → にんじん (人参)
- じいやがも → じゃがも (ジャガイモ)

結果と考察：

検証の結果、「関連性の高い単語」を集めるだけだと単語が修正されるものと、されないものがあることがわかった。「国名・野菜・料理・有名企業など」は7割以上修正できたが、「スポーツ・職業・県名など」6割以下や2割・0割というジャンルもあった。単語の文字数の違いなどはあるが、有名なものを使用したのでジャンルの間に大きな違いがあるわけ

ではないと思う。そのため、文章のような形式である方が、確率的にパターンを読み込みやすいのかもしれないと考えられる。

このほかにも「関連性の低い単語だけを集める検証」や「単語の修正率が高いジャンルに関連性の低い単語を混ぜる検証」、「関連性の高い単語を集めたジャンル2つを混ぜる検証」など様々な検証を行った。

その中で、特徴的な結果だったのが「関連性の低い単語」の影響である。「関連性の低い単語」を集めたものを「関連性の高い単語」と同じ数の10文章ほど試したが、修正率が高いものはなく、修正できたとしても2単語ほどだった。そのため、関連性の低い集合は、理解されにくかった。加えて、「修正率と関連性の高い単語」に「関連性の低い単語」を混ぜると ChatGPT の理解がかなり低下した。

・関連性が低い単語の検証例

本文

おにぎり こうもり あふりか すいせい きょうかしょ となかい ぼけもん ちちおや そふとばんく ちやいろ

タイポ文

おぎにり こもうり ありふか すせいい きよしかうよ とかない ぼもけん ちおちや そとふんばく ちいやる

ChatGPT 4o mini

修正後の文章は以下の通りです。

おぎにり こおうり ありふか せいい きょうしかう とかない ぼけん ちおちや そとふんばく ちいろ

「関連性が低い単語」とタイポグラフィカルエラーの組み合わせは、ChatGPT に大きなマイナスの影響を与えることが分かった。ただ、「関連性の高い単語」でも失敗した例「スポーツ・職業・県名など」があり、関連性が高いと1つの情報を加えるだけで、違う結果が起こるのではないかと考え、検証した。

検証の例

本文

やきゅう さっかー じゅうどう すいえい たつきゅう ぼくしんぐ まらそん らぐびー ぼとみんとん ばれーぼーる

タイポ文

やゆきう さかっー じどゆうう すえいい たきゅう ぼしんくぐ まらそん らびぐー ぼとんみんとん ばぼーれーる

・[スポーツ]と伝えてない場合 10個中3個修正。

修正した文章は以下の通りです：

やきゅう さかつう じどうう すいえい たきょう ぼしんぐ まらそん らいぐ ばとんみ
とん ばぼーれる

【予測した意味】

- ・ やきゅう (野球)
- ・ さかつう (サッカー)
- ・ まらそん (マラソン)

・[スポーツ]と伝えた場合 10個中8個修正。

以下のようにタイポグラフィカルエラーを修正しました：

1. やゆきう → やきゅう (意味: 野球)
2. さかつー → さっか (意味: サッカー)
3. じどゆうう → じどうう (意味: 自道具、もしくは自動車)
4. すえいい → すいえい (意味: 水泳)
5. たきゅうう → たつきゅう (意味: 卓球)
6. ぼしんくぐ → ぼしんぐ (意味: ボウリング)
7. まらそん → まらそん (意味: マラソン)
8. らびぐー → ラグビー (意味: ラグビー)
9. ばとんみとん → ばとんみとん (意味: バトミントン)
10. ばぼーれーる → バレーボール (意味: バレーボール)

推測した意味は、各スポーツに関連する単語として修正しました。

この検証では、単語の集まりがどのようなジャンルなのか伝えた場合、修正率がどのように変化するか試した。この結果、ただ単語を送ったときとは異なり、多くの単語を修正することができた(しかし、修正した単語の文字は、正しい単語の文字とは異なることがある)。

以上のことから、「関連性が高ければ」単語が修正されやすく、「関連性が低ければ」単語が修正されにくい傾向があることは分かった。しかし、関連性が高くても失敗するケースはあるため、ChatGPTがタイポグラフィカルエラーに完全に対応できるとはいえないと思う。

