



Title	テキストマイニングとデジタルヒューマニティーズ 2023（冊子）
Author(s)	
Citation	言語文化共同研究プロジェクト． 2024, 2023
Version Type	VoR
URL	https://hdl.handle.net/11094/97312
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

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

The University of Osaka

言語文化共同研究プロジェクト

テキストマイニングと
デジタルヒューマニティーズ
2023

浅野 元子

藤田 郁

田畑 智司

大阪大学大学院言語文化研究科

2024

言語文化共同研究プロジェクト

テキストマイニングと
デジタルヒューマニティーズ
2023

目 次

田畑 智司	プロジェクトの目的と活動	1
浅野 元子	医療系日本語論文の英語抄録の言語特徴に関する探索的研究 —ジャンルに基づくコーパス研究の一環として—	5
藤田 郁	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	25

「テキストマイニングとデジタルヒューマニティーズ」 プロジェクトの目的と活動

本共同研究は、自然言語処理、コーパス言語学・計量言語学、数理統計学、データマイニング、機械学習など、諸分野の知見を有機的に統合した方法論を開発し、テキストマイニングを応用して人文学、言語文化学の諸問題にアプローチする、すなわち「デジタルヒューマニティーズ (Digital Humanities)」の実践と理論的精緻化の可能性を探る営みである。このプロジェクトは、2001 年度に岩根 久教授、緒方 典裕助教授、および筆者の 3 名でスタートした「電子化言語資料分析の方法論」を基礎とするが、2003 年度から名称を一部改め、言語文化研究科の大学院生もメンバーに加わった。2006 年度には三宅 真紀助教の加入を得て、対象言語も英・仏・ギリシャ語に広がった。2011 年には言語文化教育論講座に着任した今尾 康裕講師が加入した。2014 年度後期から、さらに Hodošček Bor 講師が加わった。そして、2019 年度をもって退職された岩根 久教授の後任として、2020 年度に山田 彬堯講師が着任・加入した。言語文化研究科と文学研究科の統合により設立された人文学研究科には「人文学林」という分野横断組織が置かれ、デジタルヒューマニティーズ振興の役割を担っている。その人文学林から、2022 年に菅原 裕輝特任助教が、そして 2023 年に吉賀 夏子准教授が加入し、現在の陣容となっている。(職位はいずれも当時)。2016 年度から、プロジェクトの名称を、当該リサーチコミュニティの名称としてより相応しい「テキストマイニングとデジタルヒューマニティーズ」にアップデートしたが、研究の系統は創始時より常に一貫している。

「テキストマイニングとデジタルヒューマニティーズ」プロジェクトは大きく分けて二つの層で構成されている。一つは研究基盤となるコーパス、テキストアーカイヴの開発・構築、もう一つは構築したコーパス、テキストアーカイヴからのデータ抽出法研究、並びに得られた高次元の言語データの計量分析である。前者には英・仏語の文学作品や、聖書（共観福音書）などの電子テキスト化、ロシア語政治演説コーパス、近代日本文学コーパスの編纂、マークアップ言語 XML による TEI (Text Encoding Initiative: デジタル化したテキストの国際互換規格の枠組) に準拠したタグ付けなど、人文学資料のデジタル化やマークアップ法、データ符号化方法論の開発などが含まれる。一方、高次元人文学データ分析の事例として、語彙・語法、コロケーション、意味構造、語用論などのレベルにおける言語使用の実態研究、高度な数理モデルや機械学習を応用した言語分析やテキストマイニング、文学作品の言語特徴の特定や、使用域間の言語変異や文体識別問題の考察、著者推定法の精密化研究を挙げることができる。

本プロジェクト班は人文学研究科の専任教員 7 名と名誉教授 1 名（今尾 康裕、菅原 裕輝、田畑 智司、Hodošček Bor、三宅 真紀、山田 彬堯、吉賀 夏子、岩根 久名誉教授）、当研究科博士後期課程在学生 6 名（福本 広光、藤田 郁、竹森 ありさ、涌井 萌子、曹 芳慧、Camilleri Gabriele）、博士前期課程在学生 5 名（小堀 彩夏、Vogatza Dimitra、李 晨婕、肖 媛媛、于 拙）に加え、OG の黄 晨雯氏（2024 年 5 月に本研究科言語文化学専攻助教着任）、京都大学 徐勤氏（2023 年 3 月博士学位取得）、大阪医科大学 浅野 元子氏（2020 年 3 月博士学位取得）・名古屋外国語大学 杉山 真央氏（2019 年 3 月博士学位取得）、本学非常勤講師の高橋 新氏、南澤 佑樹氏（2024 年 4 月に本研究科外国学専攻助教着任）、帝塚山学院大学 八野 幸子氏（本研究科博士課程修了）、国立国語研究所の竹内 綾乃氏を主たる参加メンバーとしている。研究を遂行するために、コアメンバー以外も自由に参加できる月例の研究会・討論会などを通して、研究情報の交換、論文や開発ツール、構想段階のプロジェクトや進行中のパイロットスタディのプレビューなどを行っている。

2023 年度は、対面開催とオンライン併用のハイブリッド方式で開催した。オンラインでの研究会の開催を続けているうちに、学外からの研究会参加者が増加したこともあり、今後もハイブリッドでの開催を続ける予定である。

2023 年度「テキストマイニングとデジタルヒューマニティーズ」研究会開催記録 およびメンバーによる DH 関連学会での発表記録

第 1 回 2023 年 4 月 14 日開催

発表者・発表題目

全メンバー 2023 年度の活動計画打合せ

第 2 回 2023 年 5 月 12 日開催

発表者・発表題目

黄 晨雯 「ChatGPT とテキストマイニング」

第 3 回 2023 年 6 月 9 日開催

発表者・発表題目

藤田 郁 “The Stylistics of Alfred Tennyson’s Poetry: A Stylometric Approach”
Hodošček, Bor “YoutuberLinguisticProfile: A toolbox for linguistic analysis of Youtube videos
Work-in-progress report on Tom Scott videos”

第 4 回 2023 年 7 月 7 日開催

発表者・発表題目

杉山 真央 「戦勝記念日演説から見るプーチン大統領の人称代名詞使用」

第 5 回 2023 年 7 月 12–15 日開催 Poetics and Linguistics Association International Conference (PALA)
2023, University of Bologna, Italy

発表者・発表題目

Tabata, Tomoji “Using topic models to explore body language in Dickens’s literature and journalism”
Fujita, Iku “‘How much I love this writer’s manly style!’:
Similarities and differences between Tennyson and other poets”

第 6 回 2023 年 8 月 4 日開催

発表者・発表題目

曹 芳慧 “Characterization by dialogues in Hardy’s novels:
a preliminary quantitative study of three works”
塚越 柚季 「『リグ・ヴェーダ』の韻律情報を利用した文体比較」

第 7 回 2023 年 9 月 1 日開催

発表者・発表題目

Camilleri, Gabriele 「因子分析を用いた文学作品翻訳コーパスにおける発話キャラクターの抽出」

第 8 回 2023 年 9 月 9-10 日開催 英語コーパス学会第 48 回大会（学会創設 30 周年記念大会）
発表者・発表題目

後藤 一章	「音声認識 AI を活用した音声・映像コーパス構築ツールの開発」
菅原 裕輝	「コーパス研究は仮説検証型の科学か？：形式概念分析を用いたメタ分析」
Fujita, Iku	“Extraneous to Allusions: Stylistic differences between Tennyson and 18th-century poets”
曹 芳慧	「Hardy 作品における会話部の感情分析」
杉山 真央	「戦勝記念日演説から見るプーチン大統領の “МЫ (we)” と “ВЫ (you)”」
Tabata, Tomoji	“Exploring body language in Dickens’s fiction through topic modelling”
鈴木 大介	「現代英語における worse の文副詞用法をめぐって」

第 9 回 2023 年 10 月 6 日開催
発表者・発表題目

涌井 萌子	「『レ枢機卿のマザリナード』の帰属検証—分析方法と基準の検討—」
岩根 久	「Pierre de Ronsard (1524–1585) 作品集としての <i>Les Amours</i> の変容 (1552–1560)」

第 10 回 2023 年 11 月 3 日開催
発表者・発表題目

福本 広光	「Clinton, G. W. Bush, Obama による分離不定詞再考： コンコーダンスラインに見える特徴的用法に関する試論」
Vogatza, Dimitra	“Investigating the translational equivalency of pity in EN & JP”
王 簫影	「漢字圏と非漢字圏日本語教科書リーダビリティ研究—長文テキストを対象に—」

第 11 回 2023 年 11 月 18 日開催 国際シンポジウム デジタルヒューマニティーズと研究基盤：
欧州と日本の最新トレンド
発表者・発表題目

吉賀 夏子	パネルセッション 「人文学のためのデジタル研究基盤及び DH の人材育成に関する事例報告」
田畑 智司・	ポスター・デモンストレーションセッション
吉賀 夏子・	「大阪大学における共同研究プロジェクト
菅原 裕輝・	『テキストマイニングとデジタルヒューマニティーズ』の歩み」
ホドシチェク ボル	

第 12 回 2023 年 12 月 1 日開催
発表者・発表題目

徐 勤	
小堀 彩夏	「英訳版村上春樹作品における統計的手法を用いた文体分析」

第 13 回 2024 年 1 月 5 日開催

発表者・発表題目

吉賀 夏子 「江戸時代の佐賀地域における情報基盤の構築とその可能性」
南澤 佑樹 「スウェーデン語の背中に関連する表現について」

第 14 回 2024 年 2 月 16 日開催

発表者・発表題目

肖 媛媛 「コーパスに基づく英語政治ニュース研究 -英語母語圏と非母語圏の比較研究」
立野 寛太 「スポーツの社会言語学—結果の予測不可能性と実況の関わりについて—」

第 15 回 2024 年 3 月 6 日開催 人文学林シンポジウム

発表者・発表題目

田畑 智司・ 「デジタルヒューマニティーズの現在（いま）」
吉賀 夏子・ 一方法論的学際性にもとづくデータ駆動型の人文知の追求へ向けて—
石田 友梨・
ホドシチェク
ボル

