

論文

ChatGPT生成エッセイと英語母語話者によるエッセイの比較 —コーパス言語学の視点から—

藤原 隆史

A Comparison between ChatGPT-Generated Essays and Essays Written by Native
Speakers of English:
An Analysis from the Perspective of Corpus Linguistics

FUJIWARA Takafumi

要 旨

近年登場した生成系 AI の一つに、OpenAI 社の ChatGPT がある。インターネット登場以来の革命的技術であるとして、様々な分野において注目を集めている。言語研究分野や教育分野においても、ChatGPT の及ぼす影響は小さくないと考えられる。本研究では、ChatGPT が生成した英文エッセイを英語母語話者が書いたものと比較し、ChatGPT によって生成された英文の特徴についてコーパス言語学の視点から考察した。結果として、英語母語話者と比べて総語数が多くなること、語彙レベルが高めになること、客観的表現を用いること等の特徴があることが明らかとなった。

キーワード

ChatGPT コーパス言語学 特徴語 ICNALE 英語教育

目 次

- I. はじめに
- II. 先行研究
- III. 分析方法
- IV. 結果と考察
- V. 教育的示唆
- VI. 結論
- 謝辞
- 注
- 文献

I. はじめに

2022年11月にOpenAI社が公開したChatGPT¹⁾は、人間の言葉にかなり近い自然な言語生成能力があることから、驚きをもって受け止められている^{2,5)}。インターネット登場以来の革命的技術とも言えそうなChatGPTであるが、あまりにも急激なAIの進化とそれを取り巻く状況の変化に、脅威を感じるのも無理はない^{6,7)}。もちろん、株式市場に与える影響も無視できない⁸⁾うえに、高賃金のホワイトカラーの仕事に与えるインパクトは小さくなく、ChatGPT等のLarge Language Models(LLMs)の導入によって、米国では労働力の8割が少なくとも10%の影響を受け、さらに、約19%の労働者は業務の少なくとも50%において影響を受けるとの予測もある⁹⁾。また、Offers デジタル人材総研の報告では、9割以上の人がChatGPTの登場をポジティブに受け止めつつも、働き方への影響があると答えた人は半数以上に上った¹⁰⁾。公的機関や一般企業での導入やルール作りも進んでおり¹¹⁻¹⁵⁾、教育現場でも生成系AIへの対応をめぐる議論が交わされている¹⁶⁻²⁰⁾。

以上のような状況の中で、ChatGPTが言語研究や教育に与える影響の大きさはもはや説明の必要がないほどであろう。Mizumoto & Eguchi(2023)²¹⁾やShidiq(2023)²²⁾はChatGPTが教育環境にもたらす影響の大きさを指摘しているし、Lund & Wang(2023)²³⁾は、ChatGPTをアカデミアの世界で用いる際の有用性だけでなく、不正使用をしないことの重要性といった倫理面での課題も指摘している。また、日本の大学もChatGPTの利用についてそれぞれの方針等を示している²⁴⁾。対応は大学によりばらつきがあるが、概ねChatGPTを生かした研究や教育活動を妨げないという方向性では一致しているようである。したがって、言語研究や教育への応用についても、ChatGPTをはじめとした生成系AIの利用は避けて通れない情勢となっていると考えられる。

一方で、ChatGPTの登場から間もないこともあり、とりわけ言語研究や教育分野でのChatGPTの活用に関する分析はまだ少ないのが現状である。本研究では以上を踏まえ、ChatGPTによって生成された英語の文章の量的・質的特徴について分析する。その上で、英語教育への応用の可能性についても何らかの示唆を得ることを目的とする。具体的に

は、ChatGPTによって生成された英文エッセイをInternational Corpus Network of Asian Learners of English(ICNALE)^{25, 26)}に収録されている英語母語話者による英文エッセイとの比較を通して分析する。

II. 先行研究

本章では、ChatGPTの技術的側面に簡単に触れた後^{註1)}、ChatGPTを教育に利用することを目的とした研究、また、ChatGPTによって生成された英文自体を分析した研究について触れる。

1. 自然言語処理モデルTransformer

ChatGPTは、OpenAI社が開発したChatbotであるが、そもそもGPTとはGeneral Pretrained Transformerの頭文字をとったものであり、これはTransformer²⁹⁾という自然言語処理モデルをベースとした深層学習モデルである。Transformerは、それ以前のConvolutional Neural Network(CNN)およびRecurrent Neural Network(RNN)や、RNNをベースとしたSequence-to-Sequence(Seq2Seq)等とは異なり、Attentionと呼ばれる機構のみを用いて、入力された学習用データの特定の部分のみに注意を向けさせて学習させる深層学習モデルの一種である³⁰⁾。Transformerを用いた自然言語処理では、それまでのCNN、RNN等のモデルと比べ格段に高速な学習速度を誇り、GPT以外ではGoogleが発表したBERT³¹⁾などが代表的な深層学習モデルとして知られている。ChatGPTに用いられているGPTは、2023年5月現在でGPT-4まで進化しているが、GPTは自己回帰モデルと呼ばれておりテキスト生成を得意としている³²⁾。ただし、ChatGPTによって生成された文章には事実と異なる内容が含まれている場合があることが指摘されている³³⁻³⁵⁾点には注意が必要であろう。本研究では、ChatGPTが生成する英文テキストの言語的特徴に注目しており、英語母語話者による英文エッセイと比較することでこの点について考察する。

