The University of Osaka
Institutional Knowledge Archive

Title	計量文体論手法による英作文の語彙比較 ： 語学教育へのアプローチ
Author(s)	藤田, 郁
Citation	言語文化共同研究プロジェクト. 2024, 2023, p. 25-47
Version Type	VoR
URL	https://doi.org/10.18910/97315
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

計量文体論手法による英作文の語彙比較
語学教育へのアプローチ

藤田　郁

大阪大学大学院言語文化研究科

〒 560-0043 豊中市待兼山町 1-8
E-mail:　tabata@lang.osaka-u.ac.jp

あらまし　本稿は，日本の大学生が英語で書いた作文，機械翻訳を使って英語に翻訳した大学生による日本語作文，そして ChatGPT によって生成された英作文を比較分析するものである。計量文体論手法を適用し，学生に自ら考察させた結果も含め，分析と考察を行った。本研究では，テクストの分析と結果の提示だけでなく，テクストマイニングを通じて学生が各作文の相違点に気付き，機械翻訳や生成 AI の利用について内省することを目指している。機械翻訳と生成 AI の発展は利用する人間の批判的思考力の低下等，ことに教育の場面では問題や課題に直面している。生成 AI に限らず，機械翻訳ツール等も，非常に便利である一方，言語学習者の言語習得を妨げる可能性を否定できない。本稿は，学生が機械翻訳や生成 AI の使用について深く考え，適切な利用方法を見つけることを促すことを目的とした授業で扱った分析対象と結果を発展させたものである。

キーワード　ChatGPT, DeepL, 機械翻訳, 英作文, 計量文体論

Stylometric Methods for Lexical Comparison of English Compositions:
An Approach to Language Teaching

Iku Fujita

Graduate School of Language and Culture, University of Osaka

1-8 Machikaneyama-cho, Toyonaka, Osaka, 560-0043 Japan

Abstract　This paper analyzes three types of written English texts produced by university students in Japan, DeepL, and ChatGPT using stylometric approaches, cluster analysis, and correspondence analysis. The advent of generative AIs is astonishing and machine translators have made our lives easier these days. However, the negative impact on education has also been a concern and discussed. The author of this paper requested university students to write brief essays in English and Japanese about health in their “Data Science” class. The Japanese articles were then translated into English using DeepL. The target corpus comprises English essays written by students, translated by DeepL, and essays generated by ChatGPT. During the class, students analyzed the corpus using the same methods. This paper presents the results of the analyses and the students’ discoveries through text mining. It also discusses how these findings affect their understanding and awareness of the use of generative AIs and machine translators.

Keywords　ChatGPT, DeepL, machine translation, English writing, Stylometry

1. はじめに

本稿は，同一テーマのもとで英語学習者である日本の大学に在籍する学生の英作文と，日本語作文を機械翻訳を通して英語に翻訳した作文，及び ChatGPT (OpenAI, 2023) [1]に作例させた英作文に計量文体論手法を適用し，分析した結果を提示する。本稿で提示する分析及び結果の一部は，作文を執筆した学生らに自ら考察させた。テクストマイニング手法を通じ，各作文を客観的に分析，考察することにより，学生が各作文の相違点に気づき，機械翻訳や生成 AI の利用について内省を得られたかについても提示する。分析結果及び学生らの内省に基づき，特に語学教育の場面において機械翻訳や生成 AI との共存や活用について課題を提起したい。

近年，機械翻訳や生成（系）AI（以下生成 AI）の発展が目覚ましい。2022 年 11 月に 会話型 AI，大規模言語モデルの ChatGPT が OpenAI 社によって公開されて以降，その利用登録者は急速に拡大した。以下は，ChatGPT-4 (ChatGPT Plus プラン) に生成した ChatGPT 及び他の生成 AI に関する説明文[2]である。

> ChatGPT やその他の生成 AI は、自然言語処理を基にした技術であり、テキストや画像などのメディアを生成する能力を持っています。これらの AI システムは大量のデータを学習し、人間のような会話や文章を作成することが可能です。ChatGPT は特にテキスト生成に特化しており、質問に答えたり、物語を書いたり、さまざまなタスクを実行できます。一方、画像生成 AI は、ユーザーの説明に基づいて新しい画像を作り出すことができます。これらの技術は教育、エンターテイメント、研究など多岐にわたる分野で応用されています。

上記説明文を見る限りでは，文法の誤りも見受けられず，まるで人間が書いたかのような文章（や画像）を生成する ChatGPT の公開やその他生成 AI は，世界の人々を驚かせ，楽しませると同時に人間の領分を脅かす存在として世界を震撼させ，議論の的となっている。教育の分野や場に置いても例外ではなく，Rahman and Watanobe (2023: 1) では，教育における ChatGPT の与える影響について以下のように述べている。

> ChatGPT poses different threats to the traditional education and research system, including the possibility of cheating on online exams, human-like text generation, diminished critical thinking skills, and difficulties in evaluating information generated by ChatGPT.

教育の場における生成 AI 等の使用に関する懸案事項は押し並べて万国共通ではないかと考える。上記のような，学習者に直接関連する問題点に加え，著作権やセキュリティ上の問題点も指摘されており，大学等教育機関では使用方針を打ち出す等対応を余儀なくされた。大阪大学も例外ではなく，2023 年 4 月 17 日付で在学生へ向けた「生成 AI (Generative AI) の利用について」を公開している (`https://www.osaka-u.ac.jp/ja/news/topics/2023/04/17001`)。筆者が授業を担当したいくつかの教育機関でも，学生宛に生成 AI 利用に関するメッセージが 2023 年 4 月に発信され，いずれにおいても組織全体では生成 AI の

[1] 英作文生成は 2023 年に ChatGPT-4 を使用して行なった。

[2] 入力したプロンプト:‘Write a concise introduction about ChatGPT and other generative AIs in Japanese.’ (出力日: 12/03/2024)

利用を禁止しないが，授業や課題での利用可否はそれぞれの授業担当教員の指示に従うよう指示があった。

ChatGPT のセンセーショナルな公開とその拡大によって，教育の場での生成 AI の利用が問題になるより前から，語学教員を泣かせてきたのは Google Translate (`https://translate.google.com/`) や DeepL (`https://www.deepl.com/translator`) 等に代表される機械翻訳ではなかろうか。年々機械翻訳の精度が向上し，非常に便利なツールである一方で，学習者の言語習得を妨げる大きな要因となっていることは否めない。英語学習のための英語で書かれた教材の意味が分からないから，と機械翻訳に丸投げし，得られた日本語の翻訳文を読んで分かったつもりになるだけでは，目の前の課題や授業をやり過ごせても英語力は向上できない。また，インプットだけでなく英作文などのアウトプットをする場面においても，機械翻訳を使用する学習者は一定以上見受けられる。たとえ機械翻訳の利用を禁止し，機械翻訳によって作成された文章は教員側でわかるので，機械翻訳の使用が分かった時点で 0 点だと伝えても，機械翻訳によって作られた英文を提出してくる学生も残念ながら見受けられた。

授業態度や英語（や他の言語，科目）学習に対する動機づけ，教授法等，議論は多岐に渡る。また，機械翻訳や生成 AI をことに語学の授業で使用することの是非については様々な論があるかと考える。本稿は，教育学的知見から教授法や言語習得の過程について論じることを目的としていない。上記の経験から，教員側が気づく機械翻訳や生成 AI 等によって作成された文章と学生が書いた文章の相違点は何であるのかを具体的に学生に示すことにより，機械翻訳等の使用を抑制あるいは活用方法を考える機会を提供できるのではないかと考え，本稿の着想に至った。ただし，デジタル人文学，テクストマイニングの観点から，三種の英作文を比較分析することを目的とした授業であり，英語の授業ではないため，分析・考察後の学生の態度や行動，授業内外での機械翻訳や生成 AI の利用を追跡調査する等はしていない。

2. データ及び手法

2.1. データ

本稿で分析するデータは学生による英作文（以下 Student; ファイル名: ST），学生による日本語作文を筆者が DeepL を用いて英語に翻訳したもの（以下 DeepL; ファイル名: DL），そして ChatGPT に作成させたもの（以下 ChatGPT; ファイル名: CHN/CHS）の三種に大別される。いずれの作文も，テーマを「健康 (health)」とし，学生が作文する際には，英作文は 300 語前後（275–325 語），日本語作文は 700 文字前後（650–750 文字）の語数・字数制限を設けた。

作文の際には，作文と分析の目的を明示した上で，DeepL 等をはじめとする機械翻訳および ChatGPT 等の使用を厳禁とし，辞書はその媒体の種類（紙，電子辞書，Web 上）を問わず使用可とした。また，英作文を授業内で行ってもらい，日本語作文を授業外の課題としたため，日本語作文の提出率（81%）がわずかに英作文の提出率（88%）を下回った。本稿執筆

にあたり，作文を研究及び論文作成，分析へ使用すること及び文の一部を使用する点については受講学生の所属を含む個人が特定できる内容及び作文全文を公開しないことを条件に許諾を得ている。許諾を得られなかった学生の作文は分析の対象から除外している。また，分析に使用するデータの各ファイル名には番号が振られているが，コンピュータを用いてランダムに付与された数値であり，ファイル名内の番号が特定の学生を指し示すものではない。