第 16 回 2024 年 3 月 8 日開催

発表者・発表題目

菅原 裕輝 “Digital Sociology of Expectation”
今尾 康裕 「Mallet を統合した CasualConc の新機能」

2024 年 5 月
研究代表者 田畑 智司

医療系日本語論文の英語抄録の言語特徴に関する探索的研究 —ジャンルに基づくコーパス研究の一環として—

浅野元子

大阪医科薬科大学

〒569-8686 高槻市大学町 2-7

Email: motoko.asano@ompu.ac.jp

概要 本研究は、日本の学術誌に掲載された医療系研究論文の英語抄録に見られる言語特徴を調査することを目的とする。英語論文の執筆読解を支援する方略を探るための「ジャンルに基づく」(東條, 2015, p. 19) コーパス研究の一環である。国際誌掲載論文の抄録は広く研究されているが、英語学習において日本語による支援が提案されることを踏まえると、日本語の文脈で書かれた英語抄録の調査が必要だと考えられる。英語と日本語の抄録が付与された論文の抄録を対象に「修辭的機能」(Swales, 1990, p. 138) に注目して「フレーズの構成素」(Quirk, Greenbaum, Leech, & Svartvik, 1985, p. 68)を調査した。英語抄録では、“This study aimed to”などの「意図的なメタ談話記述」(Van Bonn & Swales, 2007, p. 98) が多く観察され、日本語抄録の「研究」などと同じ役割を果たす部分は主に名詞の“study”などが使用されており、日本語テキストの影響が見受けられた。英語抄録テキストはPubMed (National Library of Medicine, 2023)に収載されており、日本の「学術コミュニティ」(Van Bonn & Swales, 2007, p. 106)における「英語学習者のための参考例」(Hino, 2018, p. 61) となる可能性がある。また、健康リスクが予測される場合、英語では“affecting the quality of life”などと明示的に記されているのに対し、日本語抄録では「生活に幅広く関係する」など和らげた叙述がみられ、「抄録の地域性」(Van Bonn & Swales, 2007, p. 106) が示唆された。

キーワード 英語抄録, 日本語抄録, コーパス, ジャンル, ESP, EAP

Linguistic features of English abstracts for Japanese medical science research papers:

A part of a genre-based corpus study project

Motoko Asano

Osaka Medical and Pharmaceutical University

Abstract This study aims to examine the linguistic features of English abstracts in medical research papers published in a Japanese academic journal, as part of a “genre-based” (Tojo, 2015, p. 19) corpus study project for exploring support strategies for research communication. While abstracts in international journals have been studied extensively, the need for investigating English-language abstract texts prepared within the Japanese context has arisen from the benefit of using learners’ first language as a scaffolding tool in learning environment. This study focused on the “rhetorical functions” (Swales, 1990, p. 138) and examined the “constituents of phrases” (Quirk, Greenbaum, Leech, & Svartvik, 1985, p. 68). The analysis revealed frequent use of “[p]urposive metadiscoursal statements” (Van Bonn & Swales, 2007, p. 98) such as “This study aimed to” and a preference for nouns such as “study” that fulfilled roles similar to those observed in the Japanese texts, suggesting an influence of Japanese text structure. Considering that these English abstracts are accessible via PubMed (National Library of Medicine, 2023), the texts could serve as “reference sample[s] for Japanese learners of English” (Hino, 2018). English texts explicitly describe impacts like “affecting the quality of life,” in contrast to Japanese abstracts, which tended to use softened expressions when communicating potential health risks. The findings may indicate the “locality of abstract texts” (Van Bonn & Swales, 2007), potentially offering insights into the teaching and learning of disciplinary preferences of highly conventional texts in the target region.

Keywords research article abstracts, corpus, genre, ESP, EAP,

1. はじめに

本研究は、「ジャンル」(Swales, 1990)ⁱの考え方に基づいて、論文の読解と執筆を支援する方略を探るコーパス研究の一環である。「パラレルコーパスの使用」(Boulton & Cobb, 2017, p. 378, 以下、特に断りがなければ日本語訳は筆者による)などによる母語を利用した英語学習が効果的であると説かれる(服部, 2011; 今井・岡田・野島, 2012; 大津・吉田, 2023)。医療系大学における英語抄録の読解授業で公式日本語訳を用いた支援が有効であることが指摘されている(中野他, 2022)。「専門分野の専門家による言語使用」を学習することは「有益な基礎」をもたらす(Hyland & Tse, 2005, p. 138)。日本の医療系研究論文に付された英語と日本語の抄録テキストを調査した研究(浅野・松田, 2022)では、両者の間における「ステップ “Move-step”」(p. 39)レベルでの差異が明らかにされ、それぞれが異なる「学術コミュニティ」(Van Bonn & Swales, 2007, p. 106)に向けて叙述されたことが確認されている。本研究では、英語抄録テキストの特徴を、日本語抄録テキストを手掛かりに調査した。特に「特定のコミュニケーション上の機能を具現化するテキストの断片」(Moreno & Swales, 2018, p. 41)に焦点を当て、英語抄録で日本語抄録と同様の「修辭的機能」(Swales, 1990, p. 138)を果たす「テキストの断片」(Moreno & Swales, 2018, p. 40)の「フレーズの構成素」(Quirk, Greenbaum, Leech, & Svartvik, 1985, p. 68)を精査した(以下、「テキストの断片」(Moreno & Swales, 2018, p. 40)を単にテキストまたは断片, 「フレーズの構成素」(Quirk, Greenbaum, Leech, & Svartvik, 1985, p. 68)を構成素という。)

ⁱ ESP 研究と学術英語 (EAP) 執筆支援, 「ディスコース・コミュニティ」, 「ムーブ」など研究の柱となる概念については補遺に記した。

1. 1. “part-genre”としての研究論文抄録

研究論文の抄録は「情報管理のためのツール」(Lores, 2004, p. 281)としての役割を果たす“part-genres” (Swales, 2004, p. 210) の一つである。英語抄録の付加は研究を広範な読者に届ける手段とされる (Lores, 2004; Pho, 2008)。また、論文のテキストにおける「国, 社会, 文化」などの「地域的影響」(Swales, 1990, p. 66)が明らかにされて久しいが、抄録テキストは単なる「中立的な要約」ではなく「修辭的要素が多い」(Hyland & Tse, 2005, p. 126) ことも指摘される。フランスの学術誌に掲載された英語とフランス語の抄録などを分析した研究 (Van Bonn & Swales, 2007) では、英語テキストの一人称代名詞の使用などの特徴が「教育的理解」(p. 106)に寄与するとされた。2010 年以降、「英語を母語としない指導的立場の教師や研究者」(Hyland & Shaw, 2016, p. 7)が増加し、学術のための言語は「母語話者不在の特殊な談話の型」(Mauranen, Hynninen, & Ranta, 2016, p. 45)とみなされるようになってきた。Lores-Sanz (2016)は、英語以外を母語とする書き手の抄録が少ない「ムーブ」で複雑に構成されることを報告し、「規範的な IMRD 構造」(p. 77) の再考を提案している。この提案は「慣例とは異なる構造」が「ゲートキーパー」に「承認」(Flowerdew & Habibie, 2022, p. 45) される可能性を示している。

1. 2. ジャンルに基づく論文執筆支援と日本語の使用

英語論文の緒言部における「Move-step 分析」(Swales, 1990, p. 143) を行った研究で「先行研究を紹介する」際の書き方が紹介されている。

例1: The effect of . . . has been studied extensively in recent years

(Swales, 1990, p. 144, 以下, 特に断りがなければ下線による強調は筆者による)

Swales (1990)に基づいて Academic Phrasebank (Morley, 2023)が編纂されている。これは、学位論文や研究論文などを用いたコーパスから特定分野に依存しない共起語を抽出したもので、研究者に広く使用されている (Davis & Morley, 2018)。日本では、中谷 (2016)、保田 (2021a)、河本と石井 (2022)などによる論文執筆のための指南書が数多く出版されている。中谷 (2016, p. 142)は論文コーパスに依拠し「よく使われ」る動詞を挙げて日本語を付与している(例 3)。保田 (2021a, p. 190)は「英語論文執筆支援ツール AWSuM」(水本, 2017) に収められた論文からフレーズの例を示し日本語を付けている (例 4)。河本と石井 (2022, p. 165) は、医療系論文のムーブ分析に基づいて連語を示し日本語を付加している (例 5)。「Scaffolding (支援) の中に日本語」(大津・吉田, 2023, p. 40) を加えた提示は、より深い理解を導くと考えられる。

例3: . . . report 「報告する」 (中谷, 2016, p. 142)

例4: . . . has not been well studied 「十分には調査されてこなかった」 (保田, 2021a, p. 190)

例5: . . . was associated with 「～と関連していた」 (河本・石井, 2022, p. 165)

先行研究 (Van Bonn & Swales, 2007) では、英語とフランス語で書かれた抄録が調査されている。科学のためのフランス語については「特別な言語保存の取り組みがあるかもしれない」(Swales, 1990, p. 99) とされる。一方、日本では「学問・思想の基本用語」は「近代以降、西欧文明の学問・思想など」を「急速に受け入れる」過程で「翻訳のために造られた」とされる (柳父, 1982, p. ii)。柳父(1982) は日本と諸外国の環境の変遷に言及して「社会」「個人」などの「ことばを機能として使いこなすべく」(p. 37)「漢字二字の表現」によって「学問を組み立てる」趨勢を描写している (p. 124)。今村(2004) の研究では、「動詞として使われる割合が高いサ変名詞」(p. 10) が論文執筆に重要であることが明らかにされている。本研究では、英語抄録テキストの言語使用の特徴を、日本語抄録の「サ変名詞」(今村, 2004, p. 10)を手掛かりに調査した。