生成AIの確率的に次の単語を予測する特徴とジャンルでまとめられた単語の集合は、相性がよく修正できると考えていたが、私が期待したほどの結果にはならなかった。また、答えではないChatGPTの情報更新を助けるような「情報」を与えれば、出力の精度はかなり高まった。そのため、データへの依存性が強い傾向があり、正しい情報に導くためのプロンプトが重要であるといえる。

単語に焦点をあてて検証した結果、文字を修正できない傾向が強いことがわかった。単語の意味や正しい順序を知っているのに、完全に修正できない事例が多くあり、Transformer

の機能にタイポグラフィカルエラーはネガティブな影響を与えられられる。

※付録

単語（名詞）のみのケース① 単語（名詞）のみのケース②

3.4. 変換確率「0.9」の文章で検証

今までの検証では、単語にアプローチをするやり方を検証した。今回は、文章全体を検証の対象として実験した。分析対象は、自作の文章やあらかじめ作られた文章（教科書の一部など）を使用した。文章内の4文字以上の単語がタイポグラフィカルエラーになったとき、ChatGPTは単語を修正できるのか、文章の内容を理解できるのかについて調べた。

変換確率「0.9」とは、4文字以上の集まりがタイポグラフィカルエラーになる確率のことをいう。したがって、約90%の確率で単語がタイポグラフィカルエラーになるのでかなり見づらい文章になっている。下記にその例文を記載する。

元の文章と変化した文章の例

本文

「きゅうしょくのかれーをくらすのぜんいんでおなじりょうずつわけることになりましたあやかはかれーがすきではないのですこしにしてほしいとおもっています いっぽうでゆうやはおなかがすいていたのでえーこれだけとっています」

タイポ文[0.9]

「きょしうゆくのかれーをくらすのぜいんんでおなじりょうずつわこるけとになましりたあやかはかれーがすできはなのいですしこにしてほしいとおっいまもてすいぽうでゆうやはおかながすいいてたのでえーこだれけとっています」

※5 文部科学省検定済教科書 小学校道徳用 小学道徳 生きる力 5

結果と考察：

この検証では、タイポグラフィカルエラーのある確率が高い文章の意味を理解できているのか、誤字になっている単語を正しく修正できたのか調べた。その結果、完璧に元の文章に戻せたケースはなく、文字の順番を修正できる確率も低かった。先ほどの例文に対するChatGPTの出力を下記に記載する。

・ ChatGPT 4o mini の回答

「修正後の文章は以下の通りです。

きょうしゅくのかれーをくらすのぜんいんでおなじりょうずつわけるけとになましりたあやかはかれーがすきだはなのいですしこにしてほしいとおっしゃいましたいぽうでゆうやはおかながすいていたのでえーこだわりとっています

予測される意味:

- 「きょうしゅく」→「きょうしゅく」(宿泊、食事を提供する場所)
- 「ぜんいん」→「ぜんいん」(全員)
- 「わこるけと」→「わけるけど」(分けるけど)
- 「なましりた」→「なまえをしりた」(名前を知りたい)
- 「すできは」→「すきだは」(好きだは)
- 「すしいこ」→「すしにこ」(寿司にしたい)
- 「おっいまもてす」→「おっしゃいました」(言いました)
- 「すいいてた」→「すいていた」(空いていた)
- 「えーこだれけ」→「えーこだわり」(こだわり)

この例で修正できたのは1単語(ぜんいん[全員])のみで、ほかは異なる文字の集合に変換されてしまっている。また、その変化によって文章の意味も変わってしまった。そのため、ChatGPTはこの文章を正しく理解、修正することができなかったといえる。

この文章は小学生の道徳の教科書を参考に一部抜粋したものである。この文章は、不公平であるかを判断することを目的として作成された文であったため、ChatGPTに「この文章を読んで不公平だと思ったか」という質問を試みた。