ChatGPT による英語作文は，特に指定なく健康に関して書かせた作文 20 編（CHN）と，大学生として書かせた 20 編（CHS）に分けられる。作文作成の手順は，まず ‘health’ に関するエッセイのトピックを提示させ[3]，20 のトピックそれぞれを 150 語程度の作文にするよう指示した。ChatGPT は入力するプロンプトによって出力結果（の精度）が異なるため，プロンプト作成・入力等に関して本が出版される等活発に議論されているが，本稿で使用する作文は，大学生が気楽に ChatGPT に書かせて出力された作文を想定したため，あまり複雑ではないプロンプトを用いることとした。なお，学生に 300 語前後で作文をさせたのに対して ChatGPT には 150 語と指定したのは，作文出力前の試験的な出力で 300 語や 200 語で出力するよう指示をしても，指示より 100 語以上多い語数の作文が出力されてしまったためである。分析の際に各語の頻度は 1,000 語当たりの相対頻度にしたものを使用するため，各作文の語数は厳密に制限しなくて良い場合もある。しかし，300 語で起承転結が完結する作文と 600 語で完結する作文とでは，語の頻度だけでなく，内容の多様性も異なる可能性が考えられる。作文が長くなればなるほど，内容が多様になり，生起する（内容）語の種類も豊富になることが予想される。さらに，英語では，既出の語の頻繁な繰り返しを避ける傾向にあることから，内容が多様にならずとも，作品が長くなることにより語の種類が増える傾向にあると考えられる。以上のことから，分析する各作文の長さ（語数）は可能な限り近い値であることが望ましいと考えた。

以下 Table 1 に，作文（Student, DeepL, ChatGPT）ファイルをまとめた各コーパスのファイル数，総語数，異なり語数，そして Type Token Ratio (TTR; $TTR = \frac{\text{異なり語数}}{\text{総語数}}$) を示す[4]。ChatGPT のファイル数が他の Student, DeepL よりも少なく，総語数も半分以下であるが，TTR の値は 3 コーパスの中で最も高い割合となっている。

Table 1: 各コーパス情報

No.	コーパス名	ファイル数	総語数	異なり語数	TTR (%)
1	Student	77	17,518	2,095	11.959
2	DeepL	71	22,451	2,614	11.643
3	ChatGPT	40	8,713	1,580	18.133

[3] 入力したプロンプト: ‘Provide 20 essay topics related to health.’/‘Provide 20 essay topics related to health as a university student.’

[3] 入力したプロンプト: ‘Please write a 150-word essay on the topic **“The Impact of Mental Health on Physical Well-being.”**’/‘Please write a 150-word essay on the topic **“The Impact of Sleep Deprivation on Academic Performance”** as a university student.’ 注釈ない太字箇所には先に出力されたトピックを入れ替えて実行。

[4] 本稿で扱う各コーパスの作成及び頻度の計算や統計情報の抽出には CasualConc 3.0.7 (Imao, 2024) を使用している。

Table 2 は，各コーパスの上位 20 語とその素頻度と相対頻度をまとめたものである。いずれのコーパスでも機能語が上位を締めているものの，上位 20 語に一部内容語が含まれており，その数の多さは Student < DeepL < ChatGPT である。Student と DeepL では，*to* がコーパス内で最も頻度が高い語であるのに対し，ChatGPT では *to* は第 5 位に位置し，相対頻度で比較すると Student と DeepL の相対頻度の 6–7 割である。尚，Table 2 内では一人称代名詞 *I* を大文字で表記しているが，頻度の計算は全ての語を小文字にした状態で計算している。

Table 2: 各コーパス上位 20 語の素頻度及び相対頻度

	Student			DeepL			ChatGPT		
No.	語	素頻度	相対頻度	語	素頻度	相対頻度	語	素頻度	相対頻度
1	to	627	35.792	to	904	40.265	and	493	56.582
2	I	494	28.200	the	824	36.702	of	261	29.955
3	is	493	28.142	and	678	30.199	a	246	28.234
4	and	382	21.806	is	645	28.729	the	216	24.791
5	the	350	19.979	a	558	24.854	to	216	24.791
6	you	320	18.267	of	546	24.320	health	179	20.544
7	of	305	17.411	I	448	19.955	in	158	18.134
8	a	300	17.125	in	421	18.752	is	155	17.790
9	we	298	17.011	health	346	15.411	can	116	13.313
10	health	290	16.554	that	334	14.877	for	104	11.936
11	in	275	15.698	it	325	14.476	well-being	101	11.592
12	for	256	14.614	we	245	10.913	students	99	11.362
13	have	223	12.730	for	239	10.645	mental	88	10.100
14	it	205	11.702	are	218	9.710	it	73	8.378
15	our	179	10.218	be	203	9.041	stress	69	7.919
15	that	179	10.218	as	198	8.819	physical	63	7.231
17	not	164	9.362	not	191	8.507	their	63	7.231
18	are	163	9.305	healthy	175	7.795	university	63	7.231
18	so	163	9.305	have	170	7.572	as	56	6.427
20	do	142	8.106	sleep	159	7.082	are	55	6.312

DeepL, ChatGPT コーパスと比較し， Student コーパスに特徴的な語を単語頻度から見出すため，対数尤度比（Log-likelihood ratio; 以下 LLR）を計算し，LLR の値が高い上位 10 語及び値が低い下位 10 語を示したのが Table 3 – 6 である。Table 3 と 4 は DeepL コーパス，Table 5 と 6 は ChatGPT コーパスとの比較結果である[5]。

DeepL と比較した場合，Student コーパスに多く特徴的に現れる語は *you*, *so*, *think* と続き，LLR 上位 10 語中 4 語が人称代名詞（*you*, *your*, *our*, *I*）である。Student コーパスの頻度が DeepL と比較して少なく，最も特徴的な（LLR が負の値で最も低い）語は定冠詞 *the* であり，第 7 位に不定冠詞の *a* がある。下位第 4 位に *believe* がランクインしている。上位に *think* がランクインしていることから，日本語での「～だと考える/思う」に対応する英語を，学生の英作文では *think*，DeepL の翻訳では *believe* を使用する傾向にある可能性が考えられる。試しに筆者作例の日本語文「私は，健康な生活を送るためには睡眠が最も重要だと考える。」を DeepL にて英語に翻訳してみたところ，“I believe that sleep is the most important aspect of

[5] Log-likelihood の絶対値を降順で表示

a healthy life.” と，*believe* を使用した英文が出力された。また，「考える」を「思う」に変更しても同じ訳文となった。

Table 3: Student コーパスを基準とした対数尤度上位 10 位の語と値（DeepL コーパス）

単語	相対頻度 (Student)	相対頻度 (DeepL)	LLR
you	18.277	5.256	155.306
so	9.305	2.405	88.169
think	7.821	1.915	78.036
your	7.250	1.782	72.079
our	10.218	3.474	70.237
keep	3.368	0.534	46.744
can’t	2.512	0.223	46.113
lot	2.626	0.445	34.918
make	3.425	0.891	32.145
I	28.200	19.955	28.805

Table 4: Student コーパスを基準とした対数尤度比下位 10 位の語と値（DeepL コーパス）

単語	相対頻度 (Student)	相対頻度 (DeepL)	LLR
the	19.979	36.702	-100.199
one’s	0.114	3.207	-68.059
as	3.596	8.819	-44.114
believe	0.400	2.583	-34.084
would	0.457	2.628	-32.295
state	0.057	1.425	-29.619
a	17.125	24.854	-28.553
and	21.806	30.199	-27.309
of	17.410	24.320	-22.942
such	1.199	3.385	-21.008

Table 5: Student コーパスを基準とした対数尤度比上位 10 位の語と値（ChatGPT コーパス）

単語	相対頻度 (Student)	相対頻度 (ChatGPT)	LLR
I	28.200	0.000	403.570
you	18.267	0.000	260.338
we	17.011	1.607	158.595
so	9.305	0.000	132.117
think	7.821	0.000	110.975
do	8.106	0.115	105.313
people	8.106	0.115	105.313
have	12.730	1.607	105.273
your	7.250	0.000	102.850
if	7.193	0.115	92.552

ChatGPT との比較では，LLR が高く Student において多く使用されている上位 10 語には DeepL 同様人称代名詞および *think* がランクインしている。一方で，下位 10 語は，接続詞

and を除いて DeepL とは大きく異なり，内容語の中でも名詞，形容詞が多くランクインしている。Table 5 および ChatGPT に特徴的なのは，人称代名詞 *I*, *you*, *think*, *your* の頻度が 0 である点である。ChatGPT コーパス内にて他の人称代名詞と再帰代名詞の生起頻度を確認したところ，一人称および三人称単数人称代名詞 (e.g., *I*, *he*, *she* 等) 及び二人称代名詞 *you* はいずれの格も頻度は 0 であった。一人称および三人称複数の代名詞，再帰代名詞は使用されており，確認された代名詞の種類と頻度（丸括弧内スラッシュの左が素頻度，右が相対頻度）は以下の通りであった：*they* (20/2.295), *their* (63/7.231), *them* (9/1.033), *themselves* (7/0.803), *we* (14/1.607), *our* (31/3.558), *us* (1/0.115)。