2. ChatGPTの言語研究と教育分野への応用

ChatGPTの発表からそれほど時間が経っていないため、ChatGPTを教育へ応用することに関する研究、あるいはChatGPTが生成した英文に関する研究はあまり多くはない。数少ない研究のうち、Mizumoto & Eguchi (2023) では、GPT (text-davinci-003 model) を用いて英語のエッセイを採点させる試みについて論じている。英文エッセイの採点にはかなりの労力がかかるため、コンピュータを用いた自動採点 (automated essay scoring) の試みが行われてきたが、Mizumoto & Eguchi (2023) では、ETS Corpus of Non-Native Written English³⁶⁾ と IELTS TASK 2 Writing band descriptors (public version) を用いて、GPT に TOEFL のエッセイと IELTS ルーブリックを示しながらエッセイを採点させる指示 (prompt) を与えている。そして GPT による採点の結果を、元々 ETS Corpus of Non-Native Written English につけられていたプロの採点者によるスコア (low, medium, or high) と比較し、GPT による採点の信頼性を検証した。結果として、採点結果に若干の誤差があるものの、GPT を英文エッセイの採点に用いた場合、ある程度の精度があることが示唆された。

また別の研究として、竹ノ内 (2023)³⁷⁾ は、ChatGPT を利用した実用英語技能検定2級のライティング分析について報告している。この研究では、実用英語技能検定の HP に掲載されている6本の解答例を CEFR (Common European Framework of Reference for Languages)³⁸⁾ に基づいて評価するように ChatGPT に指示を与え、その評価内容を分析している。結果として、ライティングの評価点を算出させることはできなかったが、CEFR に基づく評価に関しては概ね妥当であったと述べている。ただし、ChatGPT によるライティングの評価を指導に生かすという観点では、他の分析ツールと組み合わせたり教師によるチェックを入れたりすること等が必要であるとしている。

ChatGPT が生成するテキスト自体に関する研究として、Guo et al. (2023) が挙げられる。Guo らは、英語と中国語の一般的な質問に対する人間と ChatGPT (GTP-3.5) による回答の質の違いを調査し

た。その結果分かった ChatGPT の特徴として、まとまりがあり明確な文章で回答すること、長く詳細な回答をすること、バイアスや有害な表現がないこと、質問に対する回答がデータ上存在しない場合は回答しないこと、事実とは異なる回答をすること等が報告されている。また、ChatGPT は回答の一貫性、客観性、形式的であること (Formality)、非感情的であること等から、まるで「一つの専門家集団」³⁹⁾ による回答のようであると述べられている。さらに Guo らは、ChatGPT が生成した回答の言語的特徴として、人間の回答よりも平均的な長さが長い (token 数が多い) こと、type 数が多いこと、type/token ratio (TTR) が高いこと、また、品詞の使用として名詞、動詞、限定詞、形容詞、助動詞、等位接続詞、不変化詞がそれぞれ人間の回答よりも多いことなどを報告している。

ChatGPT が生成した英文に関する研究として、Liu et al. (2023)⁴⁰⁾ も注目に値する。Liu et al. (2023) は、TOEFL や GRE 等のエッセイ課題として書かれたエッセイと同様の課題を複数の GPT モデル (gpt2-xl, test-babbage-001, text-curie-001, text-davinci-001, text-davinci-002, text-davinci-003, gpt-3.5-turbo) に与え、その結果として生成された英文の言語的特徴について調査を行っている。それによると、全体的な傾向として、GPT モデルが人間の書いたエッセイよりも長い文章を生成すること、最新のモデル程人間の書いたエッセイに肉薄するが、語彙の豊かさではまだ人間のエッセイが GPT モデルを上回っていること等を報告している。ただし、一部の GPT モデル (text-davinci-003, gpt-3.5-turbo) では人間のエッセイよりも文法的に複雑な文章を生成したと Liu らは述べている。

さらに、Kohnke et al. (2023)⁴¹⁾ は、教育における ChatGPT の有用性と欠点について考察している。ChatGPT が、目標言語における間違いの指摘と修正、様々なジャンルのテキストの生成、語の辞書の定義、例文や訳文の提示といったことを容易に行ってくれる点を評価しつつも、課題等における不正行為、英文エッセイ等のオリジナリティの問題、情報の引用元を提示しないこと、誤った情報の提示、事前学習に起因する文化的偏り^{注2)}等の欠点を指摘している。その上で、Kohnke et al. (2023) は、ChatGPT 等の生成系 AI を用いる際には、利用者は適切なデジタ

ルコンピテンスを備えているべきであるとしている。

本章では、(Chat)GPTを用いた英文エッセイ・ライティングの評価に関する研究、さらに、ChatGPTが生成したテキストと人間によるテキストの比較研究、ChatGPTの利点と欠点に関する考察について概観した。本研究では、特にChatGPTが生成した英文エッセイが、英語母語話者による英文エッセイとどのように異なっているかに注目している。上記で見たGuo et al.(2023)やLiu et al.(2023)の研究において一定の言語的特徴が示されているが、ChatGPTによって生成された英文の特徴については、まだかなりの調査・研究の余地が残されていると考えられる。

