A Study on Differences between GPT-generated and ENS-written Essays on Doing a Part-time Job and Smoking in a Restaurant: Utilizing Corpus Linguistics Methodologies

Fujiwara, Takafumi

(Citation)
Learner Corpus Studies in Asia and the World,6:39-58

(Issue Date)
2024-03-20

(Resource Type)
departmental bulletin paper

(Version)
Version of Record

(JaLCDOI)
https://doi.org/10.24546/0100487714

(URL)
https://hdl.handle.net/20.500.14094/0100487714

A Study on Differences between GPT-generated and ENS-written Essays on Doing a Part-time Job and Smoking in a Restaurant
—Utilizing Corpus Linguistics Methodologies—

FUJIWARA, Takafumi (Matsumoto University)

Abstract

The present study focused on the differences between GPT-generated essays and essays written by English native speakers, seeking some implications for English education. In the analysis utilizing the corpus linguistics methodologies, the 100 GPT-generated essays and the 400 essays written by English native speakers (contained in the ICNALE) were compared in terms of the use of some function words such as conjunctions, personal pronouns, prepositions, and modal verbs. The themes of the essays were "the pros and cons of college students doing a part-time job" and "the pros and cons of smoking in restaurants", which are the topics used in the ICNALE. The results showed some points, namely, ChatGPT tends to generate the same phrases repeatedly, using the words such as *can*, *and*, *by*, *their*, *with*, and, *for* many times. In contrast, it generates some negative keywords such as conjunctions (*if*, *that*, *or*, *when*, and *as*), personal pronouns (*I*, *they*, *my*, *it*, and *us*), prepositions (*to*, *as*, *before*, *about*, and *since*), modal verbs (*will*, *would*, *might*, and *could*), and adverb particles (*up*, *out*, and *off*). The present study analyzed the function words above in detail, yielding some implications for utilizing ChatGPT in language education in EFL classrooms.

Keywords

ChatGPT, ICNALE, Corpus Linguistics, Contrastive Analysis, Academic Writing

1. Introduction

On November 30, 2022, ChatGPT was launched by OpenAI Inc. Since then, it is no exaggeration to say that ChatGPT has continued to influence almost every sector in the world. Due to the advent of the latest AI technology and its availability to wide-ranging fields, not only natural language processing but also theoretical linguistics, corpus linguistics, translation studies, test theories, and other relevant fields are very active. Language education is no exception, and it can be said that English education in Japan is also facing a turning point. In the field of English education in Japan, researchers,

teachers, practitioners, and others try to work on the application of ChatGPT to language learning and teaching. Especially, the field of English for academic purposes (EAP) including academic writing is most likely to be affected by ChatGPT and other similar AI technologies. Under such a situation, this study aims to reveal the traits of GPT-generated essays by utilizing the techniques of corpus linguistics analysis and comparison between GPT-generated essays and essays written by English native speakers (ENS). The present study will compare the essays excerpted from the International Corpus Network of Asian Learners of English (ICNALE) and the essays generated by ChatGPT on the same themes as those from the ICNALE. By comparing both, the present study will analyze some keywords of the essays generated by ChatGPT as well as those of the essays written by ENS, offering some implications for language learning and teaching in English as a foreign language (EFL) settings such as Japan.

2. Previous Studies

This chapter reviews the previous studies on ChatGPT and its application to language learning and teaching. In addition, some studies on linguistic traits of GPT-generated texts are also reviewed. The chapter ends with the objective of the current study.

As an overall trend, some studies on the application of ChatGPT to language education are emerging though, there are not enough studies on more specific descriptions of linguistic traits observed in GPT-generated texts such as argumentative essays. That is, some studies feature linguistic traits of GPT-generated texts, but they deal with relatively general and overall features, and the scope of the linguistic items ranges broadly from content words to function words. As far as content words are concerned, it is predicted that content words will be strongly influenced by the theme of texts, and the possibility cannot be denied that the analysis of content words could be an analysis of the theme of texts and not of language itself. Function words, on the other hand, will be less influenced by the theme of texts because they occur everywhere in English texts. Therefore, the present study tries to focus not on content words, which will rely heavily on the theme of the essay, but on function words that are ubiquitous in English texts including academic essays, questions and answers, or language textbooks.

2.1 Mizumoto and Eguchi (2023)

Mizumoto and Eguchi (2023) investigate the usability of GPT (text-davinci-003 model) for automated essay scoring. In their study, Mizumoto and Eguchi (2003) analyze the accuracy of scoring by the GPT model by comparing the results of the scoring and those of human raters on ETS Corpus of Non-Native Written English (TOEFL). The rubric

used was the IELTS TASK 2 Writing band descriptors (public version). The GPT model scored the essays written for the task in TOEFL into three levels (low, medium, or high) based on the IELTS scoring criteria. The results show that though there are some slight gaps between the scores, the scoring of the GPT model was not fatally different from that of professional human raters, indicating the GPT model could be used in task scoring.