Table 6: Student コーパスを基準とした対数尤度比下位 10 位の語と値（ChatGPT コーパス）

単語	相対頻度 (Student)	相対頻度 (ChatGPT)	LLR
and	21.806	56.582	-203.554
well-being	0.171	11.592	-198.627
students	0.856	11.362	-142.200
stress	1.199	7.919	-71.524
university	1.028	7.231	-67.814
chronic	0.114	3.787	-59.102
regular	0.114	3.443	-52.844
mental	2.911	10.100	-52.729
challenges	0.057	3.099	-51.748
maintaining	0.114	3.214	-48.694

2.2. 手法

本稿で使用する手法は，階層的クラスタリング（を付与したヒートマップ；以下 HM），コレスポンデンス分析（以下 CA）である。HM の作成には `Python`，CA の分析・描画には CasualConc (Imao, 2024) を用いた。

なお，授業内で学生に分析してもらう際に使用したのは，このうち HM と CA であり，分析や結果の描画には `Python` コードを用意し，Google Colaboratory (`https://colab.research.google.com/?hl=ja`) 上で実行してもらった。

3. 結果

3.1. 授業内分析結果（受講生考察内容）

まず，相対頻度上位 50 語を用いて階層的クラスタリングを付与したヒートマップを描画した（Fig. 1）[6]。本 HM では，縦軸に作文ファイル，横軸に上位 50 の単語が配置され，各作文ファイルにおける相対頻度の値が大きければ大きいほど，セルが濃い青で示される。Fig. 1 左上方に青いセルがまとまって現れており，列のデンドログラム（図上）および行のデンドログラム（図左）のどちらも一つのクラスターを形成している。ここには，DL_32 を例外として ChatGPT ファイルがまとまって現れており，クラスターに含まれている単語は図左

[6] 本稿の図は全て以下の URL にて PDF ファイルを閲覧可能：`https://www.dropbox.com/scl/fo/ia2x3xh74fcie1agv3vko/h?rlkey=vt360tnksus371osn5t81myq0&dl=0`

から順に *the*, *of*, *in*, *as*, *a*, *also*, *more*, *physical*, *health*, *mental*, *with*, *by*, *on*, *these*, *their*, *and*, *can* である。

ChatGPT クラスターの下方，図の中央付近には，Student と DeepL が混在しているものの，Student 作文がまとまっている傾向にある。Student に多く現れる傾向にある語のデンドログラムは，*healthy*, *life*, *we*, *our*, *exercise*, *body*, *is*, *important*, *good* のクラスター，*if*, *you*, *your*, *have*, *do*, *so*, *think* のクラスター，*at*, *when*, *I*, *my* のクラスター，そして *this*, *people*, *they*, *for*, *sleep*, *time* に分割されている。いくつかの内容語が各クラスターに含まれているが，概してこれらのクラスターは，一人称複数代名詞 *we*, 二人称代名詞 *you*, 一人称単数代名詞 *I*, 三人称複数代名詞 *they* のクラスターとして見ることができる。

Fig. 1 の下方には DeepL がまとまって現れる傾向にある。DeepL の作文の傾向は，上位 50 語の一部が ChatGPT や Student のように顕著に現れるというよりは，比較的満遍なく現れている点にある。図上部，語のデンドログラムの右側に位置するクラスターには *that*, *it*, *to*, *be*, *not*, *but*, *are*, *or* が位置しており，ChatGPT には顕著に現れない一方で Student, DeepL に現れる傾向にあると見ることができ，中でもどちらかと言えばわずかに DeepL に現れる傾向にある。

Fig. 2, 3 は，コーパス上位 50 語に CA を施した結果である。 Fig. 2 は 列座標の第 1, 2 次元，Fig. 3 は行座標の第 1, 2 次元を描画している。Fig. 2 では，各コーパスを色分けして表示している（コーラル: ChatGPT，ターコイズ: DeepL，ネイビー: Student）。

Fig. 2 では，図の中心に DeepL のファイルがまとまって現れる傾向にあり，Student ファイルは図の右側第 1 象限と第 4 象限に縦に散らばって現れている。 ChatGPT は，図左側，第 1 次元 -0.4 以下，第 2 次元 0.0 付近にまとまって分布している。このことから，図の右側は人間が書いた作文が位置し，左側には機械が書いた作文が位置していると見ることができる。

Fig. 3 を見ることにより，Fig. 2 で見られたファイルの分布の傾向の要因を見てとることができる。まず，ChatGPT がまとまって現れていた第 1 次元 -0.4 以下，第 2 次元 0.0 付近には，*mental*, *their*, *physical* といった単語が位置している。この 3 語は Fig. 1 で ChatGPT のクラスターを構成する単語に含まれており，それ以外の ChatGPT クラスターの語 *the*, *of*, *in*, *as*, *a*, *also*, *more*, *health*, *with*, *by*, *on*, *these*, *and*, *can* は第 1 次元 -0.4 ～ -0.2，第 2 次元 0.25 ～ -0.25 付近にまとまって現れている。第 1 象限の中でも右上端には *you* と *your*, 第 4 象限の右下方には *I* および *my* が位置していることから，縦の軸は上に行けば行くほど一般的な人々に言及した内容の作文（generic *you*）であり，下に行けば行くほど個人的なエピソードや意見を含む作文が位置していると考えられる。

以上の分析結果を授業内で学生に自分で `Python` を実行して描画してもらい，図の読み方を説明した上で，どのようなことが読み取れるかを考察してもらった。学生および筆者の考察結果は，次節の分析結果と合わせ，第 4 節にて提示する。

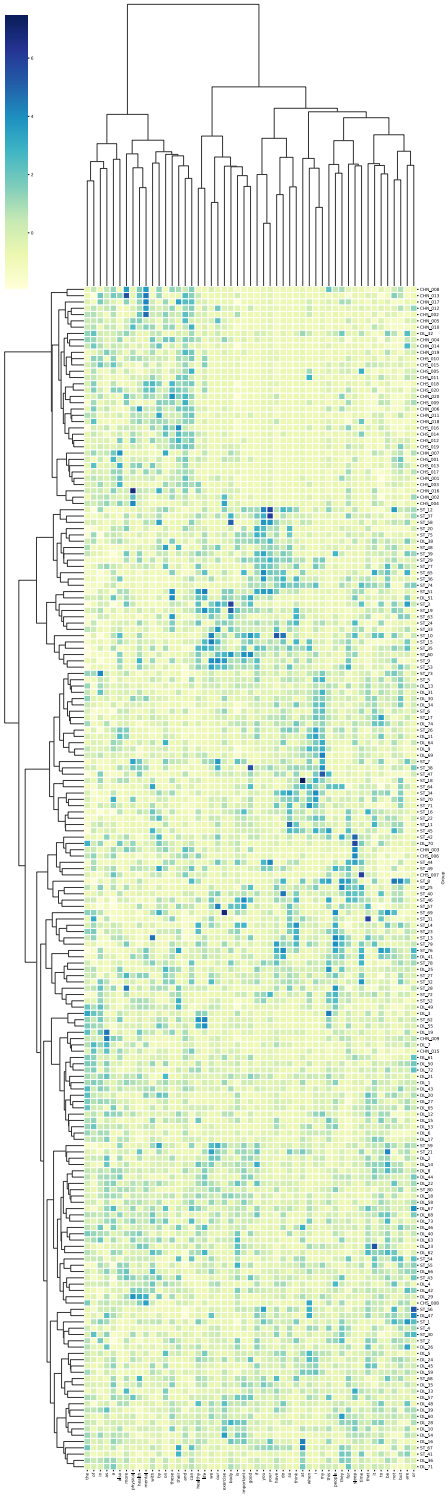

Fig. 1: 3 コーパス上位 50 語の HM 図 (Euclidian 距離, Ward 法)

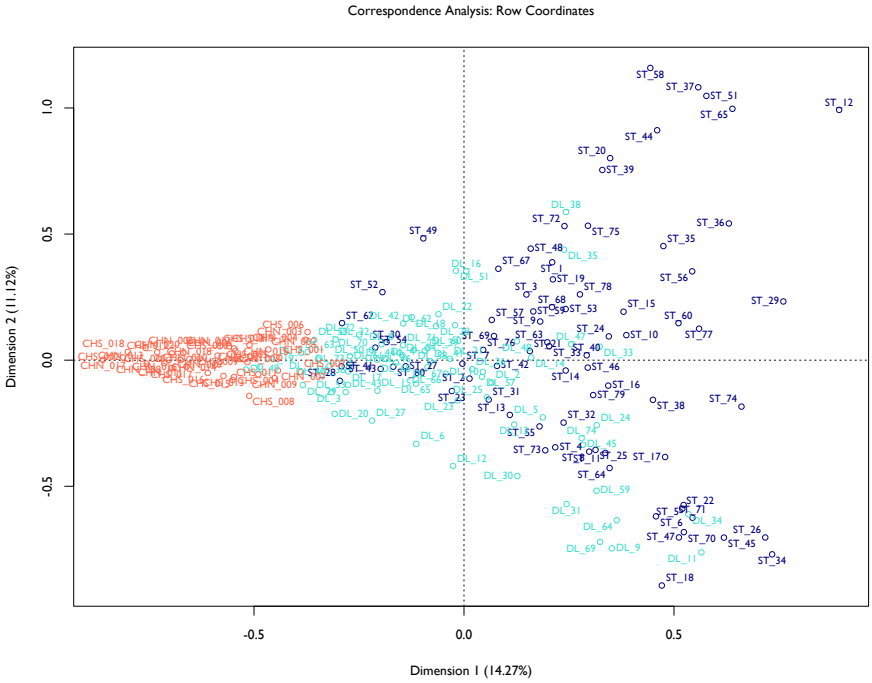

Fig. 2: 3 コーパス上位 50 語の CA (Row)