1.3. 多言語コーパスによる対照ジャンル研究

対照研究は、外国語学習の困難度 (Lado, 1957)や修辞の文化的差異 (Kaplan, 1966) などに着目して数多く行われてきた。なかでも「対照修辞学 “*Contrastive Rhetoric*”」は「学術英語教育」に「直接関連する分野」(Swales, 1990, p. 64, 大文字とイタリック体の使用は原著による)と位置付けられる。それに関連した「対照ジャンル研究 “*contrastive genre analysis*”」(Aijmer & Lewis, 2019, p. 2) では「ジャンルに繰り返し出現する特徴」(p. 4) などが比較検討されている。初期の研究には、論文における、英語母語話者と非母語話者の動詞使用の研究 (Wingard, 1981),「ヘルシンキ大学における 1988 年からの研究」(Mauranen, 1993, p. 6) などがある。このアプローチは、“*Intercultural rhetoric*”(Connor, 2004, p. 291) として発展し、「学術研究のための英語」の「スタイル」の「多様性」(Swales, 2017, p. 251),「ディスコース・コミュニティ」の「社会的慣例」(Connor & Ene, 2019, p. 53)に注目が集まっている。

さらに、多言語コーパスを用いて「言語間の違い」を調査する研究が増えている (Culo, Hansen-Schirra, Maksymski & Neumann, 2017, p. 53; Lavid Lopez, 2019, p. 159)。英語とスペイン語の小説、新聞記事などの二言語コーパスを試作して「アノテーションを施すプロジェクト」(Lavid & Moraton, 2016, p. 76)より「新聞記事、論説、編集者への書簡」各 6 編を抽出して調査した研究では「言語特有の変異」(p. 179) を明らかにしている。多言語コーパス研究では「ある言語または言語変種のテキストと対応する別の言語または言語変種のテキストによるコーパス」を「パラレルコーパス」(Borin, 2002, p. 4),「多言語テキスト間」の「主題やスタイル」,「ジャンルや機能」が同等あるいは比較可能なコーパスを「コンパラブルコーパス」(Borin, 2002, p. 5; Connor & Moreno, 2005, p. 155; 仁科, 2023, p. 60; Rees, 2022, p. 395) といふが「用語にはばらつきがある」(Kenning, 2010, p. 487) とされる。先行研究では、英語とスペイン語による論文の「考察部」の「コンパラブルコーパス」(Moreno & Swales, 2018, p. 48), 英語とフランス語の論文抄録「コーパス」(Van Bonn & Swales, 2007, p. 98) を用いてテキストをステップレベルで分析している。Van Bonn と Swales の研究では“*It is now widely acknowledged that,*”などが現地の「ディスコース・コミュニティ」(2007, p. 101) に慣例的に使用されると報告しているが、日本の医療系のディスコース・コミ

コミュニティに受容される英語の特徴を調査した研究はまだ少ない。

2. 方法

2.1. 目的

本研究の目的は、Shaw (2016, p. 252) による「コーパスを用い」た「テキスト分析」の提案に従い、先行研究 (Moreno & Swales, 2018; Tanko, 2017) を参考にして、日本の医療系研究論文の英語抄録の特徴を同じ論文の日本語抄録を手掛かりに調査することである。

2.2. 使用したコーパス

英語以外の言語による学術誌は「研究を正当化」するなどを目的として「負担が少ない小規模な学術コミュニティ」(Van Bonn & Swales, 2007, p. 97) を形成するとされる。それらの学術誌の中で「代表性、評価、アクセスしやすさ」(Nwogu, 1997, p. 121) を備える『日本公衆衛生雑誌』に掲載された 2020 年 1 年分の日本語論文 (45 編) の英日抄録(計 90 編)を対象とした。同誌は日本公衆衛生学会によって 1954 年に創刊された定期刊行物であり、h-index (Hirsch, 2012)によるジャーナル評価の日本語部門において 9 位に位置する (Google Scholar, 2023)。英語抄録は専門分野のデータベース PubMed (National Library of Medicine, 2023)で検索可能であり「広範な学術的文脈の中に位置付け」られ得る (Van Bonn & Swales, 2007, p. 105)。投稿規定に「原著」は「公衆衛生上重要で科学的な研究・調査に関する論文」(日本公衆衛生雑誌編集委員会, 2022, pp. 841–842)で「日本語の原稿」には「目的」「方法」「結果」「結論」に構造化した「400 語以内の英文抄録」を「英文」の「専門家によるチェックを受け」て付することが推奨される。学術誌のサイトからテキストを入手し、日本語抄録は WinCha (2000 R2, 奈良先端科学技術大学院) を用いて形態素に分けた (表 1)。

表 1 使用したコーパス

	総語数	異なり語数
英語抄録テキスト	15,531	2,301
日本語抄録テキスト	19,048	2,128

CasualConc (3.0.4, Imao, 2023)

2.3. 分析手順

英語抄録テキストは、品詞タグ付けを行い (TagAnt 2.0.5, Anthony, 2022), 日本語抄録の「サ変名詞」(今村, 2004, p. 10) を頻度上位 30 語まで抽出し、粗頻度と相対頻度を算出した (KH Coder, 3.Beta.05.a, 樋口, 2022; Web 茶まめ, 堤・小木曾, 2015; CasualConc, 3.0.4, Imao, 2023, Appendix 1)。最上位に付置した「研究」「関連」「調査」「分析」「生活」「実施」「支援」「検討」「参加」「認知」に対してタグ付けを行ってコンコーダンスラインを表示し、日本語と英語のテキストが

並ぶように抽出した(AntPConc, 1.2.1, Anthony, 2017)。英語抄録において、日本語の「研究」などと同じ機能を持つ部分を先行研究 (Moreno & Swales, 2018, 例 6～10; Quirk et al., 1985, 例 11～13) に基づき精査した。

例6: The data from . . . allowed us to revise. 「研究のポジティブな面に言及」

例7: to obtain a homogenous normal group for intra- and inter-therapist comparison 「根拠を提示」

例8: The results {which have been} quoted are

「縮約関係節での動詞の省略」があり「引用内容とともに扱うメタ談話項目 (Hyland, 2005)」

例9: and rewarding items . . . 「研究への貢献に言及」

例10: there may be two reasons for this 「論理的な関連機能を提示」

(例6～10: Moreno & Swales, 2018, pp. 48–50, 波括弧と丸括弧は原著による)

例11: the *escaped* prisoner [‘the prisoner who has escaped’]

例12: by signing a *peace* treaty

例13: The man *being questioned* by the police was my brother.

(例11～13: Quirk et al., 1985, pp. 413, 657, 1264, イタリック体などは原著による)

英語テキストは、構成素の「品詞 “parts of speech”」(Quirk et al., 1985, p. 68) ごとに分類した。同じ機能を持つ部分が観察されない場合、Absence と記録した。アノテーションを行い品詞別の生起数を集計する過程を、安定するまで繰り返した。

3. 結果

3.1. 英語抄録テキストの語彙項目

英語抄録テキストは「特定のトピック」に左右されにくい「機能語」が高頻度語の上位に付置した (田畑, 2005, p. 195; Anthony, 2019, p. 186; 表2)。また, study, studiesはすべて名詞として使用され (例14, 図1), studied は3件中1件が「論を進めるための語彙」(今村, 2004, p. 10, 例15), 残りの2件は研究対象者の背景となる学習歴などの専門分野に関する情報を示すと考えられた。

例14: However, to our knowledge, there have been no previous studies investigating the relationship between . . . and the quality of end-of-life care evaluated by bereaved families. (JPH023)

しかし、日本で老健施設における遺族による看取りの質の評価と . . . の関連を調査した研究はない。

例15: . . . have not been studied enough whether it is . . . (JPH009)

...について十分検討されていなかった。

表 2 英語抄録テキスト 頻度上位語

Rank	Item	Raw	Std	Rank	Item	Raw	Std	Rank	Item	Raw	Std
1	the	977	62.9	11	as	117	7.5	21	we	66	4.2
2	and	685	44.1	12	health	99	6.4	22	analysis	65	4.2
3	of	671	43.2	13	on	97	6.2	23	age	59	3.8
4	in	358	23.1	14	study	93	6.0	24	support	59	3.8
5	to	352	22.7	15	that	92	5.9	25	high	58	3.7
6	a	233	15.0	16	from	82	5.3	26	p	58	3.7
7	was	197	12.7	17	or	76	4.9	27	their	58	3.7
8	were	193	12.4	18	this	71	4.6	28	community	55	3.5
9	with	189	12.2	19	among	69	4.4	29	between	53	3.4
10	for	167	10.8	20	by	66	4.2	29	years	53	3.4

Raw: 粗頻度, Std: 1,000 語あたりに標準化した相対頻度 (CasualConc, Imao, 2023)

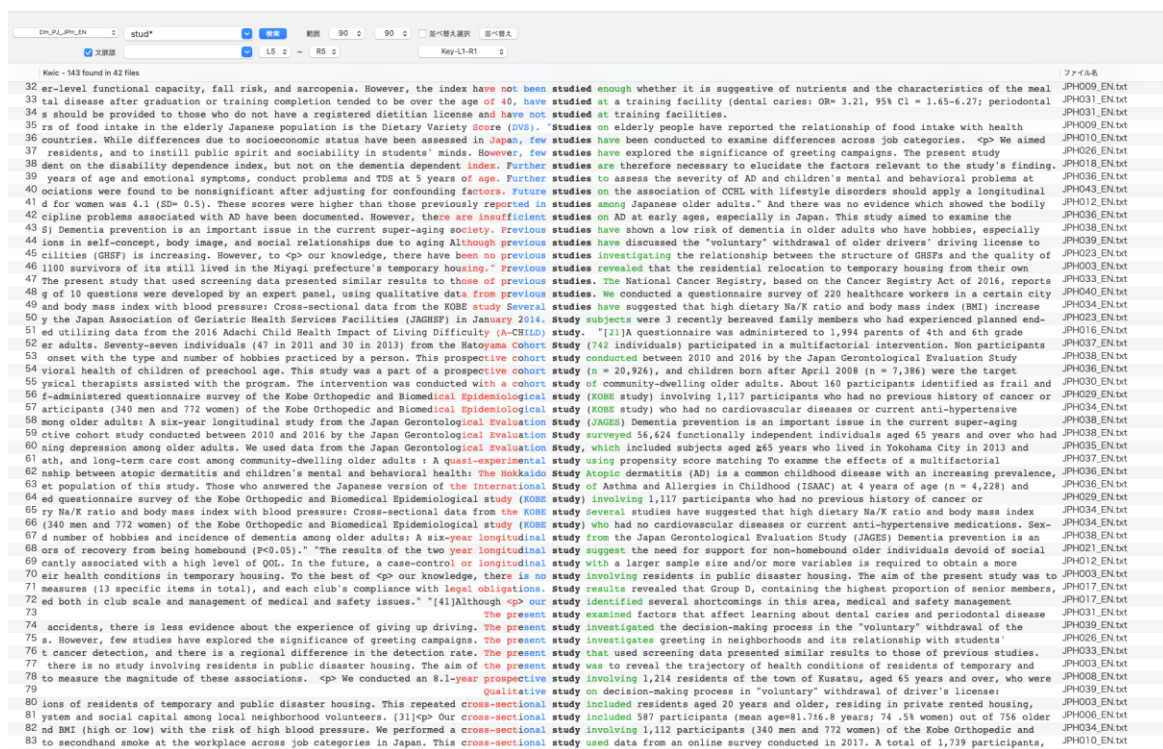


図 1 コンコーダンスラインの例

3.2. 英語抄録における「テキストの断片」の「フレーズの構成素」の語類別頻度

英語抄録において日本語抄録と同じ機能を持つテキストの構成素について、品詞別生起頻度を表3と図2に示す。「動詞」として出現した件数の相対頻度は、「検討」「実施」「関連」の順に頻度が高く「文脈に応じて」(Hyland, 2005, p. 53) 異なる記述が観察された(例16～18)。「参加」「分析」「支援」「調査」「認知」と続き、「研究」と「生活」は出現しなかった。