以上のことから、本研究では、ChatGPTによる英文エッセイと英語母語話者による英文エッセイを比較し、特に以下の点に注目して分析する。すなわち、①コーパス特性、②語彙レベル、③特徴語の3点である。次章では、これらの特徴を分析するための方法について述べる。

Ⅲ. 分析方法

本章では、本研究で用いる分析方法、分析対象、分析ツールについて述べる。

1. GPT

本研究で用いたChatGPTはGPT-3.5をベースとしたバージョンであり、2023年5月30日現在、OpenAIに登録をしたユーザーであれば無料で利用することが可能である。現時点で、上位互換のGPT-4.0をベースとしたChatGPTが利用可能だが、こちらのバージョンは有料であることと、GPT-4.0に搭載されているマルチモーダル認識等の高負荷のかかる高度な処理を伴うタスクでなければGPT-3.5でも対応可能であること等から、今回はGPT-3.5ベースのChatGPTを用いている。GPT処理性能の点については、将来的に更なる発展が見込まれるため、今後は最新バージョンのGPTモデルを用いた分析をする必要があり、本研究の今後の課題としたい。

2. コーパス

本研究が分析の対象としたデータセットは、ChatGPTによって生成された英文エッセイ30本、およびICNALEに収録されている英語母語話者による英文エッセイ100本である。ICNALEには、英語学習者によるエッセイだけでなく、比較のための英語母語話者によるエッセイも収録されている。本研究では、2つあるトピックのうち、Topic A: It is important for college students to have a part-time job.を選択した。Ishikawa(2013)⁴³⁾に基づき、ChatGPTに以下のpromptを与え、エッセイを生成させた。

(1)用いたprompt

Do you agree or disagree with the following statement? Use reasons and specific details to support your opinion. Topic: It is important for college students to have a part-time job. When writing your essay, clarify your opinions and show the reasons and some examples. The length of your essay must be from 200 to 300 words.

上記の指示を繰り返し、30本の英文エッセイを生成させた。ただし、生成は3回に分けて行ったこと(2023年5月6日、12日、13日)、また、1つのスレッドのみで連続して生成させたことにより、テキストの均質性が保たれていない可能性がある^{注3)}。以上のように、同一のトピックについて、ChatGPTによって生成された大学生がpart-time job (PTJ)をすることの是非に関する英文エッセイ(GPT_PTJ)30本(10355 words)、英語母語話者によるPTJの是非に関する英文エッセイ(ENS_PTJ)100本(22623 words)の合計130本が本研究で用いたデータセットである。

3. 分析ツール

本研究の分析には、コンコーダンスとしてAntConc 4.2.0⁴⁴⁾、ワードレベルチェッカーとしてNew Word Level Checker(NWLC)⁴⁵⁾、平均語数等の分析にVoyant Tools⁴⁶⁾を用いた。

具体的な分析としては、AntConcのkeyword分析ツールを用いて、target corpusをGPT_PTJ、reference corpusをENS_PTJに設定し、対数尤度比統計量(log-likelihood ratio, LLR)を基準として特徴語を抽出した。尚、閾値はBonferroni補正ありの $p < 0.01$ に設定した。また、同様の設定で特徴的低頻度語(negative keywords)も抽出した。

さらに、NWLCを用いたワードレベルの分析では、word listをNew JACET8000(分析の後半ではSVL12000も使用)に設定し、GPT_PTJとENS_PTJのそれぞれの英文エッセイ全文についてレベルチェックを行った。

平均語数の分析では、Voyant Toolsを用いて、GPT_PTJの30本とENS_PTJの100本の英文エッセイについて、token数、type数、TTR、1文の平均語数をそれぞれ算出した。

IV. 結果と考察

この章では、上記で述べた分析の結果とそれに基づく考察について述べる。

1. コーパス特性

本項では、各コーパスの特性について述べる。以下の表1は、GPT_PTJにおけるtoken数、type数、TTR、1文の平均語数をまとめたものである。ENS_PTJと比較するため、表の一番下にGPT_PTJだけでなくENS_PTJの平均値等も表示している。

token数の平均はGPT_PTJが345 wordsとなっており、ENS_PTJの226 wordsよりも多くなっている。しかし、そもそものライティングの指示が200 wordsから300 wordsの指定となっていることを考えると、ChatGPTは多くの場合でこの指示を無視していることになる。本研究では、このようなChatGPTの不規則な挙動を統制するのではなく、ありのままの英文生成過程を分析するという観点から、文字数オーバーの出力も分析の対象とした。なお、GPT_PTJのtoken数をみると、10番のエッセイまでは300 words以下のものもあるが、それ以降は300 wordsを超え、最終的には400 wordsを超えるものもあった。注の3で述べた通り、本研究では1つのスレッドのみで連続して指示を与えているため、

「会話履歴」がChatGPTの出力に影響した可能性がある。

Type数では、GPT_PTJが平均で約166 words、ENS_PTJが平均で約120 wordsという結果になった。また、TTRに関しては、GPT_PTJの平均が48.10、ENS_PTJの平均が53.18となった。コーパスサイズの違いを考慮に入れTTRの修正指標であるGuiraudの R を計算したところ、GPT_PTJが $R=8.93$ 、ENS_PTJが $R=7.99$ となった。すなわち、修正指標を採用した場合には、GPT_PTJの方がTTRは高いという結果になった^{註4}。この結果から、GPT_PTJのTTRの方が高く、ENS_PTJに比べ多様な語彙を用いてエッセイを構成していたことがうかがえる。後述の語彙レベルと合わせて考えれば、ChatGPTによって生成された英文は、英語母語話者と同等かそれ以上に難度の高い英文であると考えられることも可能であろう。