Though the study of Mizumoto and Eguchi (2023) is on how to apply the GPT model to language education in terms of essay scoring, and it is not directly concerned with the present study that is aimed at analyzing linguistic traits of GPT-generated essays, it is worth noting that the study of Mizumoto and Eguchi (2023) confirms the high linguistic potential of the GPT model. This means that as Mizumoto and Eguchi (2023) reveal, the GPT model could have adequate language *competence* that bears a fully-fledged examination. Therefore, the GPT model and GPT-generated essays can be comparable to essays written by humans, ensuring that the question of the present study is worthwhile to inquire about.

2.2 Gut et al. (2023)

Guo et al. (2023) analyze the quality of the words used by ChatGPT in its generation. The study used ChatGPT (GPT-3.5) to generate the answers to general questions made on an Internet question site in China. The generated answers were compared to the actual answers by humans. Based on the investigation, Guo et al. (2023) report some points, namely, ChatGPT tends to generate relatively long answers consisting of many tokens, the answers contain numerous word types, contributing to the high type/token ratio (TTR), and there are more nouns, verbs, determiners, adjectives, auxiliary verbs, coordinating conjunctions, and infinitives compared to human written answers. Concerning function words, though Guo et al. (2023) have revealed that there are more determiners, auxiliary verbs, coordinating conjunctions, and infinitives than texts written by humans, they have not adequately argued more specific individual function words and their traits in GPT-generated texts.

2.3 Mahyoob et al. (2023)

Mahyoob et al. (2023) propose a framework for the evaluation of AI-generated academic writings, including GPT-generated ones, in terms of the following six perspectives: Relatedness, Adequacy, Limitation, Authenticity, Cognition, and Redundancy. According to Mahyoob et al. (2023), by using the framework, researchers can evaluate the capabilities of generative AIs like ChatGPT to generate academic texts. The research indicates that despite the tremendous capabilities of ChatGPT, there are

some defects such as “information repetition, nonfactual inferences, illogical reasoning, fake references, hallucination, and lack of pragmatic interpretation” (p. 282). The authors conclude that though ChatGPT can generate human-like, grammatically correct academic texts with well-formed structures, it fails to “produce well-formed academic texts and articles with authentic and reliable information” (p. 290), lacking pragmatical knowledge of social norms and values. As in the other previous studies, Mahyoob et al. (2023) seem not to investigate more specific individual function words, and there is ample reason why the present study should analyze function words in detail.

2.4 Herbold et al. (2023)

Herbold et al. (2023) investigate the capability of ChatGPT models (versions 3 and 4) to generate argumentative essays on various themes, which are subjected to students in secondary schools. Herbold et al. (2023) compare the essays according to the perspectives of lexical diversity, syntactic complexity, semantic properties, and discourse properties. According to Herbold et al. (2023), ChatGPT models tend to generate relatively longer essays. Furthermore, GPT-generated essays, the authors report, outperform human-written essays regarding the point of quality. Regarding linguistic devices, ChatGPT can use suitable discourse markers to structure the plot of the essay, but at the same time, the fact that ChatGPT models can use the discourse markers suitably does not necessarily ensure the logical coherence of the essays. This is because, Herbold et al. (2023) surmise, “due to the rigid structure of the generated essays . . . instead of using discourse markers, the AI models provide a clear logical structure by separating the different arguments into paragraphs” (p. 8). Besides the report above, as a limitation, Herbold et al. (2023) admit that the human-written essays Herbold et al. (2023) collected were written by non-native speakers of English, meaning that the validity and generalizability of the study could be affected by the fact. Moreover, the investigation indeed reveals that the quality of GPT-generated essays could be comparable to the quality of those written by humans, but the authors have not mentioned individual linguistic items used in the essays. This is where the current study may come in, that is, mainly on the point of more specific words and expressions, GPT-generated essays can be worth investigating.

2.5 Objective of the Present Study

The purpose of this study is to investigate the linguistic traits of the academic essays generated by ChatGPT. To achieve this goal, the present study employs the way of comparison between GPT-generated essays and essays written by ENS. By doing this,

the current study tries to reveal the differences between them, especially some differences in the usage of function words such as personal pronouns, conjunctions, prepositions, modal verbs, etc., which will be analyzed with the aid of the methodologies of corpus linguistics. Regarding the analysis of the words in essays, as mentioned above, it can be said that the keywords used in essays depend much on the theme. Since the study tries to analyze all the essays written by ENS included in the ICNALE, the content words must be affected by the themes of the essays, namely, the pros and cons of university students doing part-time jobs and the pros and cons of smoking in restaurants. In contrast, function words are considered to be used without any thematic influence, as Biber et al. (2002) point out "individual function words tend to occur frequently, and in almost any type of text" (p. 16). This is because function words convey not meanings but relationships between content words, and such roles occur in various texts regardless of theme. This is why the study investigates the function words in the essays. By doing so, some theme-independent knowledge can be gained, and the results can have some implications for academic essay writing in EFL classrooms. Therefore, the present study sets the following research question.