例16: In this study, we investigated the relationship between ... (JPH022)

そこで、本研究では、... の関連について検討した。

例17: A questionnaire survey on . . . was conducted among . . .

(JPH002)

. . . について書面調査を実施した。

例18: However, the acceptance of support was associated with . . .

(JPH011)

しかし, . . . による支援は, . . . が関連しており, . . .

表 3 英語抄録における日本語抄録と同じ機能を持つテキストの構成素の品詞別相対頻度

Item	Verb	Noun	Adjective	Adverb	Personal pronoun	Determiner	Preposition	Other	Absence
検討	60.0	17.8	0.0	0.0	0.0	0.0	0.0	0.0	22.2
実施	56.1	21.1	1.8	0.0	0.0	0.0	0.0	0.0	21.1
関連	33.3	37.8	4.4	0.0	0.0	0.0	0.0	1.1	23.3
参加	16.7	53.7	1.9	0.0	0.0	0.0	1.9	1.9	24.1
分析	16.0	69.3	1.3	0.0	0.0	0.0	0.0	0.0	13.3
支援	11.5	69.2	0.0	0.0	0.0	1.9	0.0	0.0	17.3
調査	5.7	73.9	0.0	0.0	0.0	1.1	0.0	0.0	19.3
認知	2.3	72.1	18.6	0.0	0.0	4.7	0.0	0.0	2.3
研究	0.0	71.9	0.0	0.0	5.2	4.2	0.0	0.0	18.8
生活	0.0	72.5	8.7	1.4	0.0	0.0	0.0	0.0	17.4

100語当たりの相対頻度

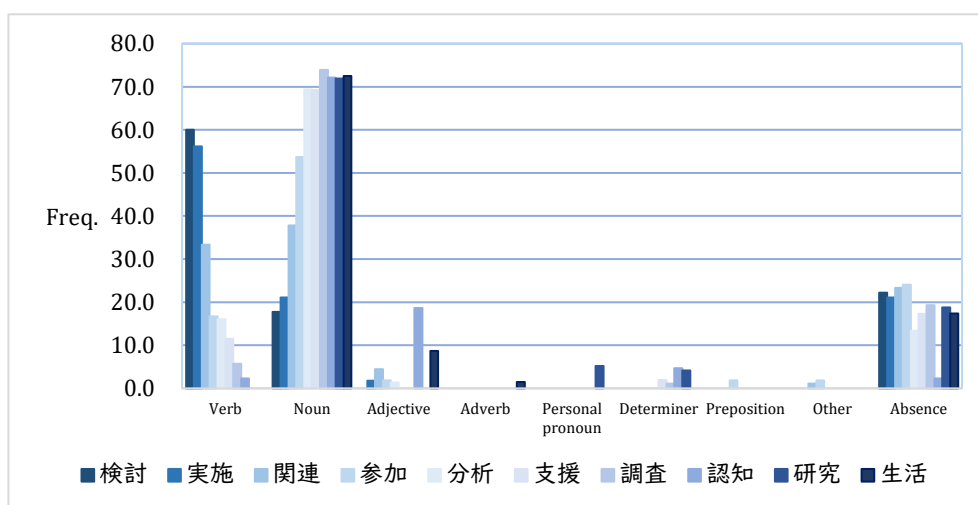


図 2 英語抄録における日本語抄録と同じ機能を持つテキストの構成素の品詞別相対頻度

「参加」「分析」「調査」と同じ機能を持つ英語のテキストは「論を進めるための語彙」(今村, 2004, p. 10)が多く, 様々な動詞が使用された (例 19~21)。「支援」「認知」はすべて「専門分野における語彙」(今村, 2004, p. 10)として用いられた (例 22, 23)。例 23 では, 健康リスクの知識について, 英文には... did not know the term と明示的で, 日本語では和らげた表現が観察された。

例19: Among the total subjects, 758 (16.8%) participated in the YWP.

(JPH035)

対象者のうちYWP参加群は、758人（16.8%）であった。

例20: The relations between these variables were estimated using . . . (JPH012)

QOLとそれ以外の項目との関連は、... を使って分析された。

例21: Sex, age, . . . , and QOL were assessed. (JPH012)

性別と年齢のほかに、... QOLが調査された。

例22: To “solve the issues in the community,” the CSCs “intervened in the community by themselves,”
“supported the autonomy of the residents,” and “connected the residents to resources.”

(JPH032 引用符などは原著による)

地域への働きかけでは、〈CSC自らが地域に働きかける〉だけでなく〈住民の主体性を支援する〉ことや〈住民と資源をつなぐ〉こともCSCは行い、〈地域の課題解決をはかる〉ことへ進んでいた。

例23: In contrast, individuals who were socially isolated and/or frail did not know the term. (JPH024)

一方、フレイル対策が必要な者ではフレイル認知度が低いという実態が明示された。

英語抄録では、日本語抄録と同じ機能を果たす「フレーズの構成素」の多くが名詞として用いられた。日本語の「研究」と同じ機能を有する構成素に人称代名詞 *we* が5件観察され、いずれも日本語の「本研究」含むテキストと同じ機能を有していた(例 24)。英語抄録は研究の目的とその重要性を2文に分けて、より明示的に述べる傾向がみられ、日本語抄録ではより複雑な論理構成で叙述された。

例24: We aimed to investigate the progress of health promotion program implementation. This included . . . ; further progress through focus on the population approach emphasized in Health Japan 21 (2nd) was also discussed. (JPH002)

本研究は、健康日本21（第二次）で重視されるポピュレーションアプローチに着目して、... さらなる推進に向けたあり方を検討することを目的とした。

例 25 は抄録の冒頭で「研究テーマの重要性」(Morley, 2023)を示している。健康リスクに関して、日本語抄録に「生活に幅広く関係する」と記し、英語では *affecting the quality of life* と明示的な叙述がみられた。日本語「コミュニティ」(Van Bonn & Swales, 2007, p. 106)に向けて表現を和らげていると推察された。例 26 においても、英語テキストでは *observation* を用いて簡潔に叙述する傾向が認められた。これらは「抄録[テキスト]の地域性」(Van Bonn & Swales, 2007, p. 106) が反映されている可能性がある。

例25: Atopic dermatitis (AD) is a common childhood disease . . . affecting the quality of life of afflicted children. (JPH036)

小児期のアトピー性皮膚炎は、. . . 生活に幅広く関係する可能性が示唆されている。

例26: However, there is a lack of long-term observation of . . . (JPH003)

しかし、. . . を長期間にわたって調査した研究はほとんどない。

英語テキストでは、日本語抄録の言語項目と同じ機能を持つ構成素が出現しない場合もあった(例 27)。各々の「文脈に応じて」(Hyland, 2005, p. 53) 記述を調整していることが示唆された。

例27: Ebetsu city officials distributed and collected the questionnaires, delivering anonymized data to the researchers. (JPH043)

調査票は、市の職員によって配布・回収が行われた。調査終了後、市から匿名化されたデータを受け取り、分析を実施した。

また、英語抄録テキストには、日本語抄録と同じ機能を有するテキストが様々な構成素で記された(例28～32)。英語抄録テキストにeligibleは1件のみ生起したが、日本語テキストに「分析対象」(例28) は12件出現し、同じ機能を持つ英語テキストの構成素は文脈により多様で、 We analyzed (JPH007), A total of . . . were included (JPH010), . . . were finally for analysis (JPH013), . . . documents were selected (JPH020) などが観察された。

例28: We used data from 4,509 eligible respondents. (JPH035)

. . . 4,509人を分析対象者とした。

英語抄録に、日本語抄録の「参加」と同じ機能を持つテキストの構成素が「方法」セクションの第1文で観察された(例 29)。英語では「研究の概略」(河本・石井, 2022, p. 88) に続いて「研究参加者について説明」(保田, 2021a, p. 80) され「修辭的単位」(Swales, 1981a, p. 41) に従って情報が提示されたが、日本語では「研究参加者について説明」し、「研究の概略」を「コホート研究である Kobe . . . study」と述べて、再び「研究参加者」について「に参加した 1,117 人に」と複雑な論理構成で示された。また、日本語のみに「健康な」と叙述された。英語と日本語のそれぞれの「コミュニティ」(Van Bonn & Swales, 2007, p. 106) を考慮した叙述と推察される。

例29: We conducted a baseline (2010, 2011) self-administered questionnaire survey of the Kobe Orthopedic and Biomedical Epidemiological study (KOBÉ study) involving 1,117 participants who had no previous history of cancer or cardiovascular diseases and were not under treatment for

hypertension, diabetes, or dyslipidemia. (JPH029)

がんや循環器疾患の既往歴がなく、高血圧、糖尿病、脂質異常症の治療中でない、健康な都市住民を対象としたコホート研究である Kobe Orthopedic and Biomedical Epidemiological study (KOBE study) のベースライン調査 (2010～2011年度) に参加した1,117人にK6日本語版による質問紙調査を実施した。