さらに、1文の平均語数については、GPT_PTJが18.54 words、ENS_PTJが27.93 wordsとなり、GPT_PTJの平均の方が少なくなった。

先行研究との比較という観点から結果を見てみると、GPT_PTJがENS_PTJよりもtoken数とtype数が多くなり、TTRも高くなったことについては、Guo et al.(2023)と同様の傾向を示しているといえる。一方で、Liu et al.(2023)の結果は人間によるエッセイの方がChatGPTによるエッセイよりも語彙的豊かさの点で上回っており、本研究はそれとは逆の結果となった。Guo et al.(2023)が一般的な質問に対するChatGPTの回答と人間の回答の比較、Liu et al.(2023)が英語テストのエッセイ課題に対するGPTモデルと人間の英文エッセイの比較であり、各研究における検証方法の条件が統制されていないため、単純な比較は難しい。今後は、実験デザインを統一して分析を行う必要があり、この点に関しては今後の課題としたい。

2. 語彙レベル

次に、GPT_PTJとENS_PTJの語彙レベルを比較した結果について述べる。以下の表2と表3はGPT_PTJとENS_PTJの各レベルの語の頻度をまとめたものである。

単語のカバー率という点からみると、ENS_PTJ

表1 GPT_PTJのtoken・type・TTR・1文の平均語数

	Token	Type	TTR	W/S*
1	245	127	51.83	22.3
2	308	130	42.20	22.0
3	266	117	43.98	20.5
4	298	130	43.62	19.9
5	308	139	45.13	20.5
6	306	122	39.86	17.0
7	293	128	43.68	15.4
8	291	138	47.42	17.1
9	326	134	41.10	19.2
10	238	119	50.00	21.6
11	325	182	56.00	18.1
12	338	182	53.84	17.8
13	355	171	48.16	17.8
14	335	177	52.83	17.6
15	336	181	53.86	17.7
16	387	188	48.57	18.4
17	363	183	50.41	18.1
18	422	200	47.39	21.1
19	383	192	50.13	19.1
20	404	194	48.02	17.6
21	374	189	50.53	17.0
22	367	188	51.22	19.3
23	366	167	45.62	18.3
24	412	188	45.63	17.9
25	404	188	46.53	16.8
26	395	192	48.60	18.8
27	411	197	47.93	17.9
28	387	183	47.28	17.6
29	353	177	50.14	16.0
30	359	178	49.58	17.9
GPT ave	345.16	166.03	48.10	18.54
ENS ave	226.23	120.32	53.18	27.93

* W/S = 1文の平均語数

表2 GPT_PTJの語彙レベル頻度表

Word Level	Type	Freq	%	Cum
PropNoun	1	16	0.15	0.15
L1	255	7851	71.29	71.43
L2	150	1783	16.19	87.62
L3	97	739	6.71	94.33
L4	40	244	2.22	96.55
L5	36	129	1.17	97.72
L6	14	67	0.61	98.33
L7	8	20	0.18	98.51
L8	4	20	0.18	98.69
NA	38	144	1.31	100
ALL	643	11013	100	100

表3 ENS_PTJの語彙レベル頻度表

Word Level	Type	Freq	%	Cum
PropNoun	56	109	0.48	0.47
L1	621	19955	87.54	88.02
L2	402	1702	7.47	95.49
L3	223	487	2.14	97.62
L4	86	163	0.72	98.34
L5	70	93	0.41	98.75
L6	41	54	0.24	98.98
L7	22	30	0.13	99.11
L8	22	34	0.15	99.26
NA	112	168	0.74	100
ALL	1655	22795	100	100

におけるカバー率はL1で88.02%、L2で95.49%、L3で97.62%、L4で98.34%となっており、L4の時点で読解の閾値と言われるカバー率98%⁴⁸⁾を超えている。一方で、GPT_PTJにおけるカバー率はL1で71.43%、L2で87.62%、L3で94.33%、L4で96.55%、L5で97.72%、L6でやっと閾値を超えて98.33%となっている。すなわち、GPT_PTJの方が、NWLCで判別可能な語彙だけで見た場合の語彙レベルが高い。

前述の通り、ChatGPTによって生成された英文と英語母語話者のそれとを比較した場合、New JACET8000の単語リストに基づくコーパス全体の読解難度はGPT_PTJの方が高くなっている。その一方で、GPT_PTJにおけるNA(not available)、すなわち語彙リストに記載のない単語に注目すると、GPT_PTJでは38語であるのに対し、ENS_PTJでは約3倍近くの112語に上っている。それぞれの単語の使用頻度は高くないものの、ENS_PTJでは読解難度の高い語彙が用いられている可能性がある。そこで、New JACET8000にてNAと判定された語について、NWLCのSVL12000の単語リストを用いて、再度語彙難易度を判定した^{注5)}。その結果、GPT_PTJで20語、ENS_PTJで49語がNAと判定された。以下の(2)と(3)に、それぞれのコーパスにおいてNAと判定された語を示す。尚、各単語の後ろの括弧内の数字は、Weblioによる語彙レベル(後述)を表している。