ChatGPT 4o mini

[本文の場合]質問：これは不公平だと思いますか？

答え：不公平だと思わない。

この場合、あやかさんとゆうやさんの食の好みやおなかのすき具合が違うので、分け方に工夫が必要です。あやかさんが少しだけ欲しいと思っているのに対し、ゆうやさんがたくさん欲しいがっているのは、それぞれの状況に応じたニーズがあるから、不公平だとは一概には言えません。しかし、みんなに分けるときに、お互いの気持ちを考えた調整が必要かもしれません。例えば、あやかさんには少なめに、ゆうやさんには多めに分ける、もしくはあやかさんの分を少し残して、他の人が調整してもいいかもしれません。

ChatGPT 4o mini

[変換確率 0.9 の場合]質問：これは不公平だと思いますか？

答え：理解できなくなっている。

この文章は、何かの出来事についての報告や意見を述べているようですが、「不公平」かどうかを判断するためには、具体的な状況や文脈が必要です。文章自体からは、誰かがカレーを食べたくないと行って寿司を希望している場面が描かれています。それが不公平かどうかは、その時の状況や関係性に依存します。たとえば、他の人がカレーを好んでいる中で、あやかさんだけが特別に寿司をお願いすることで不公平に感じるかもしれませんが、逆に個々の好みを尊重することが大切だという意見もあります。

元の文[本文]では、文章の意味を理解して回答しているのに対し、タイポグラフィカルエラーの文章[0.9]では、文章の具体的な情報がたりないと理解ができなくなっていた。このほかにも約30程度の文章をいろんな形式で試した。基本的に調べたことは、「文字の修正」と「文章の理解」についてであり、すべて元の文章と比べて理解度が低くなった。

部分的に理解できているような出力をしている結果も存在した。しかし、タイポグラフィカルエラーではない文字の集まりから予測された結果であり、確率的に抽出された結果はかなり違和感のある文章になっていた。そのため、ChatGPTはタイポグラフィカルエラーへの変換確率[0.9]の文章の内容を、読むことはできなかったといえる。

※付録

変換確率「0.9」の文章に関する資料①

変換確率「0.9」の文章に関する資料②

変換確率「0.9」の文章に関する資料③

3.5. その他の検証

「文章の理解」についてアプローチを変えて少し検証したものである。検証した事例を下記に掲載する。

4 択の知識問題

[本文]

おおさかのめいぶつりょうりたこやきのなかみにつかわれるぐざいとしていっばんてきにはいっていないものはどれですか？

- A) たこ
- B) エビ
- C) ねぎ
- D) こんにゃく

ChatGPT 4o mini

こたえは D) こんにゃく です。

たこやきのなかみには、たこ、エビ、ねぎなどがよく使われますが、こんにゃくはあまり一般的な材料ではありません。

タイポ文「0. 9」

おおさかのめぶいつりうよりたよきのみかにつれかわるぐざいとしていてんぱつきにはいいものはどれですか？

- A) たこ
- B) エビ
- C) ねぎ
- D) こんにゃく

ChatGPT 4o mini

このもんだいのせんたくしのなかで、めぶいつりうよりたよきのみかにおいてんぱつきにちょうどあうぐざいとして、もっともふさわしいのは「たこ」だとおもいます。たこはおおさかのめぶいつりうよりたよきなどでよくつかわれるし、てんぷらにもあいそうです。ですので、答えは A) たこ です。

不正解

結果と考察：

4 択問題を 10 問ほど試した。正解した数は 4 問、不正解の数は 6 問という結果だった。タイポグラフィカルエラーがない文章では 10 問中 10 問正解だった。知識を問うような問題だったため、少し難しかったと思う。また、正解した回答には、あまり精度の高くない回答が存在した。

本文

あにめぼけもんでいちばんにんきのきやらくたーはどれでしょうか？

- 1 ぴかちゅう
- 2 ひこざる
- 3 なえとる
- 4 かいおーが

ChatGPT 回答

ぼけもんのなかでいちばんにきんのきやくたーは「ぴかちゅう」だとおもいます。ぴかちゅうは、アニメやゲーム、映画でもよくみられるシンボルキャラクターですし、世界中でも人気があります。

・精度の高くない正解の回答例。

タイポ文「0. 9」

あにめ ぼもけん で いばちん にきんの きくたやらー は どでれしょうか ?