例 30 では、英語も日本語も類似した論理構成で叙述され、「研究の概略」(河本・石井, 2022, p. 88) に続いて「研究参加者について説明」(保田, 2021a, p. 80) された。英語テキストの構成素より、日本語テキストの「調査」が「専門分野における語彙」(今村, 2004, p. 10)の一部を構成することが示された。

例30: A four-year follow-up survey (2014, 2015) was conducted on 1,004 people (follow-up rate of 90%). (JPH029)

4年後の追跡調査が2014～2015年度に実施され、1,004人が参加した (追跡率90%)。

例31は「結論」セクションでの「提言の提示」(水本・浜谷・今尾, 2016, p. 25) を示し、日本語抄録は「支援を経験する」と簡潔に記され、英語では to gain experience providing supportと具体的に叙述されていた。日本語「コミュニティ」(Van Bonn & Swales, 2007, p. 106) には文脈に応じて和らげた短い記述を行い、英語の読み手に対しては明示的に示す方略が示された。英語テキストでは明示的に示す方略が例32においても観察された。日本語抄録では「健康・生活実態調査」と和らげて記されるが英語テキストでは... Child Health Impact of Living Difficulty (A-CHILD) studyと叙述されていた。

例31: In order to . . . , it is necessary . . . to implement initiatives that increase opportunities for people to gain experience providing support. (JPH011)

... には、...支援を経験する機会を増やす取り組みが必要であることが示唆された。

例32: A cross-sectional study was conducted utilizing data from the 2016 Adachi Child Health Impact of Living Difficulty (A-CHILD) study. (JPH016)

東京都足立区で実施された「足立区子どもの健康・生活実態調査」の2016年の調査データを使用し横断研究を行った。

4. 考察

日本語論文の英語抄録では、日本語抄録と同じ機能を果たす構成素が数多く見られたが、「具現化」(Moreno & Swales, 2018, p. 41) の方法には差異が認められた。英語テキストでは「ステッ

プ」(Moreno & Swales, 2018, p. 48)レベルで明瞭な論理構造が示され、日本語抄録では複雑な構成が散見された。これは、先行研究 (Lores-Sanz, 2016; Van Bonn & Swales, 2007) と類似した傾向であった。英語抄録では、日本語の「検討」「実施」「関連」と同じ機能を有する構成素が動詞として高頻度に使用され、一方で「研究」「生活」と同じ役割を果たす部分は主に名詞として用いられた。また、英語テキストでは、繰り返しを避け、明示的な叙述を行う傾向があり、日本語では「コミュニティ」(Van Bonn & Swales, 2007, p. 106) に配慮した、和らげた表現が用いられた。

英語抄録には “This study aimed to” (15件), “The purpose of this study is/was to” (10件)などの「意図的なメタ談話記述」(Van Bonn & Swales, 2007, p. 98) が顕著に多かった。本研究の対象誌は、国際誌と同等の「ディスコース・コミュニティの規模や専門的な成熟度」(Van Bonn & Swales, 2007, p. 105) を有さない可能性があり、一概に比較できないと考えられるが、これらは「日本人」著者(田中・藤井・冨浦・徳見, 2006, p. 76; 小林・田中・冨浦, 2011, p. 2) による国際誌掲載英語論文の特徴を調査した研究より得られた知見と類似した傾向を示している。田中他 (2006)の研究では、日本人著者には名詞による名詞の修飾などが高頻度に使用され、小林他 (2011) の研究では、*on the other hand*など (p. 5, イタリック体は原著による) のメタ談話標識 (Hyland, 2005)が多く観察されている。近年の「学術の文脈」のグローバル化 (Hyland & Shaw, 2016; Lores-Sanz, 2016; Mauranen, 2019; Mauranen, Hynninen, & Ranta, 2016) を踏まえると、本研究で観察された特徴は「日本語を母語とし、日本で初等、中等教育課程を経て、仕事で英語を使用するもの」(Fujiwara, 2007, p. 58; 藤原, 2012, p. 44)によるテキストとして、今後、国際的な「学術誌」に「容認され、[査読を]通過する」(Flowerdew & Habibie, 2022, p. 44)可能性があり、さらに「日本の英語学習者のための参考例 “reference sample[s] for Japanese learners of English”」(Hino, 2018, p. 61) となり得る。

二言語コーパス研究に関して、専門的で「修辭的文脈に高度に適合した」(Hyland, 2005, p. 113) テキストを分析することの困難さが報告されている (Lavid, 2017)。Lavidの研究 (2017) では英語とスペイン語の「特徴」(p. 11) を考慮しつつ質的な検討を行う必要性が生じたことにより、「アノテーションの過程」(p. 11) が複雑であったと報告されている。今後、最新のデジタル技術を活用するなど、効率的かつ頑健な方法を模索したい。そして、言語使用の「領域」での「談話」を「適切にテキスト化する “to appropriately textualize”」するための「形式 “form”」(Widdowson, 2020, p. 111)にどのような関心が向けられているかなどの潮流を踏まえて探求を続けたい。

Appendix 1: 日本語抄録における「サ変名詞」(今村, 2004, p. 10) の生起頻度

Rank	Item	Raw	per 1,000	Rank	Item	Raw	per 1,000	Rank	Item	Raw	per 1,000
1	研究	96	5.04	11	関係	41	2.15	21	介入	26	1.36
2	関連	90	4.72	12	活動	41	2.15	22	施設	26	1.36
3	調査	87	4.57	13	介護	38	1.99	23	評価	25	1.31
4	分析	75	3.94	14	回答	35	1.84	24	行動	24	1.26
5	生活	69	3.62	15	摂取	35	1.84	25	利用	23	1.21
6	実施	57	2.99	16	信頼	34	1.78	26	解析	22	1.15
7	支援	52	2.73	17	研修	31	1.63	27	平均	22	1.15
8	検討	45	2.36	18	喫煙	29	1.52	28	連携	22	1.15
9	参加	45	2.36	19	予防	28	1.47	29	影響	21	1.10
10	認知	43	2.26	20	回帰	27	1.42	30	自立	21	1.10

Raw: 粗頻度, per 1,000: 1,000 語当たりの相対頻度 (頻度上位 30 語を表示)

補遺

I. English for specific purposes (ESP) 研究のはじまりと学術英語 (EAP) 執筆支援

ESP 研究は、学習者が多用な文脈に「応用可能な方略」(Hyon, 2018, p. 115; Widdowson, 1981, p. 5) を習得できるように、特定の「機能」(Hutchinson & Waters, 1987, p. 31; Hyland, 2012, p. 18; Johns, 1997, p. 124; Swales, 2004, p. 228) を果たすテキストに焦点を当てて行われている。

初期の研究では「特定のジャンル」の「熟練した」執筆者によるテキストの理解に「修辭的関係」が重要であることが判明し、学習者支援の必要性が示されている (Lackstrom, Selinker, Trimble, 1973, p. 128; Selinker, Todd-Trimble, Trimble, 1976, p. 282; Selinker, Todd-Trimble, Trimble, 1978, p. 315)。例えば、化学分野のテキストでの“-en participle”(Swales, 1981b, p. 41)の使用については「談話の以前の部分を参照するため」だけに用いられるわけではないとされる (p. 44)。これは、Quirk et. al. (1972) を引用して「過去分詞が名詞の前にくる」のは「前文で説明された事項に言及する場合」に限られる(Swales, 1971/1983, 菅原訳 p. 181)とされた観点から、学術英語の「理解」(Swales, 1981b, p. 42)に進展がみられたことを示している。

学術の世界の国際化で「英語を選んで使用する」(Mauranen, 2019, p. 11) 傾向が強まる中で「学術の文脈」での「コミュニケーション」「English for academic purposes (EAP)」は「英語教育とその研究の重要な力となってきた」(Hyland & Shaw, 2016, p. 1; 寺内, 2022)とされる。「学術の場で遭遇し、産出することが期待される英語は、ジャンル別、発信の段階別、専門分野別など、学術の場の外でみられるものと認知的に大きく異なっている」という見解が強く主張され、学習を促すためには「時間、資源、専門家の協力と尊敬が必要」と説かれている (Hyland & Shaw, 2016, p. 2)。

「専門家と初心者」における「ジャンル」の「修辭的、特に形式的な特徴」(Shaw, 2016, p. 243) の認識の低さを解消することを目指して、「ジャンルに基づく」(東條, 2015, p. 19)コーパス研究が国内外で広く行われ (Flowerdew & Habibie, 2022; Kunioshi, Noguchi, Hayashi, & Tojo, 2012; Maswana, Kanamaru, & Tajino, 2015; Tardy, 2009; 石川, 2021; 水本・浜谷・今尾, 2016; 保田, 2021b), 執筆支援 (Robinson, Stoller, Constanza- Robinson, & Jones, 2008; Swales & Feak, 2012; Feak, 2019; 河本・石井 2018 & 2022; 中谷, 2016; 保田, 2021a; 藤岡, 2022) および EAP 教員支援 (マスワナ紗矢子・渡寛法・山田浩・飯島優雅・高橋幸・金丸敏幸・寺内一・田地野彰, 2022) に寄与していると考えられる。これらの研究では「ムーブとステップ」の「分析」や「語彙文法的特徴」の調査が主に「手作業またはコンコーダンスを用いて」(Shaw, 2016, p. 252) 行われている。

II. ディスコース・コミュニティ

「ディスコース・コミュニティ」(Swales, 1990, pp. 9 & 24; 寺内, 2010, p. 15)は「同じ目的や目標」を有し「社会修辭学的ネットワーク」を形成するとされる。構成員は「会話」や「文書」などの「コミュニケーション」を通じて「結びつけ」られ (野口, 2017, p. 253), 「目的を達成」するために「ジャンル」テキストを使用する (Hyon, 2018, p. 15)。かつては、論文などのジャンルテキストの「熟練した」執筆者として「英語母語話者」(Selinker, Todd-Trimble, Trimble, 1976, p. 282)が