(2)SVL12000でNAとなった語(GPT_PTJ)

adaptability (12), barista (20), collaboratively (na), coursework (13), employability (na), employable (17), entrepreneurship (12), extracurricular(12), holistic(12), internship(12), mentorship(19), multitasking(15), overburden (15), preparedness (12), prioritization (20), prioritize(13), resilience(12), schoolwork(12), startup(9), transferable(14)

(3)SVL12000でNAとなった語(ENS_PTJ)

appropriateness(12), bollock(19), bursary(na), carrel (17), casework (20), cobblestone (14), cogitation (29), conducive (15), consumerist (na), counterproductive (12), coursework (13), coworker (11), cupidity (26), curricular

(13), disservice (14), distracter (30), dollop (15), existentially (na), extracurricular (12), fritter (23), fulltime (15), godsend (15), goof (14), internship (12), med (19), mesmerize (15), multi (13), multitasking (15), perversion (14), PhD (22), prioritize (13), schoolwork (12), solidify(13), studious(20), tedium(17), tertiary (12), togetherness (13), trundle (19), truthfully(12), undecided(12), undivided(13), unequivocally (14), unfairness (12), ungodly (18), uninteresting (12), unmotivated (18), unproductive(12), unread(13), unwind(12)

続けて、上記の(2)と(3)で示した語を、1052万語収録のWeblio英和辞典⁴⁹⁾で検索し、Weblioによる語彙レベル^{注6)}を参照して全ての語のレベルの平均を算出したところ、GPT_PTJの平均レベルが14.2、ENS_PTJの平均レベルが15.7となった。すなわち、NWLCの語彙リストを用いて算出可能な語彙レベルのみに焦点を当てた場合はGPT_PTJの読解難度が高めに出るが、NAに分類されたものも考慮に入れた場合には、ENS_PTJがより読解難度の高い語を用いている可能性が示された。Weblioのレベル分けは、レベル1からレベル30の30段階で、レベル12で「英検1級以上の単語」とされている。

以上を総合的に考えると、GPT_PTJの方がNWLCで測定可能な範囲での読解難度は相対的に高いレベルであるが、NWLCで測定不可能な語彙に関しては、ENS_PTJの方が読解難度の高い語が用いられていることが分かる。そのような語が用いられた場合には、読者によっては読解難度が極端に上がる可能性がある。

3. 特徴語

以下の表4にAntConcで抽出したGPT_PTJの特徴語が示されている。さらに、表5には同特徴的低頻度語が示されている。両者のLLRの平均値は、特徴語が48.1、特徴的低頻度語が44.6となっており、逸脱度はほぼ対称的であると言えそうである。

表4から、GPT_PTJでは英文エッセイによく用いられる語彙が特徴的に現れている。すなわち、firstly, secondly, conclusion, moreover,

表4 GPT_PTJの特徴語

Rank	Type	LLR			
1	skills	219.0	38	contribute	39.2
2	financial	212.3	39	professionals	37.2
3	can	132.7	40	loans	37.0
4	practical	127.8	41	gain	36.3
5	academic	127.6	42	expenses	36.3
6	students	126.4	43	moreover	34.8
7	their	120.1	44	with	32.7
8	professional	119.1	45	mentorship	32.5
9	provides	114.0	45	highly	32.5
10	management	110.5	47	additionally	32.5
11	independence	107.2	48	between	32.4
12	develop	106.0	48	strong	32.4
13	and	90.2	50	responsibility	32.2
14	valuable	88.7	51	field	30.8
15	essential	85.1	52	prioritize	30.3
16	opportunities	72.2	53	knowledge	30.2
17	connections	67.0	54	fosters	30.2
18	commitments	65.1	55	apply	29.8
19	offers	63.9	56	personal	29.7
20	experience	62.6	57	career	29.0
21	communication	59.4	58	development	28.2
22	responsibilities	58.5	59	face	28.1
23	challenges	58.1	60	marketing	27.9
24	firstly	54.0	60	insights	27.9
25	crucial	52.6	60	abilities	27.9
26	provide	51.8	63	problem	27.8
27	enhances	51.1	64	job	27.0
27	solving	51.1	65	balance	26.8
29	networking	50.4	66	careers	26.5
30	secondly	50.2	67	having	26.2
31	prepares	46.5	68	colleagues	26.1
32	conclusion	46.0	69	hold	25.9
33	enhance	44.1	69	support	25.9
34	teamwork	43.6	71	theoretical	25.6
35	network	43.4	71	establish	25.6
36	future	42.6	73	success	24.8
37	reduce	39.5	74	help	24.6
			75	effective	24.0

表5 GPT_PTJの特徴的低頻度語

75	burden	24.0
77	build	23.7
78	financially	23.2
78	develops	23.2
80	them	23.1
81	balancing	22.9
82	why	22.6
83	statement	22.2
84	classroom	22.0
85	growth	22.0
85	ensure	22.0
85	helps	22.0
88	overall	21.5
89	juggling	20.9
89	interpersonal	20.9
89	reliance	20.9
89	deadlines	20.9
93	several	20.4
94	while	19.1
95	budgeting	18.6
95	textbooks	18.6
95	alleviate	18.6