- 1 ぴかちゅう
- 2 ひこぎる
- 3 なえとる
- 4 かいおーが

ChatGPT の回答

この問題は「ポケモン」のキャラクターに関するものですね！あなたの質問に対して、正しい答えは「1 ぴかちゅう」です。

「いばちんにきんのきくたやらー」というのは、きっと「いばしんにきんのきくちょう」という形にしたかったのでしょうか。ぴかちゅうは「いばしん」を持っているポケモンの代表的なキャラクターです。

正答率は低かったものの、変換確率が「0.9」の文章を読んで正解へ導いたことに驚いた。付録には載せていないが、チャットを更新した後、ChatGPT に再度失敗したものを送ると、正解するものもあった。すべて確率やデータに依存するため、今回は正答率が低かった。

今回の 4 択問題で適切な回答をできた理由は、選択肢にあると思う。選択肢がデータやプロンプトのような役割を果たし、確率的に文字を選ぶための情報を提供できていたのではないかと思う。これにより、膨大な情報から取捨選択をする必要がなくなり、対象の文章と選択肢、それに関わるデータを確率に織り込むことで正解を導き出すことができたと考えられる。

しかし、確率によって失敗するケースがある以上、タイポグラフィカルエラーのある文章で、4 択問題を解くことは得意なことではないといえる。

※付録

その他の検証 資料① その他の検証 資料②

3.6. まとめ

ChatGPT は、単語や文章を校正するとき、タイポグラフィカルエラーのある単語や集まりが含まれると適切に出力ができないことがわかった。そのため、Transformer の性質とタイポグラフィカルエラーの相性はあまり良くないように思える。Transformer の特性を利用して、「関連性の高い」確率的に予測しやすい集まりで検証した場合、適切な出力をすることがあった。そのため、ChatGPT に合わせたアプローチが重要であると考えられる。例えば、詳細なプロンプトやデータの範囲を絞る情報などが必要になる。

全体の検証で現れた特徴的な現象があった。それは、「単語の意味や文字の正しい順番を理解しているのに、文字の順序の修正が行われていないこと」である。これは、Transformer の特徴である「確率的に次の文字を予測する」がマイナスに作用していると考えられる。

その作用とは、データを利用して確率的に選ばれるため、「正しい答えが1つではなくなってしまうこと」である。これにより、ChatGPT 独自の種類の単語の形を作り出してしまったと考えられる。このため、固有名詞だとしても複数の文字配列が存在することになることがある（例えば、スポーツとスーポツは同じ運動競技の意味）。このことは、ChatGPT の弱点ではないかと思う。

以上のことから、タイポグラフィカルエラーは ChatGPT が普段行えている単語・文章の読み取りや文章校正に悪影響を及ぼし、精度の悪い出力を強いてしまうような影響があると分かった。

4. オリジナル文章を用いた評価実験

4.1. 研究の目的

本章では、LLM の純粋な文章理解能力、特にタイポグリセミアに対する耐性を評価するため、オリジナル文章を用いた実験を行う。

4.2. 実験データの作成

本実験で使用するオリジナル文章は、以下の3つの基準に基づいて作成した。

1. **表記:** 文章はすべてひらがなで記述する。これは、漢字やカタカナが混在する場合と比較して、タイポグリセミアが与える影響をより明確に捉えるためである。また、ひらがなのみの文章は、視覚的な情報が少なく、LLM の言語処理能力がより直接的に試されると考えられる。
2. **単語構成:** 文章中に4文字以上の単語をなるべく多く含むようにする。これは、本実験で用いるタイポグリセミア付与方法の一つである文字の入れ替えが、3文字以

下の単語では機能しないためである。4文字以上の単語を用いることで、より多様なタイポグリセミアを生成することが可能となり、LLMの文章理解能力をより多角的に評価できる。また、短い単語と比較して、長い単語の方がタイポグリセミアの影響を受けやすく、LLMの誤認識を引き起こしやすいという仮説に基づいている。さらに、長い単語を含む文章は、文脈情報の重要性が高まり、LLMの文章理解能力をより深く評価できると考えられる。