考えられた。「ディスコース・コミュニティ」の構成員は, Swales(1990) によって「共通の公的な目的」(p. 24) のためにコミュニケーションを行う人々と定義され, ジャンルテキストは「学術言語の変異という困難かつ重要な問題に対して示唆を与える」(p. 67) とされる。2000 年代には「英文学術誌 “Anglophone journals” への投稿を志す英語を母語としない」研究者を「英語を母語とする」研究者が支援する傾向が報告され (Swales, 2004, p. 46), 「コーパスに基づく」研究によって「学問的な上級者」が「英語母語話者」の助けを借りずに「言語に関してコメントしたり修正を促したりすることがある」ことが明らかにされている(Mauranen, Hynninen, & Ranta, 2016, p. 48)。Tribble(2017)は, ディスコース・コミュニティの主要な区分は「英語母語話者と非母語話者」より, むしろ「熟練者と初学者」(p. 40)であると指摘している。

III. ムーブとステップ

ムーブとは, 論文などにおいて「特定の働きを担う」(田地野・寺内・金丸・マスワナ・山田, 2008, p. 117) テキストの「まとまり」(水本・浜谷・今尾, 2016, p. 22) を指す。論文の緒言部において, テキストに共通の「修辭的機能」を有する談話の運びが特定され「ムーブ」(Swales, 1981a, p. 21; 1990, p. 140; 2004, p. 228) と称された。「ムーブ」を構成するテキストにおいて「ジャンルの目的を果たすために論を進める」機能を持つ「テキストの断片」を「ステップ」(Moreno & Swales, 2018, p. 40; Swales, 1990, p. 178) という。ステップは「少なくともひとつの動詞」を「定形, 非定形, 省略, 動詞句に変換可能な名詞化」などの形で有する (Moreno & Swales, 2018, p. 49) 。ステップは “embedded moves” (Tanko, 2017, p. 48) としても知られ「1 つの節に異なる修辭的機能」を示すことがある。

謝辞

本研究は科研費基盤研究(C)2018–2022 年度 (課題番号 18K02966), 令和 4 年度 (第 17 回) 鈎奨学基金研究助成金 (大阪医科薬科大学), 科研費基盤研究(C)2023–2026 年度(課題番号 23K02800) の助成を受けた。本研究を進める過程で貴重な助言をくださった多くの先生方に感謝申し上げる。本稿の誤りはすべて筆者の責任である。

文献

- Aijmer, K., & Lewis, D. (2017). Introduction. In K. Aijmer & D. Lewis (Eds.), *Contrastive analysis of discourse-pragmatic aspects of linguistic genres*. (pp. 1–9.) Springer.
- Anthony, L. (2017). AntPConc (Version 1.2.1) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>
- Anthony, L. (2019). Tools and strategies for Data-Driven Learning (DDL). In K. Hyland & L. L. C. Wong (Eds.), *Specialised English*. (pp. 179–194.) Routledge.
- Anthony, L. (2022). TagAnt (Version 2.0.5) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>
- Borin, L. (2002). . . . and never the twain shall meet? In L. Borin (Ed.), *Parallel corpora, parallel worlds. Selected papers from a symposium on parallel and comparable corpora at Uppsala University, Sweden, 22–23 April, 1999*. (pp.1–43.) Rodopi.

- Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2), 348–393.
- Connor, U. (2002). New directions in contrastive rhetoric. *TESOL Quarterly*, 36(4), 493–510.
- Connor, U. (2004). Intercultural rhetoric research: Beyond texts. *English for specific purposes*, 3, 291–304.
- Connor, U., & Ene, E. (2019). Does everyone write the five-paragraph essay? In N. A. Caplan & A. M. Johns (Eds.), *Changing practices for the L2 writing classroom*. (pp. 42–63.) The Michigan University Press.
- Connor, U., & Moreno, A. I. (2005). Tertium comparationis: A vital component in contrastive research methodology. In P. Bruthiaux, D. Atkinson, W. G. Eggington, W. Grabe, & V. Ramanathan (Eds.), *Directions in applied linguistics: Essays in honor of Robert B. Kaplan*. (pp. 153–164.) Multilingual Matters.
- Culo, O., Hansen-Schirra, S., Maksymski, K., & Neumann, S. (2017). Empty links and crossing lines: Querying multi-layer annotation and alignment in parallel corpora. In S. Hansen-Schirra, S. Neumann, & O. Culo (Eds.), *Annotation, exploitation, and evaluation of parallel corpora: TC3 I (Translation and multilingual natural language processing 3)*. (pp. 53–88.) Language Science Press.
- Davis, M., & Morley, J. (2018). Facilitating learning about academic phraseology: Teaching activities for student writers. *Journal of Learning Development in Higher Education*. 1–17.
- Feak, C. B. (2019). Writing for disciplinary communities. In N. A. Caplan & A. M. Johns (Eds.), *Changing practices for the L2 writing classroom*. (pp. 178–199.) The University of Michigan Press.
- Flowerdew, J., & Habibie, P. (2022). *Introducing English for research publication purposes*. Routledge.
- Fujiwara, Y. (2007). Compiling a Japanese user corpus of English. *English corpus studies*, 14, 55–64.
- Google Scholar. (2023). Top publications: Japanese. Retrieved from: <https://scholar.google.com/>
- Hino, N. (2018). *EIL education for the Expanding circle: A Japanese model*. Routledge.
- Hirsch, J. E., (2012). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46), 16569–16572.
- Hutchinson, T., & Waters, A. (1987). *English for specific purposes*. Cambridge University Press.
- Hyland, K. (2005). *Metadiscourse*. Continuum.
- Hyland, K. (2012). Disciplinary specificity: Discourse, context, and ESP. In D. Belcher, A. M. Johns, & B. Paltridge (Eds.), *New directions in English for specific purposes research*. (pp. 6–24.) The University of Michigan Press.
- Hyland, K., & Shaw, P. (2016) Introduction. In K. Hyland & P. Shaw. (Eds.), *The Routledge handbook of English for academic purposes*. (pp. 1–13.) Routledge.
- Hyland, K., & Tse, P. (2005). Hooking the reader: A corpus study of evaluative that in abstracts. *English for Specific Purposes*, 24, 123–139.
- Hyon, S. (2018). *Introducing genre and English for specific purposes*. Routledge.
- Imao, Y. (2023). CasualConc (Version 3.0.4) [Computer software]. Retrieved from <https://sites.google.com/site/casualconcj/>
- Johns, A. M. (1997). *Text, role and context*. Cambridge University Press.
- Kaplan, R. B. (1966). Cultural thought patterns in intercultural education. *Language Learning*, 16, 1–20.
- Kenning, M-M. (2010). What are Parallel and Comparable Corpora and How Can We Use Them? In A. O’Keeffe and M. McCarthy (Eds.), *Routledge Handbook of Corpus Linguistics*. (pp. 487–500.) Routledge.
- Kunioshi, N., Noguchi, J., Hayashi, H. & Tojo, J. (2012). An online support site for preparation of oral presentations in science and engineering. *European Journal of Engineering Education*, 37(6), pp. 600–608.
- Lackstrom, J. E., Selinker, L., & Trimble, L. (1973). Technical Rhetorical Principles and Grammatical Choice. *TESOL Quarterly*, 7(2), 127–136.

- Lado, R. (1957). *Linguistics across cultures: applied linguistics for language teachers*. University of Michigan Press.
- Lavid, J. (2017, January 19). Annotating Complex Linguistic Features in Bilingual Corpora: The Case of MULTINOT. In T. Declerck & S. Kübler (Eds.), *Proceedings of the Workshop on Corpora in the Digital Humanities (CDH 2017), Bloomington, USA* (pp. 10–18.) CEUR-WS.org 2017.
- Lavid, J., & Moraton, L. (2016). Annotating metadiscourse markers in the English-Spanish MULTINOT corpus: Preliminary steps. In L. Degand, C. Der, P. Furko, & B. Webbe (Eds.), *Conference handbook of textlink – structuring discourse in multilingual europe second action conference*, (pp. 79–81.) Debrecen University Press.
- Lavid Lopez, J. (2019). Discourse annotation in the MULTINOT corpus: Issues and challenges. In I. Doval & M. T. Sanchez Nieto (Eds.), *Parallel corpora for contrastive and translation studies: new resources and applications*. (pp. 159–182.) John Benjamins.
- Lores, R. (2004). On RA abstracts: From rhetorical structure to thematic organization. *English for specific purposes*, 23, 280–302.
- Lores-Sanz, R. (2016). ELF in the making? Simplification and hybridity in abstract writing. *Journal of English as a lingua franca*, 5(1), 53–81.
- Maswana, S., Kanamaru, T., & Tajino, A. (2015). Move analysis of research articles across five engineering fields: What they share and what they do not. *Ampersand*, 2, 1–11.
- Mauranen, A. (1993). *Cultural differences in academic rhetoric: A textlinguistic study*. Peter Lang.
- Mauranen, A. (2012). *Exploring ELF: Academic English shaped by non-native speakers*. Cambridge University Press.
- Mauranen, A. (2019). Academically speaking: English as the lingua franca. In K. Hyland & L. L. C. Wong. (Eds.), *Specialised English*. (pp. 9–21.) Routledge.
- Mauranen, A., Hynninen, N., & Ranta, E. (2016). English as the academic lingua franca. In K. Hyland & P. Shaw. (Eds.), *The Routledge handbook of English for academic purposes*. (pp. 44–55.) Routledge.
- Moreno, A. I., & Swales, J. M. (2018). Strengthening move analysis methodology towards bridging the function-form gap. *English for Specific Purposes*, 50, 40–63.
- Morley, J. (2023). Academic Phrasebank. The University of Manchester. Retrieved from <https://www.phrasebank.manchester.ac.uk/>
- National Library of Medicine. (2023). PubMed. Retrieved from <https://pubmed.ncbi.nlm.nih.gov/>
- Nwogu, K. (1997). The medical research paper: Structure and functions. *English for specific purposes*, 16(2), 119–138.
- Pho, P. D. (2008). Research article abstracts in applied linguistics and educational technology: a study of linguistic realizations of rhetorical structure and authorial stance. *Discourse studies*, 10(2), 231–250.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1972). *A grammar of contemporary English*. Longman.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. Longman.
- Rees, G. (2022). Using Corpora to Write Dictionaries. In A. O’Keeffe and M. McCarthy (Eds.), *Routledge Handbook of Corpus Linguistics. Second Edition*. (pp. 387–404.) Routledge.
- Robinson, M. S., Stoller, F. L., Constanza-Robinson, M. S., & Jones, J. K. (2008). *Write like a chemist*. Oxford University Press.
- Selinker, L., Todd Trimble, M., & Trimble, L. (1976). Presuppositional rhetorical information in EST discourse. *TESOL Quarterly*, 10(3), 281–290.
- Selinker, L., Todd Trimble, M., & Trimble, L. (1978). Rhetorical function-shifts in EST discourse. *TESOL Quarterly*, 12(3), 311–320.
- Shaw, P. (2016) Genre analysis. In K. Hyland & P. Shaw. (Eds.), *The Routledge handbook of English for academic purposes*. (pp. 243–255.) Routledge.