Rank	Type	LLR
1	i	172.4
2	have	130.2
3	my	103.6
4	do	85.4
5	very	58.3
6	at	57.3
7	that	57.1
8	will	56.3
9	is	49.9
10	of	49.3
11	people	46.6
12	all	45.2
13	to	41.2
14	school	38.0
15	good	36.9
16	some	35.6
17	not	34.5
18	way	34.4
19	the	33.8
20	they	31.1
21	most	30.1
22	get	28.6
23	should	28.3
24	much	27.2
25	new	26.5
26	want	25.1
27	jobs	24.6
28	money	24.6
29	like	24.4
30	parents	24.1
31	great	23.0
32	able	22.3
33	one	20.6
34	just	19.5
35	than	17.4

additionally 等であり、ChatGPT はこれらのディスコースマーカーを用いて定型化されたエッセイを生成していることが分かる。また、上位20位までの特徴語がどのような文脈で用いられているかをよく見てみると、ChatGPT がよく用いるエッセイのプロットが浮かび上がる。すなわち、アルバイトを通して得られる重要なスキル (valuable skills) を構築 (develop) することの大切さ、実践的な経験 (practical experience)、経済的自立 (financial independence)、時間のやり繰り (time management)、人脈を生かすチャンス (connection with professionals / networking opportunities) の重要性等について語られており、学業成績や学生としての責任 (academic performance / responsibilities / commitments) を果たしながらアルバイトをすることができれば、アルバイトがそれらをもたらして (offers) くれる、といった流れである。

表5の特徴的低頻度語に注目すると、ChatGPT は主観的な表現を避ける傾向があることが見て取れる。すなわち、1人称代名詞 (I, my) の使用が少なく、ENS_PTJ の特徴語である you, we, me, your, our, think, feel 等の人称代名詞、思考や感情を表す動詞の使用は0回である。すなわち、一般の人や特定の人物を主語にして人々の言動について言及することがなく、物事を主語としてその良し悪しを語るような文脈が多いことが考えられる。また、助動詞 can がよく用いられる一方で、will, should, could は使用頻度が低く、would に至っては一度も使用されていなかった。これらの話し手のムードを表す法助動詞があまり用いられないということから、ChatGPT が何らかの主観的な判断をするような表現をあまり用いていないことが考えられる。

その他の興味深い点として、前置詞の使用状況が挙げられる。ChatGPT は on, in, for, with 等の前置詞をそれなりに使用しているが、of の使用頻度が低く、at に至っては5回しか使用していない。また、so, because, if 等の接続詞の使用も0回であり、因果関係を表す表現や何かを仮定・想定するような話は出てこない。さらに、等位接続詞の and、限定詞の their がよく用いられており、Guo らの観察と一部一致している。基本動詞 have, do の使用頻度が低いことも特徴的である。

4. まとめ

本章では、ChatGPT によって生成された英文の特徴を見てきた。コーパス特性については、token 数の平均、type 数、TTR の点で先行研究を支持する結果が得られたが、1文中の平均語数では英語母語話者のエッセイの方が多いという結果であった。語彙レベルに関しては、ChatGPT のエッセイでは相対的に高い語彙レベルではばつきがない一方で、英語母語話者のエッセイでは、使用頻度は低い語彙レベルの高い語が用いられていることが分かった。特徴語については、ChatGPT が英文エッセイでよく用いられるディスコースマーカーを頻繁に使用すること、主観的な表現をあまり用いないこと、人物に関する表現よりも物事に関する表現が多いことなどが分かった。

V. 教育的示唆

本研究では ChatGPT が生成した英文エッセイの特徴について考察してきたが、本章では ChatGPT の教育への応用という観点からいくつかの点について述べる。

まず、ChatGPT が生成する英文は、英語母語話者による英文よりも長くなる傾向がある。本研究では文字数を指定していたが、ChatGPT は多くの場合その指示を無視して英文を生成した。与える指示を工夫することで、ある程度統制することは可能かもしれない。また、語彙レベルについても、何の指示も与えずに生成させた場合は、英語母語話者が使用するものよりも高いレベルの語彙を用いる可能性がある。本研究では触れていないが、想定される学習者のレベルに合わせて prompt を工夫することで、一定程度使用語彙をコントロールすることは可能であろう^{註7}。特徴語については、人間によるエッセイよりも客観性が高まり、物事について述べる文脈が多くなる可能性がある。特に、人称代名詞や法助動詞、筆者の感情を表す語句の使用が減る可能性がある。また、本研究の分析では、因果関係や条件文などを表現する語句の使用が見られなかった。ただし、指示を工夫することや話題によってはこれらの語句を使用させることも可能かもしれない。

以上で述べたことは、本研究が使用したデータセッ

トを基にした考察であるが、今後のGPTのバージョンの変更やAIによる深層学習の進捗、指示の与え方等、様々な要素によって出力結果が変化していく可能性がある。確実に言えることは、生成系AIによって生成された英文を、その使用目的や対象とする学習者の特性に合っているか、教師が必ずチェックすべきであるということである。

VI. 結論

本研究では、2022年11月に一般に公開されたChatGPTに関して、生成された英文エッセイの言語的特徴を分析した。学習者コーパスICNALEに収録されている英語母語話者による part-time job の是非に関する英文ライティングのデータを用いて、ChatGPTが同様のトピックについてどのような英文を生成するかを比較し分析した。①コーパス特性、②語彙レベル、③特徴語の3つの観点で分析を行い、ChatGPTが英語母語話者よりも相対的に長い文章を生成すること、語彙レベルが学習者用語彙リストでの判定で高めになること、客観的な表現を用いて文章を構成する傾向があること等を明らかにした。