3. **文字数:** 文章の長さは100~200文字程度とする。これは、実験の実施と結果の分析を効率的に行うための措置である。短すぎる文章では文脈情報が不足し、長すぎる文章では実験の実施に時間を要する。100~200文字程度という長さは、文脈情報と実験効率のバランスを考慮した結果である。

これらの基準に基づいて作成されたオリジナル文章の例を以下に示す。

- 「あたたかいひだまりのなかで、かわいらしいこねこたちが、まるくなってすやすやねむっている。」
- 「きいろいちゅーりっぷが、たくさんならんで、うつくしいはるのこうえんをかざっている。」
- 「おおきなくじらが、あおいおおなかのなかで、ちいさなえびたちとくらしている。」

これらの文章は、日常生活で使われる自然な表現を用いながらも、上記の基準を満たすように作成されている。

4.3.実験方法

LLMに解答してもらう方法は、4つの選択肢の中から正解を選ぶ形式とした。この形式を採用する理由は、正誤判定が容易であり、実験結果の定量的な分析に適しているためである。また、今回使用するLLMはChatGPTとGeminiの2つを使うことにする。

具体的な実験手順は以下の通りである。

1. オリジナル文章に、文字の入れ替えをプログラムで意図的に付与する。タイポグリセミアの付与パターンは、単語の先頭と末尾の単語だけ入れ替えずにその間の単語だけを入れ替えるようにしている。
2. タイポグリセミアが付与された文章と、正解を含む4つの選択肢をLLMに提示する。選択肢は、正解の文章と類似した表現を用いながらも、意味が異なるように作成する。
3. LLMが選択した解答と正解を比較し、正誤を判定する。

4. 上記の手順を、2つのオリジナル文章とタイポグリセミアのパターンに対して繰り返し行う。

4.4. 評価指標

本実験では、LLMの正解率を主要な評価指標とする。正解率は、LLMが正しく解答した回数を、実験の総試行回数で割った値である。ここで、「正しく解答した」とは、事前に設定した正解データとLLMの出力が完全に一致する場合、もしくは、事前に定義した評価基準（例えば、意味的に同等である、特定のキーワードを含むなど）を満たす場合を指す。

5. 実験結果

5.1. 用意した文章と問題

実験に用いる文章として、100~200文字程度の文章を2つと選択肢問題を用意した。文章の左側に書かれている数字が、1に近づくほど文字の入れ替わりの程度が強くなることを表している。

一つ目の文章

基本 「ゆうやけが とても きれいです むらさきいろとおれんじいろのぐらでーしょんがすばらしいですね。ゆっくりとながめるじかんはほんとうにぜいたくです。」

0.5 「ゆやうけが とても きれいです むいさきらろとおれんじいろのぐーしょでらんがすらしばい ですね。ゆくっりとなめがるじかんはほんとうにぜいたくです」

0.7 「ゆやうけが とても きれいです むいきさらろとおれんいじろのぐよしでーらんがすばらしい ですね。ゆっくりとなめがるじかんはほうんとにぜたいくです」

0.9 「ゆやうけが とても きれいです むさらいきろとおいじんれろのぐしょでらんがすしばらい ですね。ゆくっりとながめるじかんはほとうんにぜたいくです」

問題

問題1：この文章は何について書かれていますか？

- あ：朝の風景
- い：夕焼けの風景
- う：夜の風景
- え：雨の日の風景

問題2：夕焼けの色はどのように表現されていますか？

- あ：あかいろ と きいろ
- い：みどりいろ と あおいろ
- う：むらさきいろ と おれんじいろ
- え：しろいろ と ぐろいろ

問題3：「ゆっくりとながめるじかんはほんとうにぜいたくです」とありますが、この文章からどのような気持ちが伝わってきますか？

- あ：急いでいる気持ち
- い：悲しい気持ち
- う：ゆったりとした気持ち、満足した気持ち
- え：怒っている気持ち

二つ目の文章

基本「むかしむかしとてもふしぎなもりがありましたそこにはおおきなきのしたにちいさいいえがありかわいいどうぶつたちがすんでいましたあるひひとりのおんなのこがもりのなかにまよいこんでしまいましたちよっぴりこわいきもちであるいているとまえからうさぎがあらわれましたうさぎはにこにこわらっておんなのこをいえまでみちびいてくれたのです」