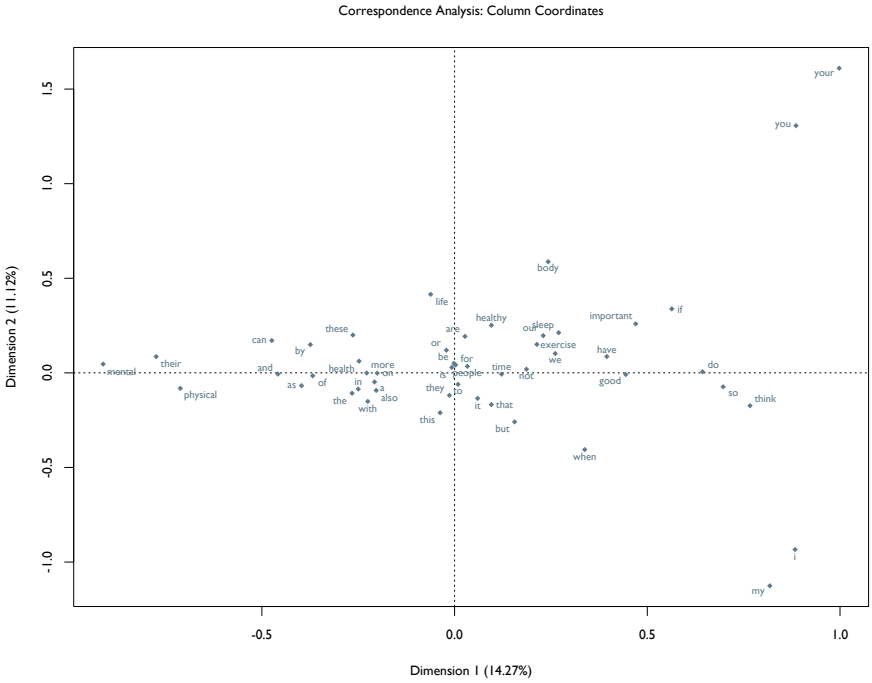

Fig. 3: 3 コーパス上位 50 語の CA (Column)

3.2. その他分析結果

次に，授業内で学生に提示しなかった分析結果についても示す。前節では，コーパス上位50語のみを対象としてHMおよびCAを行ったが，ここでは分析の範囲を拡大し，HMは上位100語（Fig. 4），CAは上位500語（Fig. 5, 6）を用いて分析を試みる。

Fig. 4は，Fig. 1同様，縦軸に作文ファイル，横軸に上位100の単語が配置され，各作文ファイルにおける相対頻度の値が大きければ大きいほど，セルが濃い青で示される。Fig. 4では，Fig. 1よりもChatGPTとそれ以外の作文が明確にグループ分けされており，図上部のクラスターはDL_70のファイルを除き，全てChatGPTファイルで構成されている。また，ChatGPTファイルは本クラスターのみに現れ，他のクラスターには点在していない。ChatGPTファイルに特に顕著に現れている語は，*their*, *students*, *university*, *these*, *and*, *well-being*, *can*, *stress*である。

ChatGPTのクラスターの下のクラスターは，DeepLの5ファイル（DL_46, DL_38, DL_35, DL_51, DL_2）を除き，Studentの作文で構成されている。このクラスターは，図上部の語を分けるクラスターが大きく2分割されている左側のクラスターに属する語が多く現れる傾向にあると言え，このクラスターには先にも取り上げた人称代名詞*I*, *my*, *you*, *your*, *we*, *our*のほか，これらの人称代名詞と共起しやすいであろうbe動詞や，*when*, *if*, *because*等の従属接続詞，形容詞*important*, *good*や，*make*, *get*, *take*, *need*, *eat*といった動詞の原形が含まれる[7]。

Studentクラスターの下は，主にDeepLファイルで構成されるクラスターである。本クラスターに含まれるStudentファイルは以下の5ファイルである：ST_72, ST_80, ST_43, ST_62, ST_28。本クラスターは，図上部の語を分けるデンドログラム右手に位置するクラスターのうち，ChatGPTに顕著な*their*, *students*, *university*等を除く語が顕著に現れる傾向がある。このクラスターには，*they*, *them*といった人称代名詞のほか，助動詞*may*, 等位接続詞*or*, *however*, 冠詞*the*, *a*, *an*, 前置詞*in*, *of*, *on*等が位置している。

更に下のクラスターはStudentとDeepLが混在しているクラスターとなっている。傾向としては，左側の語のクラスターが多く現れると言えるものの，右側のクラスターに位置する語も一定数現れている。最下部に位置しているST_8～ST_71までの15ファイルは，特に一番左のクラスターに含まれる語（e.g., *when*, *am*, *was*, *I*, *my*, *make*, *me*, *not*, *but*, *day*, *every*, *first*, *at*, *very*）が多く使用されているファイルである。

[7] 本稿では動詞等，語のレマ化は行っておらず，作文内で使用されている動詞の形がそのまま現れていると言える。

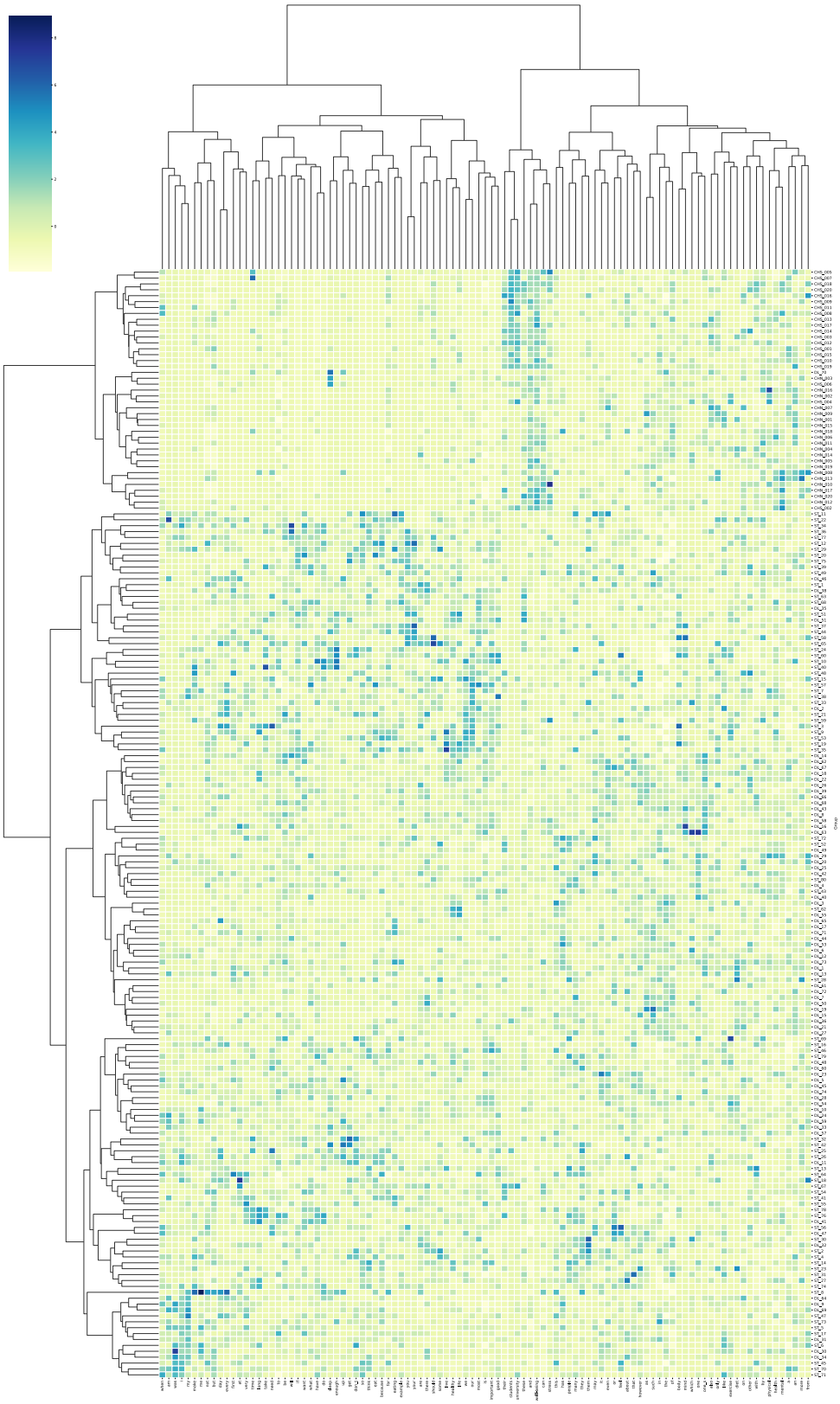

Fig. 4: 3 コーパス上位 100 語の HM 図 (Euclidian 距離, Ward 法)