Research question: What are the linguistic traits of function words in GPT-generated and ENS-written essays on two writing topics of the pros and cons of university students doing part-time jobs and the pros and cons of smoking in restaurants?

3. Method

As stated above, the present study adopted a methodology of corpus linguistics along with ChatGPT. In this chapter, the details of the analysis will be presented, namely, the version of GPT, the target of the analysis, and the concordancer.

3.1 The version of GPT

The current study utilized the free version of GPT-3.5 Turbo to generate the essays. At the time of writing this paper, the latest version of GPT-4.0 is available, but there are some reasons to use version 3.5. First, GPT-3.5 Turbo is easy to access, meaning that users can use it for free. This could be one of the advantages for researchers to use GPT-3.5 Turbo, who perhaps have a limited budget for research projects. Second, text generation may not require the latest version of GPT, which can deal with other than texts such as pictures. In other words, GPT-3.5 Turbo will do well enough in dealing with simple text data. In addition, GPT-3.5 could be relatively less susceptible to updates when compared to GPT-4.0. OpenAI explains that the training data used for GPT-3.5 is up to September 2021, whereas that of GPT-4.0 was updated to the latest version in

December 2023[1]. This could mean that the linguistic ability of GPT-3.5 has stopped to change, whereas that of GPT-4.0 was altered relatively recently. Though this is consequential, the update could have affected the outputs of GPT if this study had chosen GPT-4.0 because the essays were generated on multiple occasions (to be explained in the next section in detail). The author judged that GPT-3.5 is more stable in terms of the invariability of the outputs than GPT-4.0, selecting the free version of GPT for the analysis. In any case, however, suppose the basic linguistic proficiency of GPT is maintained up to a certain level, either model will do enough to analyze linguistic data, and which version to choose could be a matter of preference.

3.2 The target of the analysis

The present study aims to compare the GPT-generated essays and essays written by ENS to analyze the linguistic characteristics of the essays. To achieve this goal, two themes that are used in the ICNALE were selected and the instruction was used as the prompt given to ChatGPT. The themes of the essays in the ICNALE are on two topics, namely, the pros and cons of university students doing part-time jobs and the pros and cons of smoking in restaurants. The actual prompt was adopted from Ishikawa (2013) as follows.

> (1) The 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 / Smoking should be completely banned at all the restaurants in the country. 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.
>
> (Ishikawa, 2013, p. 97)

The prompt is an altered version of the ICNALE, that is, the original instruction contains some directions such as “You can use 20 to 40 minutes for each essay” or “You must use MS Word or a similar word processor” and these directions do not apply to ChatGPT that can generate texts immediately after giving the prompt.

In this study, the GPT-generated essays were generated in a single chat one by one because it is said that ChatGPT can refer to the previous conversation, and this can influence the outputs of ChatGPT. By ordering ChatGPT to generate an essay in a chat 50 times for each theme, the author gained 50 essays on the pros and cons of university students doing part-time jobs and 50 essays on the pros and cons of smoking in

restaurants. The essays were generated in three separate sessions, i.e., three times in May 2023, June 2023, and January 2024. Consequently, the corpora were (i) 100 GPT-generated essays (token: 35484, type: 1954, TTR: 5.5) and (ii) 400 essays written by ENS included in the ICNALE (token: 90613, type: 5637, TTR: 6.2). In the keyword analysis, the present study sets (i) as the target corpus and (ii) as the reference corpus. Furthermore, it should be noted that all the essays generated by ChatGPT on the pros and cons of university students doing part-time jobs were arguing from a "pro-part-time work" position, and all the essays on the pros and cons of smoking in restaurants were arguing from an "anti-smoking" position. Though this implies that ChatGPT could have a biased generation algorithm or a data set, no extra instructions were given to ChatGPT to induce a position in the debate in order to observe simple responses to the same instructions given to ENS in the ICNALE.

3.3 The concordancer

The present study utilized AntConc 4.2.0 as the concordancer. Specifically, by utilizing the keyword analysis tool of AntConc 4.2.0, some of the (negative) keywords were extracted from the corpora. In the settings of the keyword analysis, the Likelihood Measure of Log-Likelihood (4-term) was selected, and the threshold was set at $p<.01$ (6.63 with Bonferroni correction). By using this setting, unusually frequent/negative keywords were gained. Note that this paper uses the term keywords to refer to unusually frequently used words, and the term negative keywords to refer to unusually less frequently used words.

4. Results

4.1 Keywords

In this section, the results of the keyword analysis are listed in Table 1. The keyword analysis gained six unusually frequent function words based on the Log-likelihood ratio, namely, one auxiliary verb, one coordinate conjunction, three prepositions, and one personal pronoun. Note that the items in the top column of the tables below stand for the following: Freq_Tar = the frequency in the target corpus, Freq_Ref = the frequency in the reference corpus, Range_Tar = the number of the GPT-generated essays in which the word occurred, Range_Ref = the number of the ENS-written essays in which the word occurred, and LL = Log-likelihood ratio.