0.5 むかしむかしとてもふしぎなもりがありましたそこにはおきおなきのしたにちいさいいえがありかわいどうぶつたちがすんしいまでたあるひひとりのおんなのこがもりのなかにまこいよんでしいままたちびよっぴりこわいきもちであるいているとまえからうさぎがあましわられたうさぎはにこにこわらっておんなのこをいえまでみちびいてくれたのです

0.7 むかかむしし とても ふしぎ な もり が あましりた そこには おきおな きの した に ちさいな いえ が あり かいわい どつたうぶち が すでましんいた あるひ ひり との おなのんこ が もり の なか に まよいこんで しまいました ちっぴより こわい きもち で あるいてると まえ から うさぎ が あらしれまわた うさぎ は ににここ わらって おなのんこ を いえ まで みいちびて くれた の です

0.9 むむかかしし とても ふしぎ な もり が あしりまた そこには おきおな きの した に ちさいな いえ が あり かいわい どぶうたつち が すでんましいた あるひ ひり との おなのんこ が もり の なか に まいよんこで しまいました ちびよっり こわい きもち で あいるいてると まえ から うさぎ が あれしらわまた うさぎ は ににここ わらって おのんなこ を いえ まで みいちびて くれた の です

問題1：この物語の始まり方は、昔話でよく使われる表現で始まっています。それはどのような表現ですか？

あ：「あるところに」

い：「むかしむかし」

う：「そのとき」

え：「それから」

問題2：森には何がいましたか？（複数回答可）

あ：おおきな き

い：ちいさな いえ

う：かわいい どうぶつたち

え：たかい やま

問題3：女の子が森で感じた「ちょっぴり こわい きもち」の原因として考えられることは何ですか？

あ：おながが すいていたから

い：もりに まよいこんで しまったから

う：あつい から

え：ねむい から

問題4：女の子は最終的にどうなりましたか？

あ：もりの なか で ずっと ひとり で くらした
 い：うさぎ に みちびかれて いえ に たどりついた
 う：もりの そと へ じぶん で でて いった
 え：ほかの どうぶつ に おそわれた
 問題5：この物語全体を通して、どのような印象を受けますか？
 あ：とても かなしい おはなし
 い：ちょっぴり こわい けど あたたかい おはなし
 う：とても げんき が できる おはなし
 え：なぞ が たくさん ある おはなし

5.2. 実験結果及び考察

本実験の結果、概ね高い正解率が確認された。図1に示すように、対象としたLLM、特に Gemini においても数問の誤答が見られたものの、全体的な傾向としては正解率が非常に高いことが示された。Gemini において見られた誤答は、もう一度やり直したところ正解したため、単純なLLM特有のミスであり、LLMが一般的に陥りやすいエラーパターンに該当し、意味理解の根本的な欠如を示すものではないと考えられる。これらの例から、LLMは個々の文字レベルの変動には影響を受けるものの、文脈全体から意味を推測する能力を有していることが示唆される。

GPT					Gemini				
	0.9	0.7	0.5	0		0.9	0.7	0.5	0
問題1	○	○	○	○	問題1	○	○	○	○
問題2	○	○	○	○	問題2	○	○	×	○
問題3	○	○	○	○	問題3	○	○	○	○
	0.9	0.7	0.5	0		0.9	0.7	0.5	0
問題1	○	○	○	○	問題1	○	○	×	○
問題2	○	○	○	○	問題2	○	○	○	○
問題3	○	○	○	○	問題3	○	○	○	○
問題4	○	○	○	○	問題4	○	○	○	○
問題5	○	○	○	○	問題5	○	○	○	○

(図1 正答率)

本実験の結果は、実験開始前の予想とは大きく異なるものであった。当初、文章の意味を大きく損なうレベルのタイポグリセミアに対しては、LLMは全く対応できない、すなわち正解を導き出すことは不可能であると予想していた。しかし、実際の結果は、LLMがタイポグリセミアに対して一定の耐性を持っていることを明確に示している。