以下 Fig. 5 および Fig. 6 は，3 コーパス上位 500 語を対象として CA を行った結果であり，順に列座標の第 1, 2 次元，行座標の第 1, 2 次元を描画している。Fig. 5 の色分けは，Fig. 2 に準じる。Fig. 5 では，DeepL のファイルがやや第 3 象限に広がりつつ中心にまとまっている。Student のファイルは第 1, 4 象限に縦に広がっている一方，ChatGPT ファイルは第 2, 3 象限で 0.0 の線上にまとまって現れている。

Fig. 6 で対応する単語を見ると，Fig. 3 と同様に第 1 象限の右上には *your*, *you*，第 4 象限右下には *I*, *my* が位置しているほか，上位 50 語には含まれていなかった *she*, *school*, *was*, *me*, *am* も現れている。Fig. 3 の分析対象には含まれていなかった語を分析の対象にしても，ファイルの分布は大きく変わることはなく，Fig. 5 および Fig. 6 も第 1 次元の値が負の値は（図左側）機械が介入した作文が位置すると言え，第 2 次元の値が正の場合は作文の内容が一般的な内容に言及しているのに対し，負の場合は個人的な内容やエピソードを含んでいると考えられる。

分析対象の語を上位 50 語から 500 語に増やすことにより起こる顕著な違いは，分析対象となる語に内容語を含む頻度の低い語が分析に多く含まれる傾向にある点である。しかし，上位 50 語と 100 語で描画した HM（Fig. 1, 4），上位 50 語と 500 語で描画した CA の結果（Fig. 2, 3, 5, 6）に，それぞれ大きな違いは見られない。むしろ，分析対象語の範囲を拡げた方がより各コーパスの特徴を的確に捉えた結果を得られた。次節では，本節で提示した結果をまとめ，筆者による考察を施しながら，受講生がどのような点に気づき，内省を得たのかについて論及する。

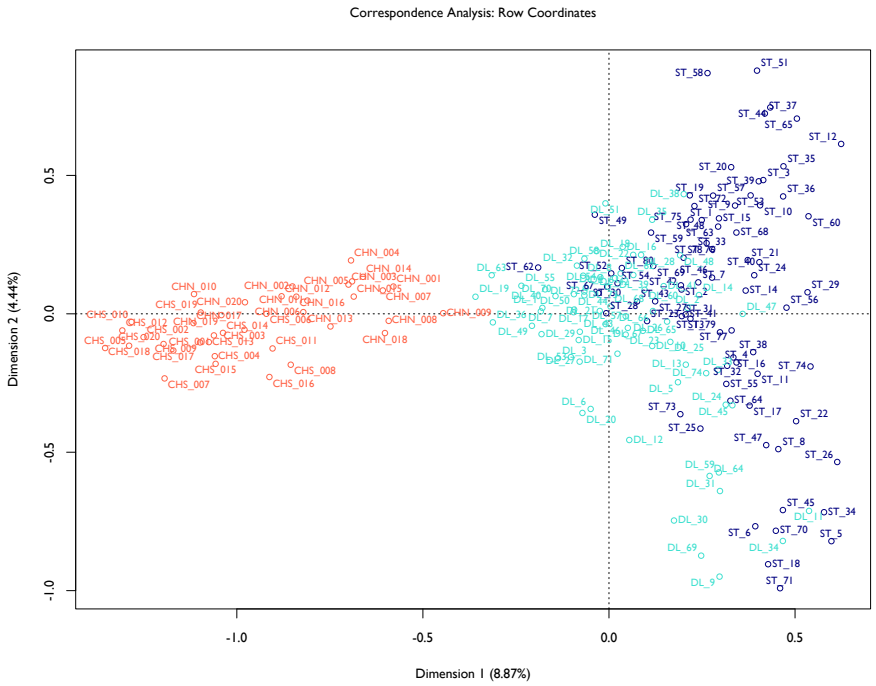

Fig. 5: 3 コーパス上位 500 語の CA (Row)

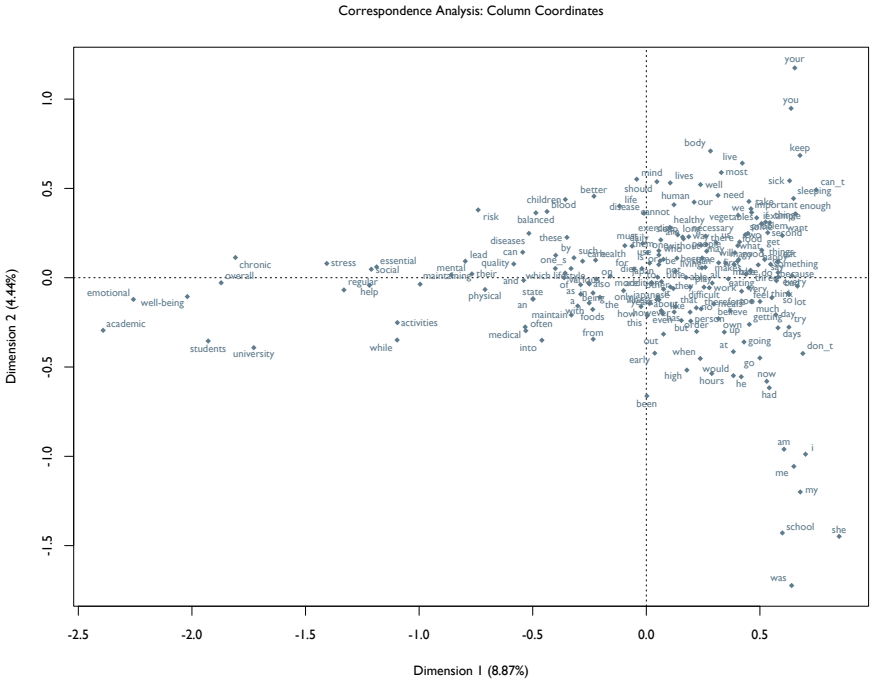

Fig. 6: 3 コーパス上位 500 語の CA (Column)

4. 考察

本節では，前節で提示した階層的クラスター付きヒートマップとコレスポンデンス分析の結果をより詳細に考察していく。

Fig. 1–3 では，ChatGPT の生成した作文には，*the*, *of*, *in*, *as*, *a*, *also*, *more*, *physical*, *health*, *mental*, *with*, *by*, *on*, *these*, *their*, *and*, *can* が，特に Student コーパスと比較して顕著に用いられることがわかった。これらの語には，*the*, *a* といった冠詞や，*of*, *in*, *with*, *by*, *on* といった前置詞がその大半を占めている。

第二言語（英語）習得において，形態素の習得順序は以下 Fig. 7 の順序が普遍的な “Natural Order（「自然な順序」）” (Krashen, 1977) であるとされてきたが，一方で必ずしも Fig. 7 の順序に則らないという主張，報告も近年なされている。また，形態素の習得順序は学習者の母語の影響を受けるという主張もなされている。白井 (2008: 28) は，冠詞の習得について Krashen の 自然な順序に触れながら，日本語を母語とする英語学習者の習得は Krashen の自然な順序よりも遅いことを示唆している。また，Lightbown and Spada (2013: 45, 47) も，特にスラブ諸言語，中国語や日本語を母語とする学習者は，上級者であっても冠詞の（正確な）使用には苦慮する点を指摘している。これら形態素の習得順序は，「実際はかなり母語の転移によって左右される」（白井，2008: 29）とされ，上記形態素の中に前置詞は含まれていないものの，前置詞がない日本語を母語とする学生の作文内での前置詞の使用頻度や種類が，英語学習者以外が作成した作文と比較して異なっていそうであると考えられる。

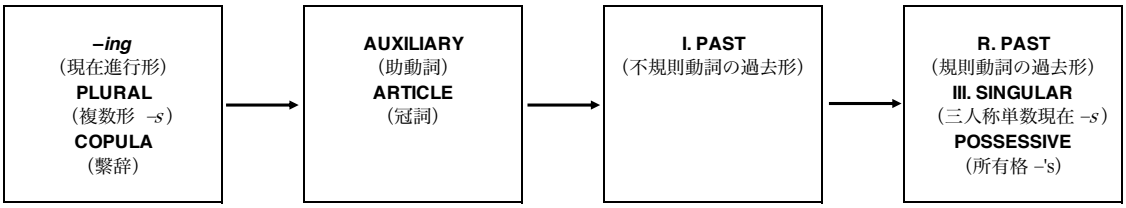

Fig. 7: 形態素の習得順序（Krashen, 1997: 149, Fig. 1 改）

ChatGPT ほど顕著ではないものの，DeepL ファイルにも冠詞や前置詞は使用されており，2 や 3 に見られる変数の位置からも，Student 作文よりは DeepL 作文に冠詞や前置詞の頻度が多い。冠詞や前置詞の頻度の差について複数の受講生が指摘をしている。英作文の際に英語学習者である自分達は「機能語をつけ忘れてしまうことが影響している」のではないか，日本語を母語とする英語学習者の「作文における弱点」あるいは「癖」が冠詞や前置詞にあるのではないか，と意見を述べる学生や，冠詞や前置詞が Student 内に少ないことは，学習者の英文法の「レベルがあまり高くない」ことを表しているのではないか，とする声もあり，学習者の英作文の際の注意点に学生自ら目を向けることができた。