Table 1
Keywords of GPT

Type	Rank	Freq_Tar	Freq_Ref	Range_Tar	Range_Ref	Keyness (LL)	Keyness (Effect)
can	3	569	503	98	251	296.176	0.031
and	23	1496	2531	100	400	158.278	0.076
by	36	268	253	81	178	125.595	0.015
their	129	464	792	95	294	46.071	0.025
with	160	291	468	97	260	36.961	0.016
for	173	572	1075	100	354	34.374	0.031

4.2 Negative keywords

Table 2 lists the results of the negative keyword analysis. The negative keyword analysis gained the following main five grammatical groups, namely, conjunctions (*if*, *that*, *or*, *when*, and *as*), personal pronouns (*I*, *they*, *my*, *it*, and *us*), prepositions (*to*, *as*, *before*, *about*, and *since*), modal verbs (*will*, *would*, *might*, and *could*), and adverb particles (*up*, *out*, and *off*). Among these, 341 of *that* were used as conjunctions and 4 of *that* were used as determiners, 40 of *as* were used as conjunctions and 95 of *as* were used as prepositions, and 400 of *to* were used as prepositions and 520 of *to* were used as *to*-infinitives.

Table 2

Negative keywords of GPT

Type	Rank	Freq_Tar	Freq_Ref	Range_Tar	Range_Ref	Keyness (LL)	Keyness (Effect)
i	1	142	1865	97	369	573.9	0.008
if	2	1	551	1	295	353.008	0
they	3	78	1083	45	323	344.122	0.004
that	4	345	2083	100	391	271.277	0.018
my	7	10	409	10	180	201.663	0.001
to	14	920	3488	100	400	126.527	0.046
it	17	259	1314	94	382	119.352	0.014
or	24	76	533	49	283	86.934	0.004
will	28	66	475	52	245	80.3	0.004
when	29	14	227	13	168	78.792	0.001
as	32	135	724	76	298	74.089	0.007
up	37	6	140	6	116	57.739	0
would	56	102	477	36	237	34.982	0.006

out	60	17	151	15	118	32.84	0.001
might	68	2	66	2	54	30.657	0
before	74	2	60	2	55	27.065	0
about	81	36	209	33	148	24.959	0.002
since	83	2	56	2	51	24.691	0
could	85	7	85	6	65	24.427	0
us	98	4	61	3	48	20.414	0
off	101	1	39	1	38	18.963	0

5. Discussion

This chapter consists of the main two parts, i.e., 5.1 Keywords and 5.2 Negative keywords. The latter has five sub-sections: Personal pronouns, Conjunctions, Prepositions and *to*-infinitive, Modal verbs, and Other negative keywords. Furthermore, in the following discussion, the current study will discuss the characteristics of GPT-generated essays in comparison with descriptions in some academic writing textbooks. This is because some of the features of GPT-generated essays show the same tendency as those in academic writing textbooks. That is, the current study hypothesizes that the linguistic features obtained from the analysis of large amounts of linguistic data, whether analyzed by the authors of academic writing textbooks or by the GPT, should to some extent exhibit similar linguistic features. In addition to this, it will be revealed that GPT-generated essays tend to follow the norms in textbooks more strictly than ENS. This might have something to do with the tendency that GPT always tries to generate sentences that adhere to the norms that were gained statistically from the pre-training data, while humans could be influenced by emotions and be inconsistent in their linguistic expressions.

5.1 Keywords

In Table 1, six unusually frequent function words in the essays of ChatGPT are listed, and the frequency of the modal verb *can* is the highest of all the function words. ChatGPT used the modal verb *can* unusually frequently compared to ENS. This could be attributed to ChatGPT's repetitive use of some fixed phrases. For example, ChatGPT tends to generate the following fixed phrases: "A *can* help B do C", "A *can* provide B with C", "A *can* create B", "A *can* lead to B", "A *can* contribute to B", or "A *can* be B". ChatGPT continuously used these phrases even in different essays, indicating that ChatGPT could have some patterns based on the training data of GPT. Moreover, it could be reasonable to surmise that *can* was used as a hedging device, which can avoid making 100 %

assertions and leave room for rebuttal of claims.