謝辞

本研究は、JSPS 科研費 JP22K00746 および JP20K00841 の助成を受けたものである。この場をお借りして感謝申し上げる。

注

- 注1 さらに詳しい技術的な説明については、近江 et al.(2021)²⁷⁾や坂本(2022)²⁸⁾等を参照のこと。
- 注2 GPTモデルは英語文化圏のテキストを大量に読み込み学習を行っているため、ユーザーによっては文化的な偏りを感じる場合があることが指摘されている⁴²⁾。
- 注3 ChatGPTはスレッドの「会話履歴」をテキスト生成に反映させることができるため。この点については、今後の研究で改善を試みる予定である。
- 注4 石川(2021)が述べている⁴⁷⁾通り、GPT_PTJの平均コーパスサイズが相対的に大きいため、修正前のTTRが下がったと考えられる。また、TTRの別の修正指標である Herdan のC値を求めたところ、GPT_PTJがC=0.87、ENS_PTJがC=0.88となり両者の差はほとんどなかった。
- 注5 NWLCで利用可能なワードリストの中で、SVL12000のリストには12000語が含まれており、New JACET8000の8000語よりも多いためSVL12000を用いた。また、SVL12000で再度チェックするにあたり、New JACET8000でNAとされた語の中から明らかな重複やスペルミスと思われる語等は排除している。
- 注6 Weblioによる各見出し語のレベル設定はWeblio独自のものであり、学術的根拠が示されていないことには注意が必要である。尚、(2)と(3)においてnaと表示されているものはWeblioによるレベル分けがされていない単語であり、高難度語である場合が多い。
- 注7 例えば、ChatGPTはCEFRのレベル分けに基づく語彙を使用するように指示することが可能である。ただし、正しく生成できるかについては今後検証していく必要がある。

文献

- 1) OpenAI, ChatGPT, <https://chat.openai.com/> (閲覧日2023.5.6)。
- 2) 「坂村健の目：対話AIの先にある怖さ」『毎日新聞』2022年12月15日付、東京朝刊、p.16。
- 3) 「憂楽帳：ChatGPTの衝撃」『毎日新聞』2023年2月2日付、東京夕刊、p.7。
- 4) 「ChatGPT『歴史転換、日本も生成AI開発を』東大・松尾氏—ChatGPTの衝撃 私はこう見る②」『日経速報ニュースアーカイブ』2023年2月16日付。
- 5) 「チャットGPTの衝撃、どう見るどう使う 専門家に聞く」『朝日新聞』2023年5月9日付、朝刊、科学1。
- 6) 「マスク氏ら1000人超『AI開発、半年停止を』『人類に深刻なリスク』」『朝日新聞』2023年3月31日付、朝刊、3総合。
- 7) 「『チャットGPT、全世界が実験台』『データ集中と監視、強まる恐れ』元グーグルの専門家が警告」『朝日新聞』2023年4月12日付、朝刊、1総合。
- 8) 「チャットGPTに対抗 グーグル対話型AI『回