しかし，ChatGPT と比べて DeepL や Student に冠詞が少ない要因は，英文法や冠詞の用法が十分に習得できていないことのみに起因するのではなく，また ChatGPT より冠詞が少ないことがすなわち冠詞の用法を含む英文法のレベルの低さに直結するとは限らない。ChatGPT に現れず，DeepL や Student で用いられている *my*, *your* が，冠詞の頻度を分ける要因の一つである可能性がある。言わずもがな，人称代名詞所有格と冠詞は共起し得ず（e.g., my life/the life/*the my life/*my the life），また人称代名詞所有格と冠詞は文法的には置き換え可能であっても意味や指示対象が変わってくる（e.g., my thought/the thought/a thought）。人称代名詞所有格と冠詞を併用するのは明らかな誤用であるが，それ以外は文法的誤りではなく，語用が自然か否か（ネイティブの語用に近いか否か）や適切であるかが異なる。

Student，DeepL の各コーパスの *my*, *your* が使用されている文脈の例を，以下 Fig. 8–10 にコンコーダンスラインにて示す。 *my* や *your* が使用されている箇所を見ると，*the my life や*my the life のような誤用は見受けられない。くわえて，“my father/friends/brother/cousin” 等，冠詞より *my* を付与した方が文脈上適切であると考えられる用例も多く見受けられる。

your の使用についても，明らかな文法的誤用の例は見当たらず，作文の読み手を想定しつつ，特定の人を指示しない generic *you(r)* を用いて記述したと推定できる。

DeepL の作文における *my* や *your* の使用については，翻訳前の日本語文に *my* や *your* に辞書的意味で相当する「私の」,「あなたの」等が記述されているか否かも着目すべき点である。Student コーパスにおける語の選択，使用には書き手である学習者の意図が介入するが，機械翻訳の場合は翻訳文を出力する際には書き手の意図は介入し得ず，基本的には翻訳元の文にある表現に従って訳出されることが予想される。しかし，DeepL ファイルで *my* や *your* が使用されている箇所と該当する日本語作文を比較すると，「私の」や「あなたの」が使用されていない箇所でも DeepL は訳文生成の際に *my* や *your* を使用していることがわかった。

以下例文は， DeepL における *my* (Fig. 10) のコンコーダンスライン 1，3 行目が現れている文（例文 (1a), (2a)）と，それぞれの元となった日本語文（(1b), (2b)）である（太字は筆者による）。

(1a) I did not have any particular worries about **my health**. However there was one thing that bothered me.

(1b) 大学生活は高校までと比べて睡眠時間も長く確保できるし、自身の体力のなさを痛感する機会もなかったからか、特に**健康**についての悩みもなく過ごしていた。しかし、たった一つだけ気になることがあった。

(2a) **My mother and sister** do strength training almost every night.

(2b) **私の母と姉**が、ほとんど毎日夜に筋トレをしていることだ。

```
ed my decision because they cared about my life Actually the level of medical care
MICHO This is fruit vinegar and affects my health and beauty Next I have protein d
 talk with a friend Not only I but also my father is interested in health He is co
the piano with force on my fingers Also my friend have stressed she often speaks b
days maybe worse My parents told me and my brother and sister to do exercise not t
 it may make our get ill For this I and my friends like eating many sweets for exa
eteria that is run by the university At my university Mie University supplies all
 muscle to keep my health and do better my kendo performance And more protein has
tal Therefore I must tell that of bribe my family in order to be treated in the ca
tudy in Vietnam as exchange student But my parents and my friends opposed my decis
g The dog is took a walk twice a day by my mother SoI will do in place of her Seco
 keep my health better but also to care my mental Cooking our own meals bring our
itude that oppose lookism should change my surrounding and even the world
 my own meals or some sweets to control my feelings when I faced something bad Whi
ition you should have good meals I cook my own meals at least four days a week I d
 has power of healing In my case I cook my own meals or some sweets to control my
tch youtube videos very much These days my favorite youtuber suffer from mental il
ry to exercise these days I can develop my muscles by exercising I can have less t
deadline But if I get up early I can do my assignment in the morning and get time
anese workers are very busy For example my cousin leaves his home at 7 am and retu
en I finish the cooking I feel very few my bad feelings and I feel very happiness
of body health It is very important for my health to exercise Recently I_m into st
nd beauty Next I have protein drink for my muscle I want to build muscle to keep m
not realize it I should be grateful for my parents Currently I do some things to k
angry This thing impact bad effects for my usual communication and I feel down
 scenes and buy some souvenirs and give my close friends to them And I like play v
n other day I think that exercises give my health good effects Second I drink vine
es you should be careful of your habits My eating habits are so bad I don_t eat an
mind Healthy mind makes our lives happy My method to remove stress is getting enou
mmunication with others surprisingly If my eyelids do not open completely I was re
```

Fig. 8: Student における *my* のコンコーダンス 30 項目（L1 のアルファベット順，L1・R1 の重複除く）

(1) では *my* が訳出されているが，(1)' では「私の健康」と直接「健康」を修飾する語は記述されていない。「大学生活」から「過ごしていた。」までが一文であり，途中に「自身の」

が使用されていることから，翻訳文で文を分割する際に必要になった主語を *I* とし，そして *health* の前に *my* が挿入された文を生成した可能性も考えられる。 (2) は，日本語文 (2)' においても *my* に相当する「私の」が記述されている例である。日本語文に「私の」が使用されていようといまいと，上記 2 例においては DeepL の翻訳文における *my* の使用は誤りではなく，文脈や指示対象を考えても適切であると言える。したがって，Student と DeepL において *my* が使用され，ChatGPT には見られない要因の一つは，文法的な誤りというよりは寧ろ作文内において実体験や自分の意見を記述するか否かであることが，以上のことから考えられる。

```
                                      Your health is important thing in your life
 be something bad and you worried about your future In the end everyone wants to be
ons lead to good health Also activating your brain such as reading or drawing can pr
 have many mistakes So breakfast affect your performance Finally I will tell you abo
 for mindhealth Your mindhealth afffect your body Body snd mental influence each oth
 can help not only your mental but also your body One research says hugging can prev
you get up early and you walking around your house You feel good morning window And
ur friend Therefore by looking again at your lifestyle a little you can live healthy
n you cannot work or study well because your brain cannot get oxygen Dietary Fiber i
is important thing in your life because your feelings depend on your health I want t
g your hands is the basic way you boost your immunity to disease Keeping your mental
inding our happiness  You may be broken your heart  However do not mind  it is uncou
ay So you can say your body are made by your eating What do you eat last night If yo
oney For example when you get an ill by your wrong life habit you have to go to the
of your own physical health if you cook your own meals Make your own meals is not on
want to stay healthy you should discard your stereotypes I accept that physical heal
g for a walk Furthermore you can not do your housework by yourself If you have a gre
 is good way to relax I recommend doing your original magic For example making smile
ke medicine You feel free and can enjoy your life In the other hand the people who a
ormation So crimes may involve you Even Your eyes copy scene that you in Explain C:
 go anywhere It is because you may feel your body hard to move They have to choose m
t skip breakfast It is important to fix your life rhythm It should be well-balanced
bad too For instance it is very bad for your health of mind and body not to sleep mu
lcium These nutrients are essential for your healthy life Vitamin is contained in ve
style a little you can live healthy for your lifetime I think that keeping our body
lieve stress For example it is good for your mind health to watch your movie play sp
he most basic of basics Try looking for your styles
you should think about your health from your twenties A course I think health is imp
u can get a salary and more you can get your meals too It is a kindness system for u
 cause heart sick Adult people who have your child should very careful when your chi
```

Fig. 9: Student における *your* のコンコーダンス 30 項目（L1 のアルファベット順，L1・R1 の重複除く）

一部の Student, DeepL 作文に顕著に現れる *you* および *your* が *I*, *my* と異なる点は，対応する「あなたは/が/の/を」が日本語作文での使用が少ないと予想される点である。現に，日本語作文内において，「あなたの」の出現ファイル数と頻度は 2 ファイル中 3 回，「あなたは/が」は 3 ファイルのみに出現し，頻度は合わせて 9 である。しかし，翻訳した DeepL ファイルには少なからず *you* および *your* が使用されており，そのファイル数と頻度は 118 回/41 ファイル（*you*），40 回/17 ファイル（*your*）となっている。*you*（Fig. 11）と *your*（Fig. 12）のコンコーダンスそれぞれ 30 項目を以下に示す。

Fig. 12 のコンコーダンスのうち第 1, 2 行目に挙がっている *your* が使用されている英文と，機械翻訳にかける前の日本語作文から対応する文を以下に引用する。

(3a) **Your health** is the most important thing in **your life**.