The second-highest item is the conjunction *and*. This appears to be consistent with Biber et al. (2002), which explains that the conjunction *and* is used more frequently in fiction and academic texts than in conversations and news. Biber et al. (2002) also maintain that, in academic writing, the conjunction *and* inclines to be used at the phrase level. Additionally, Wallwork (2011) points out that using *and* frequently at the clause level can make sentences too long, and this should be avoided. If ChatGPT is more faithful to the grammatical norms or tendencies above than ENS, it would explain why *and* was used more often than ENS essays. If this is the case, compared to ENS, ChatGPT could have generated many phrases in which *and* was used to parallel noun phrases that come in subject or object position in a sentence. For instance, the following patterns are frequently observed in the essays generated by ChatGPT: "both smokers *and* non-smokers", "such as respiratory problems, cardiovascular/heart diseases, *and* even cancer", "health *and* well-being", "personal *and* professional development/growth", "well-being *and* academic performance" and so on. The point here is that, as mentioned before, ChatGPT has a tendency to use some fixed phrases so often and this could lead to the high frequency of the occurrence of the conjunction *and*. In any case, as we do not know what data ChatGPT pre-trained on, we cannot draw any conclusions at this time, but the high frequency of occurrence of *and* should be taken into account if GPT-generated essays are to be used as model essays.

The next point worth noting is that the three prepositions, namely, *by*, *with*, and *for*, are used unusually frequently in the essays by ChatGPT. This high frequency of occurrences seems to be because of the repetition of fixed phrases containing the concerned prepositions. The following phrases are part of the examples excerpted from the corpus.

(2) Phrases containing *by*

"*by* working part-time", "*by* engaging in part-time work", "*by* implementing a complete ban on smoking", "*by* eliminating smoking from restaurants", "*by* banning/prohibiting smoking".

(3) Phrases containing *with*

"I agree *with* A", "A aligns *with* B", "A provides B *with* C", "A equips B *with* C", "individuals *with* respiratory conditions", "families *with* children", "commitments with", "associated *with*".

(4) Phrases containing *for*

"it is important/essential *for* college students to", "benefits/rewarding experience/valuable experience *for* college students", "A prepares B *for* C", "*for* instance", "*for* example", "*for* + *beneficiaries*/*experiencers*".

Although there are more patterns other than the excerpts, the phrases above were continuously used in the essays by ChatGPT. By looking closely at these phrases, it can be noticed that there are some fixed patterns in the generated essays. Therefore, when using a GPT-generated essay as a model essay, one should take into account the possible bias in the use of prepositions.

The fourth-highest grammatical item is the personal pronoun *their*. The high frequency of *their* could be derived from the generalization of the expressions in academic writing. That is, academic writing requires generalizing expressions in sentences, and it could be possible that the third-person plural personal pronoun *their* was used more frequently for this purpose. For example, the personal pronoun *their* referred to plural nouns such as "students, families, parents, customers, individuals, and non-smokers" in the GPT-generated essays. On the other hand, no personal pronouns such as *you*, *he*, and *she* appear in the GPT-generated essays. As a matter of fact, in the essays by ENS, personal pronouns such as *you*, *he*, or *she* are used in no small numbers. Specifically, *you*, *he*, and *she* appear 367, 42, and 18 times, respectively. From these observations, it can be said that ChatGPT tends to avoid using singular personal pronouns and use plural personal pronouns to make the sentences more general. In fact, the personal pronoun *they* is used 78 times in the GPT-generated essays.

5.2 Negative keywords

5.2.1 Personal pronouns

One of the traits of the negative keywords in the GPT-generated essays is that ChatGPT tends to avoid personal pronouns including the first-person singular, the second-person singular/plural, and the third-person singular. On the other hand, it is pointed out that the first-person pronouns can be used in particular academic fields such as engineering and hard science (Swales and Feak, 2012, p. 22), and ChatGPT generated some sentences that contain the first-person plural *we* occasionally (n = 141)[2)]. In addition to this, it is said that, in academic writing, personal pronouns tend to be omitted to achieve impersonality (e.g. Akbas and Hardman, 2017; Hyland, 2002). As speculations, the training data of GPT-3.5 might have contained impersonalized argumentative essays, and the training data might contain fewer informal expressions in the register of the

written texts such as academic essays/papers and technical books/journals. This could explain why ChatGPT avoids using personal pronouns listed in Table 2.

5.2.2 Conjunctions

As for the use of conjunctions, there seem to be clear differences between GPT-generated essays and ENS-written ones. For instance, the conjunctions of *if*, *that*, *or*, *when*, and *as* were used only sparingly, and the conjunctions *so*, *because*, *since*, *after*, and *before* were never used in the GPT-generated essays. ChatGPT seems to avoid using expressions of condition, causality, and forward-backward relationships in time[3]. Concerning the condition *if*, one possible explanation for the low frequency of occurrence is as follows. That is, if a conditional expression is used, one could be regarded as talking about assumptions and the argument could be less persuasive. This could lead to ChatGPT to avoid using the condition *if*.

Furthermore, some expressions giving explanations such as *because*, *as*, and *since* tend to be eliminated from the GPT-generated essays. ChatGPT never used *so*, *because* and *since* as a conjunction, and the conjunction *as* was counted only 37 times in the essays. Wallwork (2011) notes that by using the conjunction *since*, "readers are forced to carry an idea in their head before they understand how this idea relates to the idea in the main clause" (p. 42). Besides this, according to Wallwork (2011), to improve reader comprehension/readability, sentences should not be too long, and long subordinate and main clauses with *since* and *as* should be avoided. As a result, a main clause and a subordinate clause should be divided by using other linking devices. It could be possible to account for the low frequencies of the conjunctions in the GPT-generated essays by adopting the explanations above.