- Swales, J. (1971). *Writing scientific English: A textbook of English as a foreign Language for students of physical and engineering sciences*. Nelson. (スウェイルス, J. 菅原基晃 (訳) (1983). 科学英語の書き方 日経サイエンス)
- Swales, J. M. (1981a). *Aspects of article introductions*. The Language Studies Unit, Aston University.
- Swales, J. M. (1981b). The function of one type of particle in a chemistry textbook. In L. Selinker, E. Tarone, & V. Hanzeli (Eds.), *English for academic and technical purposes: Studies in honor of Louis Trimble*. (pp. 40–52.) Newbury House Publishers.
- Swales, J. M. (1990). *Genre analysis*. Cambridge University Press.
- Swales, J. M. (2004). *Research genres*. Cambridge University Press.
- Swales, J. M. (2017). Standardisation and its discontents. In M. Cargill & S. Burgess (Eds.), *Publishing research in English as an additional language: practices, pathways and potentials*. (pp. 239-253.) The University of Adelaide Press.
- Swales, J. M. & Feak, C. B. (2012). *Academic writing for graduate students* (3rd ed.). The University of Michigan Press.
- Tanko, G. (2017). Literary research article abstracts: An analysis of rhetorical moves and their linguistic realizations. *English for academic purposes*, 27, 42–55.
- Tardy, C. M. (2009). *Building genre knowledge*. Parlor Press.
- Tribble, C. (2017). ELFA vs. genre: A new paradigm war in EAP writing instruction? *English for academic purposes*, 25, 30–44.
- Van Bonn, S., & Swales, J. M. (2007). English and French journal abstracts in the language sciences: Three exploratory studies. *English for academic purposes*, 6, 93–108.
- Widdowson, H. G. (1981). English for specific purposes: Criteria for course design. In L. Selinker, E. Tarone, & V. Hanzeli (Eds.), *English for academic and technical purposes: Studies in honor of Louis Trimble*. (pp. 1–11.) Newbury House Publishers.
- Widdowson, H. G. (2020). Historical perspectives on ELF. In J. Jenkins, W. Baker, & M. Dewey (Eds.), *The Routledge handbook of English as a lingua franca*. (pp. 101–112.) Routledge.
- Wingard, P. (1981). Some verb forms and functions in six medical texts. In L. Selinker, E. Tarone, & V. Hanzeli (Eds.), *English for academic and technical purposes: Studies in honor of Louis Trimble*. (pp. 53–64.) Newbury House Publishers.
- 浅野元子, 松田紀子 (2022) 「研究論文における英語と日本語のアブストラクトは読み手を意識しているか」 『英語コーパス学会大会予稿集 2022』 37–42.
- 石川有香 (2021). 「工学系英語論文要旨の Move 分析」 『ジャンルとしての工学英語』 大学教育出版
- 今井むつみ, 岡田浩之, 野島久雄 (2012). 『新 人が学ぶということ: 認知学習論からの視点』 北樹出版
- 今村和宏 (2004). 「社会科学系基礎論文におけるサ変名詞のふるまい-作文指導への指針と「専門用語化指数」の試案-」 『専門日本語教育研究』 6. 9–16.
- 大津由紀雄, 吉田研作 (2023). 「第 1 章 大津由紀雄, 吉田研作」 ことばのまなび工房 (監修) 若林茂則 (編) 『英語の教室で何ができるか』 (pp. 1–53.) 開拓社.
- 河本健, 石井達也 (2018). 『トップジャーナル 395 編の型で書く医学英語論文』 羊土社
- 河本健, 石井達也 (2022). 『ライフサイエンストップジャーナル 300 編の型で書く英語論文』 羊土社
- 小林雄一郎, 田中省作, 富浦洋一. (2011). 「ランダムフォレストを用いた英語科学論文の分類と評価」

『情報処理学会研究報告』 6. 1-8.

田地野彰, 寺内一, 金丸敏幸, マスワナ紗矢子, 山田浩 (2008). 「英語学術論文執筆のための教材開発に向けて: 論文コーパスの構築と応用」 『京都大学高等教育研究』 14. 111-121.

田中省作, 藤井宏, 富浦洋一, 徳見道夫. (2006). 「NS/NNS 論文分類モデルに基づく日本人英語科学論文の特徴抽出」 『英語コーパス研究』 13. 75-87.

田畑智司 (2005). 「コーパスに基づく文体論研究」 齊藤俊雄・中村純作・赤野一郎 (編) 『英語コーパス言語学 改訂新版』 (pp. 183-206.) 研究社

堤智昭, 小木曾智信 (2015). 「歴史的資料を対象とした複数の UniDic 辞書による形態素解析支援ツール『Web 茶まめ』」 『じんもんこん 2015 論文集』 179-184. <http://id.nii.ac.jp/1001/00146542/>

寺内一 (2010). 「ESP の歴史と定義」 寺内一, 山内ひさ子, 野口ジュディー, 笹島茂 (編) 『21 世紀の ESP』 (pp. 3-16.) 大修館書店

寺内一 (2022, 10 月 15 日). 「ポストコロナとコミュニケーション形態の変化—ジャンルの重要性—」 大学英語教育学会 2022 年度第 2 回関西支部講演会 オンデマンド開催

東條加寿子 (2015). 「大学英語教育の中のジャンル分析」 『大阪女学院大学紀要』 12. 17-26.

中谷安男 (2016). 『大学生のためのアカデミック英文ライティング』 大修館書店

中野愛実, 宮崎佳典, 浅野元子, 藤枝美穂 (2022). 「医学系論文抄録の日英対訳コーパスを活用した医学英語教育支援システム開発」 『教育システム情報学会第 47 回全国大会講演論文集』 15-16.

仁科恭徳 (2023). 『パラレルコーパス言語学の諸相』 開拓社

日本公衆衛生雑誌編集委員会 (2022). 「日本公衆衛生雑誌投稿規程」 『日本公衆衛生雑誌』 69(10). 841-846.

野口ジュディー (2017). 「科学論文のスタイル」 豊田昌倫, 堀正広, 今林修 (編) 『英語のスタイル』 (pp. 252-260.) 研究社

服部真弓 (2011). 「英語リーディング授業での母語有効活用を考察する」 『高専教育』 34. 423-428.

樋口耕一 (2022). KH Coder Available from <https://kncoder.net/>

藤岡真由美 (2022). 「大学院学術目的の英語 (EAP) ライティング授業: 研究にもとづいた授業実践の必要性和課題」 『大阪府立大学高等教育推進機構論文集 言語と文化』 21. 21-32.

藤原康弘 (2012). 「コーパス言語学と国際英語関連分野 (EIL, WE, ELF)」 の学際的領域—英語使用者コーパスの必要性— 『外国語研究』 45. 21-52.

マスワナ紗矢子, 渡寛法, 山田浩, 飯島優雅, 高橋幸, 金丸敏幸, 寺内一, 田地野彰. (2022). 「日本の大学における EAP 教員を対象とした自己評価ツールの開発」 *JAAL in JACET Proceedings* 4. 72-77.

水本篤 (2017). 『英語学術論文執筆支援ツール AWSuM マニュアル』 Available from http://mizumot.com/lablog/wp-content/uploads/2016/12/AWSuM-Manual_J.pdf

水本篤, 浜谷佐和子, 今尾康裕 (2016). 「ムーブと語連鎖を融合させたアプローチによる応用言語学論文の分析－英語学術論文執筆支援ツール開発に向けて」 『英語コーパス研究』 23. 21–32.

保田幸子 (2021a). 『英語科学論文をどう書くか』 ひつじ書房

保田幸子 (2021b). 「科学論文における主観性：アカデミック・ディスコース概念の再考」 『日本教育工学会論文誌』 45(1). 1–13.