(3b) **健康はあなたの人生**の中で最も大切なことである。

(4a) If *we* were healthy *we* could go out whenever *we* felt like it without worrying about aching joints or frequent urination and *we* could indulge in *our* favorite foods without worrying about blood sugar or cholesterol levels. **You** can go out and eat whatever **you** like without worrying about **your blood sugar or cholesterol levels**.

(4b) 健康であったならば、思い立った時に、身体の節々の痛みや頻尿に煩わされずに外出することもできるし、**血糖値やコレステロール値**を気にせずに好きなものを頬張ることもできる。

```
d not have any particular worries about my health However there was one thing that
Among these I would like to write about my own experiences with exercise and diet
g care of my own health mentioned above My mother and sister do strength training
ence of an exercise habit also affected my eating habits As a college student I ha
day Above all what I think has affected my healthy lifestyle is the part-time job
 usually walk about 6000 steps Although my step count may be low compared to the a
o do the things I want to do freely and my activities would be restricted In order
 to choose between my mental health and my physical health I would give priority t
st of my time sitting my back aches and my shoulders are quite stiff Thus I can sa
while I sleep my organs do not rest and my sleep becomes shallow I also try not to
I obviously eat less than I used to and my stomach often gets upset Even when he i
ool and it was exam season and I was at my desk all the time hardly leaving the ho
third moderate exercise Looking back at my recent life I realized that none of the
ng choices even at the time I go to bed My parents are the ones whose work is affe
l exhaustion If I had to choose between my mental health and my physical health I
ul in a bad way During the school break my parents told me to get some exercise to
becomes shallow I also try not to bring my smartphone into the room where I sleep
 a severe form of hay fever and I broke my ribs due to the strong cough that I had
ave a healthy mouth I had been brushing my teeth well and refraining from eating s
hard I am feeling listening to music by my favorite artists cheers me up From the
 unhealthy lifestyle which is caused by my laziness Therefore I must discipline my
g methods In my case this means calling my mother once every few weeks and laughin
fective stress reliever I find it calms my mood and gives me a sense of well-being
use I was able to get more sleep during my college years than I had in high school
I will feel better and be able to enjoy my life more
 I did not have the opportunity to feel my own lack of physical fitness I did not
and sweet foods Furthermore when I felt my vision was deteriorating I went to see
 big fan of sleep First I try to finish my meals at least three hours before sleep
able juice is really doing anything for my health When I buy vegetables at the sup
re a main dish side dishes and soup for my tired body after a hard day_s work and
```

Fig. 10: DeepL における *my* のコンコーダンス 30 項目（L1 のアルファベット順，L1・R1 の重複除く）

```
aim extreme health methods First of all you should carefully examine the informatio
uce the burden on your organs and allow you to maintain a healthier body and lifest
y important to have a hobby that allows you to release your mind This is because it
ur immune system may be compromised and you may become ill Often the way we maintai
omach ache and your legs feel tired And you think to yourself I was leading a norma
p does not mean that you slept well and you will not feel refreshed when you wake u
 coronavirus infection two years ago As you all know we occasionally catch a cold b
ow You are living a normal life just as you always have But suddenly you feel sick
 do for the same reason as the first As you can see health and diet are strongly re
ver had the experience I describe below You are living a normal life just as you al
ling any slight discomfort in your body you should first review these three points
y and lifestyle Even with all this care you may still get sick However regular heal
ant to go through the effort of cooking you have the option of eating out If you ch
fter 8 hours of shallow sleep So how do you get deep sleep I have researched this s
should also eat a well-balanced meal Do you have many likes and dislikes If so it i
Your head feels hot your mouth gets dry you have a stomach ache and your legs feel
I didn_t do anything unusual Eventually you will crawl into bed and rest not knowin
melancholy and depression in everything you do In the worst case scenario it can ev
it a state in which you have everything you need including a perfectly balanced die
d is to exercise If you do not exercise you are likely to become obese The body bec
f time With a little effort to keep fit you will be able to lead a good life
 as your mind sinks further and further you naturally lose your appetite and find i
 is not how long you live but how happy you live And I believe that what is necessa
ve Second health cannot be perfect Have you ever had the experience I describe belo
 that it is actually good to be healthy You don_t realize how much you appreciate g
other words if you want to stay healthy you must continue to engage in a multitude
other words if you want to stay healthy you must continue to engage in a number of
other words if your body is not healthy you will not be able to do the things you w
not many people who will notice or help you so you have no choice but to self-repor
urs of sleep is good Proper sleep helps you think clearly and improves the quality
```

Fig. 11: DeepL における *you* のコンコーダンス 30 項目（L1 のアルファベット順，L1・R1 の重複除く）

(3a) の例文では，元の日本語文 (3b) では「あなたの」が「人生」を修飾している箇所（“your life”）に *your* が使用されているのに加え，原文では「あなたの」が修飾していない，文頭の「健康 (*health*)」にも *your* が付与されている。(4a) の DeepL 翻訳文では，“If we were healthy [. . .]” から始まり，“about blood sugar or cholesterol levels.” までが一文となっており，一文で書かれた (4b) の日本語文に対応している（cf. 例文 (1a), (1b)）。この “If we were healthy [. . .]” の文内では，一貫して人称代名詞は一人称複数の *we* あるいは *our* が使用されている。対して元となっている日本語文 (4b) 内では，対応する日本語の人称代名詞は使われておらず，主語は明示されていない。Fig. 11 の 2 行目に示している *your* は，(4a) の “You can go out [. . .]” から始まる二つ目の文の文末に太字で示している “ [. . .] your blood sugar or cholesterol levels.” の箇所である。(4a) の二文目は，前文の下線部 “we could indulge in our favorite foods [. . .]” と重複している。この重複箇所において，前文では *we* あるいは *our* が使用されている部分に *you*, *your* があてがわれており，それぞれの英文の中での人称代名詞が一致している[8]。(4b) の日本語文内のように，翻訳元の日本語文において *we/our* や *you/your* に対応す

[8] DeepL において訳出文が重複して現れる現象は筆者も何度か目にした経験があり，翻訳元の文の長さや複雑さに要因があるように推測しているが，明確な理由は不明である。

る人称代名詞が使われていなくとも，DeepL は人称代名詞を補って訳出してくれるものの，人称代名詞の選択には揺れがあることが分かった。

```
                                      Your health is the most important thing in y
hatever you like without worrying about your blood sugar or cholesterol levels In ad
         They say you never think about your own health when you are healthy and I t
th gets dry you have a stomach ache and your legs feel tired And you think to yourse
d are feeling depressed At this time as your mind sinks further and further you natu
llow sleep is blue light If you look at your smartphone or computer for a long time
t thing is to be able to choose between your desires It is also important to learn h
s your lifestyle Being healthy enhances your quality of life As mentioned earlier th
d disadvantages Find a method that fits your lifestyle Being healthy enhances your q
kfast Having breakfast is important for your daily rhythm You should also eat a well
hat hungry something that looks bad for your health is super appealing I am sure tha
blems Therefore it is important to have your own stress-relieving methods In my case
l This is because no matter how healthy your body is if you have something on your m
denly you feel sick Your head feels hot your mouth gets dry you have a stomach ache
p going out and so on In other words if your body is not healthy you will not be abl
uch meals No matter how much you eat if your diet is unbalanced you are not maintain
ou are feeling any slight discomfort in your body you should first review these thre
le let_s say you have a very bad day in your daily life and are feeling depressed At
r health is the most important thing in your life If you want to live a healthy life
a regular basis is thought to help keep your mind healthy Based on the above we beli
e about or simply lead an ordinary life your body and mind must be healthy However I
If you have a good basic rhythm of life your risk of getting sick is smaller If you
 further and further you naturally lose your appetite and find it difficult to get t
body and to the brain Along with moving your muscles exercise also protects your men
nd your days without being conscious of your health is not a matter of course but a
ink clearly and improves the quality of your work and study These are just a few of
y your body is if you have something on your mind it cannot be said to be healthy in
est days will help reduce the burden on your organs and allow you to maintain a heal
 mood This is very important to protect your own mental health Going to school and w
ing your muscles exercise also protects your mental health I am currently engaged in
```

Fig. 12: DeepL における *your* のコンコーダンス 30 項目（L1 のアルファベット順，L1・R1 の重複除く）

(4a) の翻訳文は，文法的な誤りではないものの，仮定法過去の使用が不自然である。(4b) の前後の文では，作文を書いた学生自身や誰か特定の人物が健康ではない状況については言及がなく，健康とはどのようなものか，健康な状態でいれば何ができて，健康を損ねるとどのような不都合があるのかについて説明されている。(4b) の文および周辺の文脈を読む限り，書き手の学生は健康でいることがあり得ないことと捉えている[9]と読める記述はない。したがって，書き手の意図を汲み，文脈を考慮するのであれば，仮定法過去の使用は不適切であると考えられ，読み手が違和感をもつ要因となり得ると考えられる。