Regarding the conjunction *or*, though the conjunction *or* "occurs most often in academic writing" (Biber et al., 2002, p. 228), ChatGPT generated sentences including *or* less frequently contrary to this explanation. As mentioned in the previous section, the conjunction *and* was used unusually frequently in the essays, whereas *or* did not appear so often compared to *and*. This may need further investigation in greater detail of the corpora, including another comparison between, for example, the corpora and the British National Corpus (BNC). By comparing the corpora in question with another reference corpus such as BNC, it could be possible to know another possibility, say, overuse of it in the ENS-written essays. A brief exploration into the comparison between the ENS-written essays and BNC gained a result of frequent use of *or* in the ENS-written essays (LL = 89.77). However, some more detailed investigations should be addressed in the future because this may be beyond the scope of the present study.

Concerning the conjunction that can express causality, the conjunction *so* can be used mainly in conversations, whereas it is thought to be relatively less frequently used in academic writing (Biber et al., 2002). This seems to explain why ChatGPT generated no sentence with *so* in the essays. Moreover, in academic writing, *so* can be substituted by *therefore* when denoting causality. In fact, *therefore* was used 19 times. On the other hand, the conjunction *because* is a formal form of expressing causality and used in academic writing. Nonetheless, it was not used in the GPT-generated essays. This may be difficult to fully explain. This is because the present study only analyzes a limited number of essays on two topics, and it is necessary to look at the results of the analysis when ChatGPT is asked to generate more diverse academic texts that span a broader range of disciplines. One possible speculation could be that, as some of the previous studies point out, ChatGPT will generate sentences that contain false information or hallucinations and this could be concerned with this matter. Argumentative essays should be persuasive and based on solid information, otherwise, a statement about causality using *because* would contain false information and could harm society. That could affect the outputs of ChatGPT generation.

As another point, the conjunction expressing forward-backward relationships in time, namely, *after* and *before* was not used. These conjunctions can be used to express a chronological sequence of something, which can be seen in narratives or procedures. This could account for the reason why conjunctions expressing forward-backward relationships in time were not used in argumentative essays. Regarding the conjunction *when*, all of which were used not as relative adverbs or interrogative adverbs but as subordinating conjunctions, though the same explanation could be applied, perhaps it could be affected by the topics of the essay. This must be further analyzed in detail in the future.

Moreover, the conjunction *that* was used less frequently in the essays. Other than linguistic explanations in academic writing, this could be accounted for by thinking of the reference corpus, the ENS-written essays. The analysis of AntConc yielded the results of the overuse of *that* after verbs that denote an opinion, namely, *think that* (n = 246), *believe that* (n = 96), and *feel that* (n =74)[4]. From the observations, ENSs are inclined to use these verbs in academic writing, whereas ChatGPT seems to avoid using the verbs with *that*: *think that* (n = 0), *believe that* (n = 21), and *feel that* (n = 0).

5.2.3 Prepositions and *to*-infinitive

As observed before, ChatGPT generated numerous fixed phrases in the essays. Such phrases may consist of a variety of elements. Biber et al. (2002) point out that “phrasal-

prepositional verbs consist of a lexical verb combined with both an adverbial particle and a preposition" (p. 132). Biber et al. (2002) list the following phrases as an example: "*look forward to* has the lexical verb *look*, the adverbial particle *forward*, and the preposition *to*)" (pp. 132-133). This means that the usage of prepositions can be affected by co-occurring elements such as prepositional complements and phrasal-prepositional verbs, making it difficult to analyze the whole picture fully. Not only the preposition *to* but also the other prepositions listed in Table 2 could be affected by other elements of a sentence. Consequently, in this section, the present study will list the most frequent phrases ChatGPT generated in the essays, looking at some patterns in the phrases. The following (5) to (9) are parts of the extracted phrases frequently used in the essays. The numbers in the brackets denote the frequency of occurrence of each phrase, and regarding *before* and *since*, all the sentences observed are listed. Note that although the prepositions listed below are relatively numerous within each preposition, please keep in mind that they are NEGATIVE KEYWORDS when compared to the reference corpus.

(5) *to*

"contribute *to*" (73), "lead/led/leading *to*" (57), "exposed *to*" (33), "exposure *to*" (30), "risk *to*" (17), "contrary *to*" (14), "subjected *to*" (11), "not only *to*" (9), "but also *to*" (9), "due *to*" (8), "classroom *to*" (7), "appeal/appealing *to*" (4), etc.

(6) *as*

"such *as*" (54), "students *as*" (20), "society *as*" (9), "serve *as*" (8), "part-time *as*" (4), "as well *as*" (3), etc.

(7) *before*

a. Therefore, it is important for students to assess their own priorities and consider their personal circumstances *before* deciding whether to take on a part-time job during college.

b. It's essential for students to assess their own needs and priorities *before* making a decision about taking on a part-time job.