- 答に誤り』で株急落』『週間エコノミスト』2023年2月28日付, p.2.
- 9) Eloundou T, Manning S, Mishkin P, Rock D, “Gpts are gpts: An early look at the labor market impact potential of large language models”, arXiv preprint arXiv 2303.10130 (2023).
 - 10) Offers デジタル人材総研, デジタル人材の意識調査レポート ChatGPT が与える業務への影響, p.5(2023), <https://bit.ly/41jXQOW> (閲覧日2023.5.29).
 - 11) 「損保、対話AIで効率化 東京海上、契約照会に回答案 情報保護と両立手探り」『日本経済新聞』2023年4月19日付, 朝刊, p.7.
 - 12) 「チャットGPT試験導入 関東の一部自治体で先行」『日本経済新聞』2023年5月2日付, 地方経済面東京, p.5.
 - 13) 「伊藤忠がチャットGPT導入 消費者向け活用検討」『日本経済新聞』2023年5月12日付, 朝刊, p.15.
 - 14) 「長野県庁でチャットGPTを試験使用」『日経産業新聞』2023年5月23日付, p.4.
 - 15) 「生成AI利用、初の条例 神戸市が制定 試験運用へ」『日本経済新聞』2023年5月25日付, 朝刊, p.5.
 - 16) 「『ChatGPT、学校現場の影響は』 対話型AIめぐり、県議会で初の質問／岐阜県」『朝日新聞』2023年3月9日付, 朝刊岐阜全県, 1地方.
 - 17) 「(時時刻刻)AI時代『もう後戻りできない』生成された回答、学生が添削する課題も」『朝日新聞』, 2023年4月21日付, 朝刊2総合.
 - 18) 「AI活用の禁止場面提示へ 学校での指針づくり開始」『日経速報ニュースアーカイブ』, 2023年5月16日付.
 - 19) 「教育、各国共通の課題にどう挑む 英国は学校予算最大に 一歩目のキーガン英教育相インタビュー」『日経速報ニュース』2023年5月23日付.
 - 20) 「生成AI、大学の活用術 立命館大学は学生が英訳分析」『日経速報ニュース』2023年5月27日付.
 - 21) Mizumoto A, Eguchi M, “Exploring the potential of using an AI language model for automated essay scoring”, Research Methods in Applied Linguistics 2(2), 100050(2023).
 - 22) Shidiq M, “The Use of Artificial Intelligence-Based Chat-GPT and Its Challenges for The World of Education; From The Viewpoint of The Development of Creative Writing Skills”, In Proceeding of International Conference on Education, Society and Humanity 1(1), pp.353-357(2023).
 - 23) Lund BD, Wang T, “Chatting about ChatGPT: how may AI and GPT impact academia and libraries?”, Library Hi Tech News 40(3), pp.26-29(2023).
 - 24) 「生成AI、大学に変革迫る」『日本経済新聞』2023年5月28日付, 朝刊, p.2.
 - 25) Ishikawa SI, “The ICNALE and sophisticated contrastive interlanguage analysis of Asian learners of English”, Learner corpus studies in Asia and the world 1(1), pp.91-118, (2013).
 - 26) ICNALE: The International Corpus Network of Asian Learners of English. <http://language.sakura.ne.jp/icnale/> (閲覧日2023.5.28).
 - 27) 近江崇宏・金田健太郎・森長誠・江間見亜利, 『BERTによる自然言語処理入門—Transformerを使った実践プログラミング—』オーム社(2021).
 - 28) 坂本俊之, 『作ってわかる! 自然言語処理 AI BERT・GPT2・NLPプログラミング入門』C&R研究所(2022).
 - 29) Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I, “Attention is all you need”, Advances in neural information processing systems 30.(2017).
 - 30) 布留川英一, 『BERT/GPT-3/DALL-E 自然言語処理・画像処理・音声処理人工知能プログラミング実践入門: 最先端のフレームワークの実力を試そう!』ボーンデジタル, p.27(2021).
 - 31) Devlin J, Chang MW, Lee K, Toutanova K, “Bert: Pre-training of deep bidirectional transformers for language understanding”, arXiv preprint arXiv 1810.04805(2018).
 - 32) 布留川英一, 『BERT/GPT-3/DALL-E 自然言語処理・画像処理・音声処理人工知能プログラミング実践入門: 最先端のフレームワークの実力を試そう!』ボーンデジタル, p.31(2021).
 - 33) Guo B, Zhang X, Wang Z, Jiang M, Nie J, Ding Y, Yue J, Wu Y, “How Close is ChatGPT to Human Experts? Comparison Corpus, Evaluation, and Detection”, arXiv preprint arXiv 2301.07597, p.6(2023).
 - 34) Retdian N, 「工学教育におけるChatGPTを考える」『工学教育』71(3), p.1(2023).
 - 35) 「深層学習では見えぬ論理 囲碁AIに思わぬ盲点、人間に敗れる」『日本経済新聞』2023年5月23日付, 夕刊, p.12.
 - 36) ETS, Corpus of Non-Native Written English, <https://catalog.ldc.upenn.edu/LDC2014T06> (閲覧日2023.5.30).
 - 37) 竹ノ内朋子, 「ChatGPTを利用した英検2級のライティング分析」『外国語教育メディア学会(LET) 関西支部メソドロジー研究部会 報告論集』14, pp.63-69(2023).
 - 38) Council of Europe, Common European Framework of Reference for Languages (CEFR), <https://www.coe.int/en/web/common-european-framework-reference-languages> (閲覧日2023.5.30).
 - 39) Guo B, Zhang X, Wang Z, Jiang M, Nie J, Ding Y, Yue J, Wu Y, “How Close is ChatGPT to Human Experts? Comparison Corpus, Evaluation, and Detection”, arXiv preprint arXiv 2301.07597, p.6(2023).
 - 40) Liu Y, Zhang Z, Zhang W, Yue S, Zhao

- X, Cheng X, Zhang Y, Hu H, “ArguGPT: evaluating, understanding and identifying argumentative essays generated by GPT models”, arXiv preprint arXiv 2304.07666 (2023).
- 41) Kohnke L, Moorhouse BL, Zou D, “ChatGPT for Language Teaching and Learning”, RELC Journal 00336882231162868 (2023).
- 42) 同上, p.9.
- 43) Ishikawa SI, “The ICNALE and sophisticated contrastive interlanguage analysis of Asian learners of English”, Learner corpus studies in Asia and the world 1(1), p.97 (2013).
- 44) Anthony L, AntConc (Version 4.2.0) [Computer Software], Waseda University (2022). <https://www.laurenceanthony.net/software> (閲覧日2023.5.23).
- 45) 水本篤, 「New Word Level Checkerの概要」『外国語教育メディア学会 (LET) 関西支部メソドロジー研究部会報告論集』12, pp.1-24 (2022).
- 46) Voyant Tools. <https://voyant-tools.org/> (閲覧日2023.5.30).
- 47) 石川慎一郎, 『ベーシックコーパス言語学第2版』ひつじ書房, p.143 (2021).
- 48) Schmitt N, Jiang X, Grabe W, “The percentage of words known in a text and reading comprehension”, The Modern Language Journal 95(1), pp.26-43 (2011).
- 49) GRAS GROUP, Weblio英和辞典・和英辞典, <https://ejje.weblio.jp/> (閲覧日2023.5.31).