しかし，当該の仮定法過去の不自然な使用は，本分析の結果に明示的に現れていない。本稿の分析は HM と CA のいずれも語単位での分析であるため，仮定法過去やその他の文法事項が適切に作文内で使用されているかや，書き手の意図を適切に反映しているかまでは見

[9] 仮定法過去は，「現在の事実に反対の仮定」や「現在または未来について [. . .] 仮定の条件を表す」ものであり，「内容は客観的にありうることでも，あり得ないことでも良いが，「そういうことはまずありえないが」という話者の気持ちを含むことが多い」（江川，1991: 254）。

出すことができない。したがって，受講生の考察結果にはこの仮定法過去の不自然な使用等についての指摘は見られなかった。このような語用や表現の詳細に目を向けるためには，コンコーダンスラインを入念に読み，翻訳文であれば翻訳元の文との比較考察も必要である。

作文内での人称代名詞の使用は，ChatGPT と他 2 つのコーパスではその特徴が異なる。先にも指摘したように，ChatGPT では *my* や *you* は使用されていない。一人称代名詞複数の *we* は 14 例，*our* は 31 例，コーパス内に使用が認められるが，特定の人々を指示するのではなく，*you(r)* 同様に一般的人々について言及する，いわば *people* に近い用法である。

コーパス間での人称代名詞使用の違いについては，複数の受講生が指摘しており，中には，Student においての人称代名詞の特徴に加え，*people* の頻度が他のコーパスよりも多く，Student の特徴であると考察している学生もいた。書き手個人のエピソードが作文内に含まれるか否かは，作文が人間によって書かれたものなのか，AI が生成したものなのかを分ける一つの判断要素と言え，生成 AI がどんなに自然（に見える）文章を生成できても，「人間」とは異なることを象徴していると言える。この点について，ある受講生は以下のように指摘した。

> ChatGPT には「私（の）」「あなた（の）」といった人称が使われづらいことから、学生の作文両方[10] は書き手・読み手を意識したものが多いのに対し、ChatGPT は特定の個人を想定しない普遍的な文章を書くのだろうと考えられる。

上記のほか，Student や DeepL は主観的な立場で書かれているのに対し，ChatGPT は客観的な文章であるといった指摘も複数ある。論文やレポートのような，私見や個人の経験の記述が一般的ではない文章を対象とした場合には，必ずしも本稿の分析の結果と一致しないことが予想されるが，書き手の意見を述べる等比較的自由度の高い作文では，人称代名詞の使用差は，人間と 生成 AI の記述した文章差を見出す際の手がかりになると考えられる。

人称代名詞に限らず，ChatGPT では使用される語が作文に跨って似通っているのに対し，学生が書いた Student, DeepL で使用される語は作文によって異なる点を CA から見出し，Student, DeepL の作文には書き手の癖や個性が現れると指摘する受講生もいた。

また，作文内で使用される語の種類と難易度についても，複数の指摘があり，それぞれの指摘ではいずれも使用されている語の難易度が Student < DeepL < ChatGPT の順で高い傾向にある点が言及されている。特に，英語学習の早い段階で習うであろう *good* や *important*, *body* は Student コーパスに顕著に現れている。難易度の指標ではないが，*Longman Dictionary of Contemporary English* では，*good*, *important*, *body* のいずれも話し言葉・書き言葉として使用頻度の高い 1,000 語（S1, W1）[11] と示されている。対して，ChatGPT に顕著に現れる傾向にある（Fig. 2, 3, 5, 6） *physical* は S2，W1 ，*mental* は S2，W2 となっている。*mental* や *physical* が ChatGPT に特徴的な語である点において，語の難易度の他，ChatGPT は健康について書くよう指示されると，精神の健康と身体の健康を分けて詳細を論じる文章を生成す

[10] 筆者注：学生の作文両方とは，Student と DeepL を指す。

[11] 話し言葉は S, 書き言葉は W で示され，使用頻度の高い 1,000 語は 1，2,000 語は 2，3,000 語は 3 で表示される。

るのに対し，学生の書いた作文では，精神と身体を分けずに健康に過ごすためには何が大切か，何が体に良いとされるのかが，*good*, *important*, *body*, *sleep* の使用によって記述されている傾向にあることを HM や CA の結果から指摘する学生も複数見受けられた。

本考察の結果，同じテーマについて学生に書いてもらった英作文と機械翻訳で日本語文を英訳した作文，ChatGPT によって生成された作文を計量文体論手法を用いて分析した結果は，各コーパスの特徴を捉えることができたと言える。また，分析の結果を受講生に考察してもらうことにより，テクストマイニング初学者の受講生であっても，各コーパスの特徴を見つけ，説明できることが示唆された。

5. おわりに

本稿は，同一のテーマのもとで英語学習者が書いた英作文と，日本語作文を機械翻訳を通して英語に翻訳した作文，そして ChatGPT に生成させた英作文に計量文体論手法のうち階層的クラスタリングとコレスポンデンス分析を用いて分析を行った。分析の結果を視覚化して提示するだけでなく，分析および結果の一部を，作文を執筆した学生らに自ら考察してもらうことにより，各作文の特徴や相違点に学生が気づき，機械翻訳や生成 AI の利用について内省を得られたかについて概観した。

分析の結果，学生の英作文，DeepL による翻訳文，ChatGPT による英作文それぞれに特徴があり，視覚化した図を用いることで，学生らもその特徴に気づくことができた。中には，ChatGPT の文を引用し，使用されている語の難易度が高いことを指摘し，語の意味を辞書で調べた上で，自分が同じ内容を述べるならばどのような語を使用し，どのような文を記述するかを例示した受講生もいた。また，DeepL の翻訳文を詳細に観察し，誤りを指摘したり，自分が書く英文との違いを指摘した学生もいたり，各作文の違いを概観したことで，生成 AI や機械翻訳の出力結果を単体で見ていた時には気づけなかった “機械らしさ” や “人間らしさ” に気づき，これらツールの利用の仕方を見直したいという感想を述べる学生もいた。分析を通して得られた学生の考察や内省は，学生が自ら考えて動く自律学習を促進することにもつながるのではないかと期待する。

本研究の結果，階層的クラスタリングやコレスポンデンス分析を用い，コーパス内頻度上位 50，100，500 語を分析することで，各コーパスの特徴を捉えることができることを明らかにした。本分析対象の中に，母語話者による作文は含まれていないため，本稿の結果が人間と機械/AI を分ける要素を説明できているとは言えない。また，本稿で扱ったのは同一テーマの作文であったが，分析する対象のテクストデータのジャンル（e.g., アカデミック/小説/ニュース等）によっても結果は大きく異なることが予想されるため，本稿の結果が他のジャンルにも適用できる可能性は低い。しかし，母語話者の作文データを追加したり，学習者の作文もレベル別に分類しながら数を増やしたり，ChatGPT への入力プロンプトを変えてみたりし，コーパスサイズを大きくしながら同様の分析を行い、各コーパスの特徴を把握することで，ある作文が人間によって書かれたのか，ChatGPT 等生成 AI の可能性が高いのか，機械翻訳を使用した可能性があるのか，人間によって書かれたのであれば，書き手は

学習者なのか母語話者なのか等を識別したり，確率を導き出したりする指標を提示できるのではないかと考える。

日々進化を続ける生成 AI をはじめとする科学技術を利活用しつつも，翻弄されることなく，語学の学習を促していけるような研究結果の提示や使用方法などの検討を今後も継続して行なっていく。

Bibliography

[1] DeepL. DeepL (翻訳ツール）. Available online at https://www.deepl.com/translator (Last accessed 29 March 2024.)

[2] 江川泰一郎. (1991) 『英文法解説』改訂三版. 東京：金子書房.

[3] Google. Google Translate. Available online at https://translate.google.com/ (Last accessed 29 March 2024.)

[4] Imao, Y. (2024) CasualConc (Version 3.0.8) [Computer software]. Retrieved from https://sites.google.com/site/casualconc/.

[5] Krashen, S. (1977) Some issues relating to the monitor model. *On Tesol, 77*(144-158).

[6] Lightbown, M. P. and N. Spada. (2013) *How Languages are Learned*. 6th ed. Oxford: Oxford University Press.

[7] *Longman Dictionary of Contemporary English*. (2014) 6th ed. London: Pearson.

[8] 西尾章治郎. (2023)「生成 AI (Generative AI) の利用について」Available online at https://www.osaka-u.ac.jp/ja/news/topics/2023/04/17001 (Last accessed 29 March 2024.)

[9] OpenAI. (2023/2024) Personal communications with ChatGPT-4. Available online at https://openai.com/ (Last accessed 29 March 2024.)

[10] Rahman, M.M. and Y. Watanobe. (2023) “ChatGPT for Education and Research: Opportunities, Threats, and Strategies.” *Applied Sciences 13*, 5783.

[11] 白井恭弘. (2008) 『外国語学習の科学：第二言語習得論とは何か』東京：岩波書店.