(8) *about*

"concerns *about*" (13), "message *about*" (6), "learn *about*" (5), "worry/worrying *about*" (3), "decisions *about*" (2), etc.

(9) *since*

a. For instance, countries like Ireland, New Zealand, and Italy have witnessed significant improvements in public health *since* introducing smoking bans.
b. *Since* then, there has been a significant decrease in heart attacks and respiratory problems among both smokers and non-smokers.

Concerning *to*-infinitives, according to Biber et al. (2002), "over 60 per cent of all to-clause occur in post-predicate position controlled by a verb" (p. 342). This could mean that the frequency of occurrence of *to*-infinitive may depend on the verb collocate with. In other words, to understand the more detailed conditions of occurrence of *to*-infinitive and the reasons for its infrequency, it is necessary to investigate the co-occurring content words, the verbs, in detail. However, it must be said that this is beyond the scope of the present study. This is why, along with the untouched points concerning the prepositions above, the present study would like to discuss the details in a future issue.

5.2.4 Modal verbs

According to Wallwork (2011), in academic writing, some impersonal expressions should be avoided. This is why modal verbs can be utilized to substitute, for instance, the personal pronoun *it*, which is tinged with an impersonal impression. Wallwork (2011) lists the following examples.

(10) ORIGINAL VERSION (OV)	REVISED VERSION (RV)
It is necessary/mandatory to use X.	X *must* be used.
	X is necessary/mandatory.
It is advisable to clean the recipients.	The recipients *should* be cleaned.
It is possible that inflation will rise.	Inflation *may* rise. (p. 82)

In addition to this, modal verbs can be used to tone down the level of probability (Wallwork, 2011). Despite these norms in academic writing, ChatGPT tends to use modal verbs less frequently than ENSs. Biber et al. (2002) may explain this. That is, each modal verb tends to appear in different registers. Biber et al. (2002) report that *will*, *can*[5], *would*, *could*, and *might* are more likely to appear in conversation than in academic writing. In academic writing, they insist, *may*, *should*, and *must* appear more frequently than in conversation. Regarding the results of the current study, as Biber et al. (2002) report, ChatGPT generated *will*, *would*, *might*, and *could* unusually less frequently, and this is consistent with Biber et al. (2002).

5.2.5 Other negative keywords

The present study reveals that three adverb particles were used unusually less frequently in the GPT-generated essays. In English, an action or occurrence can be expressed in two ways, i.e., a verb + particle/preposition and a single verb (Swales and Feak, 2012). According to Swales and Feak (2012), writers tend to use the latter way in a written academic style. This could be why ChatGPT avoided using the adverb particles in its generation. From the observation above, it would be possible to utilize an essay generated by ChatGPT as a model essay in EFL classrooms, subject to confirmation by the teacher.

6. Conclusion

The present study has investigated the linguistic traits of the GPT-generated essays by utilizing the corpus linguistics methodologies and the comparison between the target corpus and the reference corpus, the ENS-written essays included in the ICNALE. The focus was put on the function words in the corpus. The analysis revealed that there are some keywords and negative keywords that occur unusually frequently or less frequently in the corpus. The following are the main findings of the present study.

Keywords

ChatGPT tends to generate some fixed phrases repeatedly, and this may contribute to unusually frequent occurrences of the keywords such as the modal verb *can*, the conjunction *and*, the prepositions (*by*, *with*, and *for*), and the personal pronoun *their*. The generation is likely to be affected by some factors, namely, the register (academic or not), the level of occurrence (phrase or clause), and the generalization of expressions in sentences.

As an implication for academic writing, if teachers want to use a GPT-generated essay as a model essay, the usage of continuously generated linguistic items such as mentioned above should be checked. Otherwise, excessive usage of particular words and phrases can cause the learners to be familiar with an unbalanced vocabulary and phrases. Therefore, the outputs of ChatGPT should be checked correctly by teachers.

Negative keywords

Regarding the negative keywords, the current study gained the results of less frequently used function words in the GPT-generated essays. There are five main groups of function words, namely, the conjunctions (*if*, *that*, *or*, *when*, and *as*), the personal

pronouns (*I*, *they*, *my*, *it*, and *us*), the prepositions (*to*, *as*, *before*, *about*, and *since*) and *to*-infinitives, the modal verbs (*will*, *would*, *might*, and *could*), and the adverb particles (*up*, *out*, and *off*). As with negative keywords, these function words may be influenced by several factors. That is, the register (academic, narrative, procedures of something, or others), the impersonality of the expressions, the possibility of the overuse of the grammatical items in the reference corpus (the ENS-written essays), the topic of the essay (related to the register), co-occurring elements, and a stylistic characteristic.