この結果は、LLMが単語や文字の表面的な情報だけでなく、文脈や意味に基づいた言語処理を行っていることを示唆している。つまり、タイポグリセミアによって個々の単語が変化しても、周囲の文脈から本来の意味を推測し、正解を導き出すことが可能になると考えられる。これは、LLMが単なる文字列マッチングではなく、意味に基づいた言語

処理を行っていることの強力な証拠となる。深層学習モデルがどのように文脈情報を捉え、タイポグリセミアのようなノイズに対して頑健な表現を獲得しているのか、そのメカニズムの解明は今後の重要な研究課題である。

この LLM のタイポグリセミアへの耐性は、自然言語処理の分野において重要な意味を持つ。例えば、ユーザーが入力するテキストには、タイプミスやスペルミスが頻繁に発生する。LLM がこのようなエラーに対して耐性を持つことは、より自然で使いやすいインターフェースにつながっていると考えられる。

5.3. まとめ

生成 AI (大規模言語モデル、LLM) がタイポグリセミア、すなわち単語内の文字順序が入れ替わっていても最初の文字と最後の文字が正しければ人間が比較的容易に読めてしまう現象をどの程度理解できるのかを検証した本研究の結果から、LLM はタイポグリセミアに対して一定の耐性を持つという結論に至った。

従来の自然言語処理では、文字列の一致に基づいて単語や文章を処理する方法が主流であった。しかし、本研究で対象とした LLM は、単なる文字列マッチングに留まらず、文脈や意味に基づいた高度な言語処理を行っていることが示唆された。実験では、様々なレベルのタイポグリセミアを含む文章を LLM に入力し、文章の意味を正しく理解できるかを評価した。その結果、LLM は比較的高い精度で文章の意味を理解することができていた。この結果は、LLM が単語を構成する個々の文字だけでなく、周囲の文脈情報を用いて意味を推測する能力を有していることを示している。タイポグリセミアによって文字の位置が変化しても、前後の単語や文章全体の流れから本来の意味を推測し、正しい解釈を導き出すことが可能になっていると考えられる。この能力は、LLM が深層学習によって獲得した高度な言語表現力によるものと考えられる。

さらに、本研究は LLM のノイズやエラーに対する耐性を示す事例である。現実のテキストデータには、タイプミスやスペルミス、表記の揺れなどが頻繁に含まれる。LLM がタイポグリセミアに対して耐性を持つことは、このようなノイズを含むデータに対しても高い精度で対応できることを意味し、実用的な自然言語処理アプリケーションにおいて非常に重要な利点となる。例えば、チャットボットや音声アシスタントなどのユーザーインターフェースにおいて、ユーザーの入力ミスをある程度許容することで、より自然で使いやすいインターフェースを提供することが可能になる。ただし、本研究では比較的簡単な文章に焦点を当てており、より複雑なパターン、例えば歴史の単語や倫理感が欠如している文章に対しては、LLM の性能は低下する可能性が考えられる。また、使用する LLM のアーキテクチャや学習データによっても結果が異なる可能性がある。

今後の研究では、これらの点を考慮し、より詳細な検証を行う必要がある。具体的には、様々な種類の文章を網羅的に検証すること、更に異なる LLM を比較すること、人間との比較実験を行うことなどが考えられる。

結論として、本研究は LLM がタイポグリセミアに対して一定の耐性を持つことを明らかにした。この知見は、LLM が単なる文字列処理ではなく、意味に基づいた高度な言語処理を行っていることの証拠となるとともに、実用的な自然言語処理アプリケーションにおける LLM の有用性を示唆するものである。今後の研究によって、LLM の言語理解能力のメカニズムがさらに解明されることが期待される。

6. おわりに

本研究は、タイポグラフィカルエラーが生成 AI の出力にどのような影響を与えるのかを検証した。単語や文章にタイポグラフィカルエラーが含まれる場合、ChatGPT が普段行っている単語・文章の読み取りや文章校正に悪影響を及ぼし、精度の悪い出力を強いてしまうような影響があると分かった。

4. オリジナル文章を用いた評価実験では、「LLM はタイポグリセミアに対して一定の耐性を示すことが明らかになった。」という結果だった。とても興味深い結果だが、検証やデータに偏りがあるのではないかと感じる。そのため、この結果のみで生成 AI がタイポグラフィカルエラーに対応しているとは言い難い。

ただ、プロンプトのようなデータの範囲を絞る情報を与えることで結果が飛躍的に向上した事例もあった。3.3.単語（名詞）のみのケースでは、どんな集まり（ジャンル）を伝えることで、修正できなかった単語を修正することができた。これに似た現象が4択問題でも起こっているのではないかと考えられる。文章に対する選択肢という比較によってデータの範囲を限定することに成功し、正しい出力ができたのかもしれない。

しかし、Transformer の機能とタイポグラフィカルエラーの相性はあまり良くない。なぜなら、確率的に抽出された関連性の高い文字の配列であれば、文字の順序が違っていても正しい単語だと認識してしまうからである。さらに、ChatGPT には確率的に選ばれた関連性の高い文字の配列から、新しい単語を作り出してしまう事例もあった。

以上のことから、タイポグリセミア現象は、人間にはできて、現在の生成 AI にはできない事例の一つであるといえると考えられる。しかし、生成 AI の文章校正などの能力を検証で見ると、将来の生成 AI はこの問題を解決できるとも思った。

6.1. 今後の課題

今後の課題として、人間と生成 AI を比較する検証を行うことや有料版にある高度な生成 AI を利用した検証を行うのも面白いと思う。タイポグラフィカルエラーのある文章に対して、適切な回答をした文章を対象に、半角スペースで文節を区切らないより分かりにくい文章で試すなどの検証を行ってみるのもよい。また、今回はほとんど漢字を含まないテキストを対象としたが、漢字を含ませた場合にどのように変化するかなども興味ふかい。

今回の検証の課題として、人間・生成 AI・文章・その他の何かに焦点を当て、もっと具

体的な比較ができれば良いと感じた。生成 AI の進化のスピードはかなり速いので、時間（時期）を図って検証を行うことも重要であり、長期的な実験計画が必要だったと思う。

謝辞

本研究では、タイポグラフィカルエラーを含む例文の生成においては、中央大学国際情報学部の飯尾淳教授によるプログラムを使用させていただきました。深く感謝いたします。

参考文献

※ 1 「生成 AI の技術動向と影響」 厚生労働省

<https://www.mhlw.go.jp/content/11601000/001125241.pdf>

※ 2 「あなたの仕事が劇的に変わる!? ChatGPT を使いこなし最前線」

チャンネル名 note 最終閲覧日 2024/12/29

<https://www.youtube.com/watch?v=ReoJcerYtuI&t=2640s>

※ 3 「AI ってなに? ③【詳細解説】生成 AI ChatGPT って何? 仕組みからわかる「真の恐ろしさ」【Gemini/Grok/Copilot/Claude/LLM】」チャンネル名 たてはま/CGBeginner

<https://www.youtube.com/watch?v=7tcSj9YCZQ>

※ 4 「GPT とは何か Transformer の視覚化 | Chapter 5, Deep Learning 」

YouTube, 最終閲覧日 2024/12/29 チャンネル名 3Blue1BrownJapan

<https://www.youtube.com/watch?v=KlZ-QmPteqM>

写真 1 (3:22) ,2 (3:44) ,3 (15:01)

※ 5 著作者：藤永芳純 島恒生 ほか 4 6 名

出版：日本文教出版株式会社

「116 日文 道徳506 文部科学省検定済教科書 小学校道徳用 小学道徳 生きる力 5」

付録

3.2. タイポが与える単語修正や文章理解への影響



タイポが与える単語
修正への影響①.pdf



タイポが与える単語
修正への影響②.pdf



タイポが与える単語
修正への影響③.pdf

3.3. 単語（名詞）のみのケース



単語（名詞）のみ
のケース①.pdf



単語（名詞）のみ
のケース②.pdf

3.4. 変換確率「0.9」の文章で検証



変換確率「0.9」の文
章に関する資料①.p



変換確率「0.9」の文
章に関する資料②.p



変換確率「0.9」の文
章に関する資料③.p

3.5. その他の検証



その他の検証
資料①.pdf



その他の検証
資料②.pdf