Furthermore, some implications for academic writing regarding negative keywords should be mentioned. As argued above, the infrequency of the use of some negative keywords is a trait tied to the characteristics of the register. Although this study dealt with argumentative essays, the linguistic characteristics resulting from differences in the register should always be considered when outputting in English. This means that when using outputs of ChatGPT and other generative AIs as model texts, teachers should always pay attention to the relationship between the target words and phrases and the register in which these expressions are used. Moreover, the linguistic items targeted by teachers must be considered in relation to the words that collocate with them. This is because some function words such as conjunctions and prepositions could be influenced by content words such as verbs. If teachers want to make learners familiar with some target words and phrases, teachers should adjust the vocabulary and expressions used in the model essay by devising prompts and changing the theme of the essay.

The overall trend of the linguistic traits is congruent with the explanations/norms of the grammar textbooks for academic writing. This might be because the outputs based on the statistical processing of the training data might be consistent with the analyses and descriptions in the textbooks, as mentioned at the beginning of Chapter 5. On the other hand, the outputs of ChatGPT occasionally differ from the writings of humans. This could be assumed to be because the generation of ChatGPT should be rigorously based on statistical processing, whereas humans are more susceptible to emotions and sporadic preference for wording. This point will be discussed somewhere in another opportunity in the future.

Moreover, it must be noted that, since ChatGPT and its pre-training data are a black box, what has been described in this study is only one description. In other words, the homogeneity of the outputs of ChatGPT cannot be assured. As long as it is impossible to open the box, this can be a limitation for the present study and studies that use ChatGPT for any analysis. Furthermore, the issue of replicability does matter. That is, even if one

can replicate the current study to some degree, due to the fact of the uncertainty of the outputs, it is next to impossible to replicate the results completely. If the version of GPT is updated in the future, it is not unlikely that the outputs will change. Nevertheless, I believe that this study was able to provide some insight into the feasibility of using GPT-generated essays as model essays for learning and teaching academic writing in an EFL environment.

As the prospect for future studies, the issues left untouched in Section 5.2.3, as well as the underuse of some conjunctions such as *because* and *or* will be addressed in the future. Furthermore, by conducting some investigations into the use of ChatGPT in generating English texts, I would like to work on providing more detailed implications for language teaching and learning.

Notes

[1] https://platform.openai.com/docs/models/gpt-3-5-turbo

[2] The sentences that contain the first-person plural *we* can be seen ONLY in the essays about the pros and cons of smoking in restaurants. This can be a peculiar trait of ChatGPT and if the situation varies, the outputs could be altered. Besides this, it should be noted that, depending on the discipline or the requirements of the journal, there may be some fine rules in using the first-person plural *we*. (Wallwork, 2011)

[3] Note that *before* (n = 2) and *since* (n = 2) in Table 2 was used as prepositions.

[4] A further analysis of comparing the ENS-written essays with BNC revealed that the three verbs were used unusually frequently: *think* (LL = 761.712), *believe* (LL = 252.390), and *feel* (LL = 250.079).

[5] The present study observed an unusual frequency of *can* in the GPT-generated essays. This is not consistent with Biber et al. (2002) and needs further investigation to this point.

Acknowledgement

This study was supported in part by JSPS KAKEN Grants No. JP22K00746 and No. JP20K00841. The author would also like to express his gratitude to the two reviewers for their thought-provoking comments and suggested revisions to this study. However, any errors are the responsibility of the author.

Bibliography

Akbas, E., & Hardman, J. (2017). An Exploratory Study on Authorial (In) Visibility across Postgraduate Academic Writing: Dilemma of Developing a Personal and/or

Impersonal Authorial Self. *Online Submission*.

Biber, D., Conrad, S., & Leech, G. (2002). *Student grammar of spoken and written English*, Longman.

Guo B, Zhang X, Wang Z, Jiang M, Nie J, Ding Y, Yue J, Wu Y. (2023). How close is chatgpt to human experts? comparison corpus, evaluation, and detection. *arXiv preprint arXiv:2301.07597*.

Herbold, S., Hautli-Janisz, A., Heuer, U., Kikteva, Z., & Trautsch, A. (2023). A large-scale comparison of human-written versus ChatGPT-generated essays. *Scientific Reports*, *13*(1), 18617.

Hyland, K. (2002). Authority and invisibility: Authorial identity in academic writing. *Journal of pragmatics*, *34*(8), 1091–1112.

Ishikawa, S. I. (2013). The ICNALE and sophisticated contrastive interlanguage analysis of Asian learners of English. *Learner corpus studies in Asia and the world*, *1*(1), 91–118.

Mahyoob, M., Al-Garaady, J., & Alblwi, A. (2023). A proposed framework for human-like language processing of ChatGPT in academic writing. *International Journal of Emerging Technologies in Learning (iJET)*, *18*(14).

Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics*, *2*(2), 100050. DOI: https://doi.org/10.1016/j.rmal.2023.100050.

Swales, J. M., & Feak, C. B. (2012[1994]). *Academic writing for graduate students*. University of Michigan Press.

Wallwork, A. (2011). *English for writing research papers*. Springer.