柳父章 (1982). 『翻訳語成立事情』 岩波書店

調査対象論文題目

JPH001	Policy recommendation for the regulation of heated tobacco products based on evidence review of their health effects and influence on tobacco control	JPH023	Factors related to bereaved family's satisfaction with end-of-life care at geriatric health services facilities
JPH002	Current status and issues around the process of health promotion program implementation by the municipal office of health and welfare: Survey results of all municipalities of six prefectures	JPH024	Awareness of the term "frailty" and its correlates among older adults living in a metropolitan area
JPH003	Trajectory of evacuees' health condition after the Great East Japan Earthquake: Health survey of residents in prefabricated temporary housing and private rented housing in the Miyagi prefecture	JPH025	Elements that cancer peer supporters working in Japanese hospitals consider to be important in helping them perform their role
JPH004	Correlations among resilience against violent behavior toward the self and others, rumination, and anger in high school students	JPH026	School students' greeting behavior and its association with their community attachment and helping behavior
JPH005	Development of the Japanese version of the cognitive failure scale (the Short Inventory of Minor Lapses) for Japanese women involved in child care	JPH027	Development of livelihood difficulties assessment scales for clinicians
JPH006	Association between frailty in community-dwelling older adults certified as requiring support in the long-term care insurance system and social capital among local neighborhood volunteer	JPH028	Medical expenses for diabetes care in Japan: Analysis of inter-prefecture differences
JPH007	Development of the family empowerment scale for parents with toddlers	JPH029	The search for a predictor of deterioration of the nonspecific stress index K6 among urban residents: The KOBE study
JPH008	Impact of lifestyle-related diseases and frailty on the incidence of loss of independence in Japanese community-dwelling older adults: A longitudinal study on aging and health in Kusatsu	JPH030	An implementation study of a program supporting frailty-prevention community activities using the "Community-as-Partner" model
JPH009	Food diversity and its relationship with nutrient intakes and meal days involving staple foods, main dishes, and side dishes in community-dwelling elderly adults	JPH031	Factors that affect dietitians and registered dietitians' learning about dental caries and periodontal disease
JPH010	Differences in nonsmokers' exposure to secondhand smoke at the workplace across job categories in Japan: A cross-sectional study	JPH032	The process of community building by Community Social Coordinators: The possibility of enhancing social capital in a disaster-affected city
JPH011	Examining attitudes toward mutual support in daily life and their associated factors within a community-based integrated care system: Findings of the "Survey to enrich the lives of Miyamae ward residents"	JPH033	Secular trends of breast cancer detection rate in Japan: Age, period, and region effects from 2004 to 2015
JPH012	Jogging / running activity and quality of life in the elderly	JPH034	Association of the combined estimated 24-h urinary Na/K ratio and body mass index with blood pressure: Cross-sectional data from the KOBE study
JPH013	The analysis of newspaper articles' reporting of legislation relating to cancer registries	JPH035	Effectiveness of walking point projects with incentives for walking time, physical function, and depression among older people: inverse probability of treatment weighting using propensity scores
JPH014	Access to grocery stores and nutrition / food intake observed in the national health and nutrition survey: Focusing on the substitution-complementary relation	JPH036	Relationship between atopic dermatitis and children's mental and behavioral health: The Hokkaido Study
JPH015	Relation between concrete activities and results of activities planned in conferences of the regional and occupational health fields in secondary medical care zones	JPH037	Effects of a multifactorial intervention for improving frailty on risk of long-term care insurance certification, death, and long-term care cost among community-dwelling older adults: A quasi-experimental study using propensity score matching
JPH016	Factors related to the passive attitude of caregivers on dental care visit when elementary and junior high school children are diagnosed with dental caries	JPH038	Types and number of hobbies and incidence of dementia among older adults: A six-year longitudinal study from the Japan gerontological evaluation study (JAGES)
JPH017	Recent trends and challenges for older adults at community sport clubs in Japan: Analysis of FY 2016 survey results on comprehensive community sports clubs	JPH039	Qualitative study on decision-making process in "voluntary" withdrawal of driver's license: Transitions in self-concept, body image, and social relationships due to aging
JPH018	Life prognosis of persons with dementia: Observation of persons registered for long-term care insurance in a community in Japan	JPH040	Feelings of difficulty experienced by home healthcare support workers in collaborative practice with medical professionals: Scale development
JPH019	Survey on tobacco sales in newly registered insurance pharmacies and pharmacies in convenience stores in Japan	JPH041	Factors related to work-family conflicts of employees in small and medium-sized businesses
JPH020	Concept analysis of "Go Jo" (mutual help)	JPH042	Characteristics of older people willing to offer lifestyle support to other community members living in rural areas with heavy snowfall
		JPH043	Association of health literacy with hypertension, diabetes, and dyslipidemia: A cross-sectional survey of a regional Japanese community

JPH021	Changes in homebound status and related factors in community-dwelling older adults participating in physical checkups over two years	JPH044	Examination of the effects of a role-playing training program for the improvement of planning and management skills of public health nurses
JPH022	Relationship between prolonged media usage and lifestyle habits among junior and senior high school students	JPH045	Marriage and fertility rates of Japanese women according to employment status: An age-period-cohort analysis

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、テキストの分析と結果の提示だけでなく、テキストマイニングを通じて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。テキ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、エンターテインメント、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し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 をこに語学の授業で使用するもの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、テキ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

No.	Student			DeepL			ChatGPT		
	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 を見ることで、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。

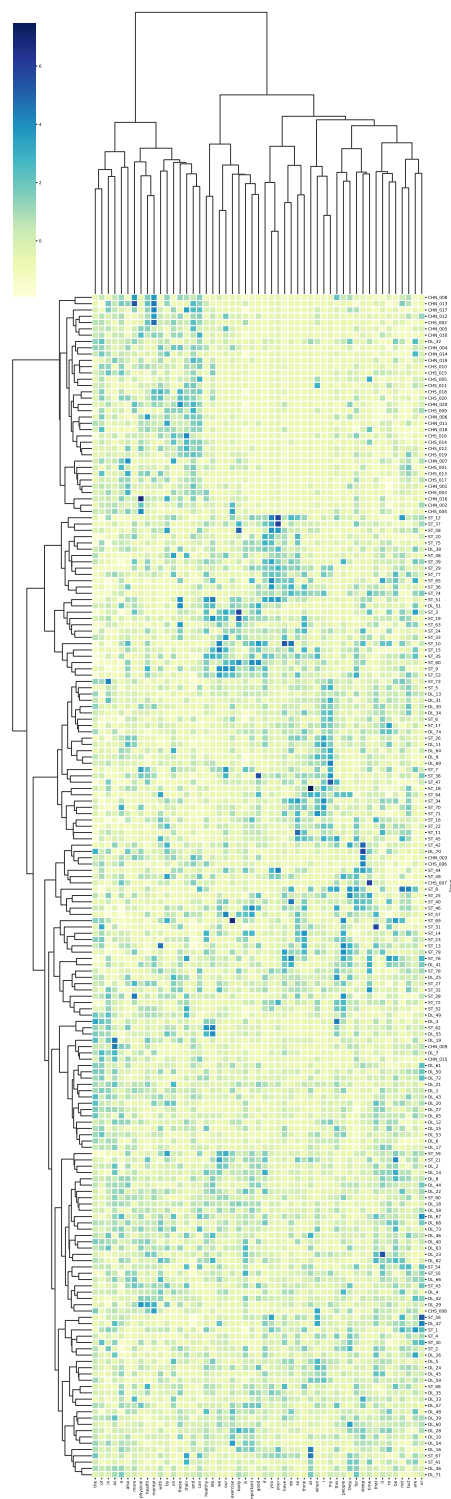


Fig. 1: 3 コーパス上位 50 語の HM 図 (Euclidian 距離, Ward 法)

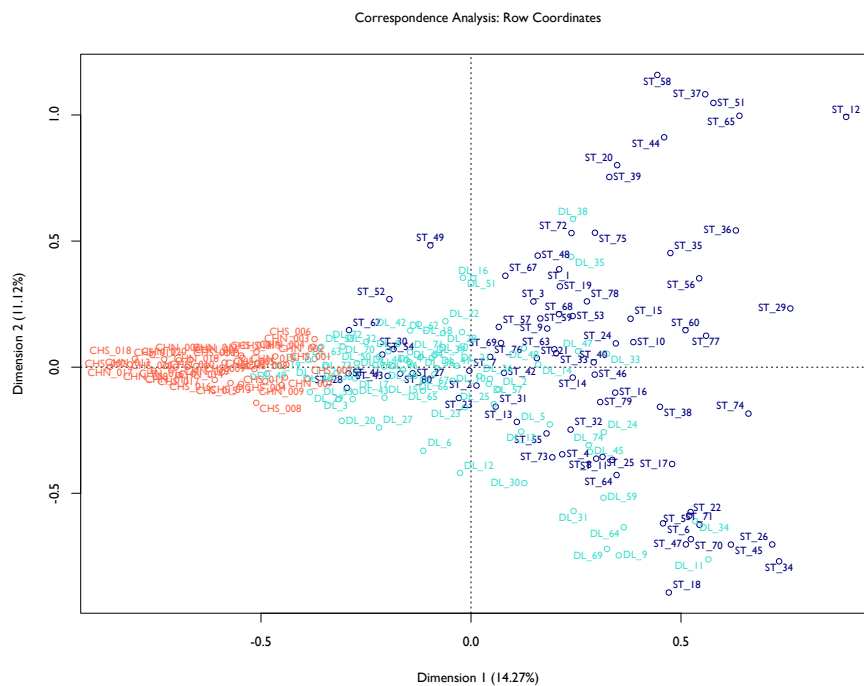


Fig. 2: 3 コーパス上位 50 語の CA (Row)

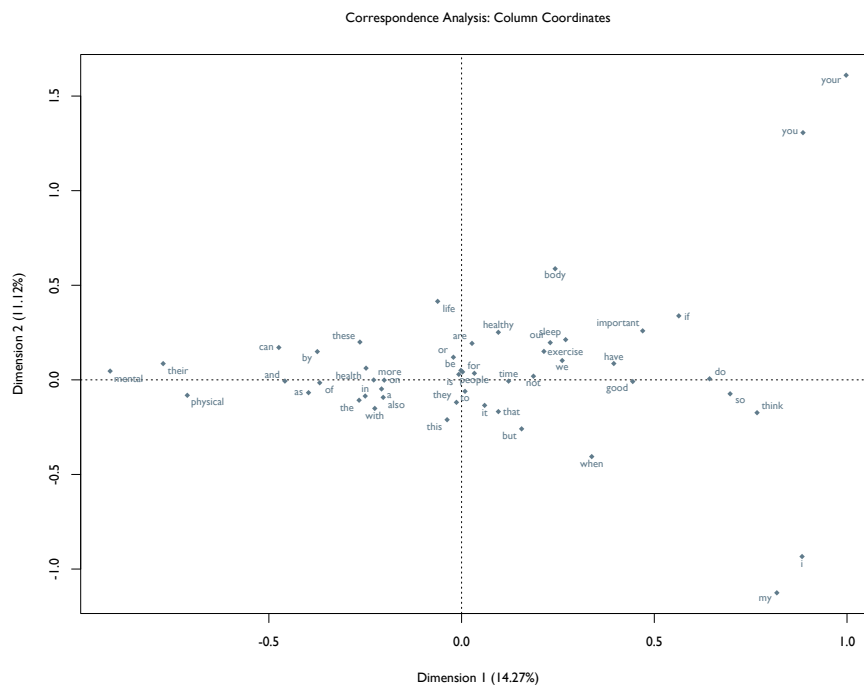


Fig. 3: 3 コーパス上位 50 語の CA (Column)

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も現れ、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に起しやすであろう *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。

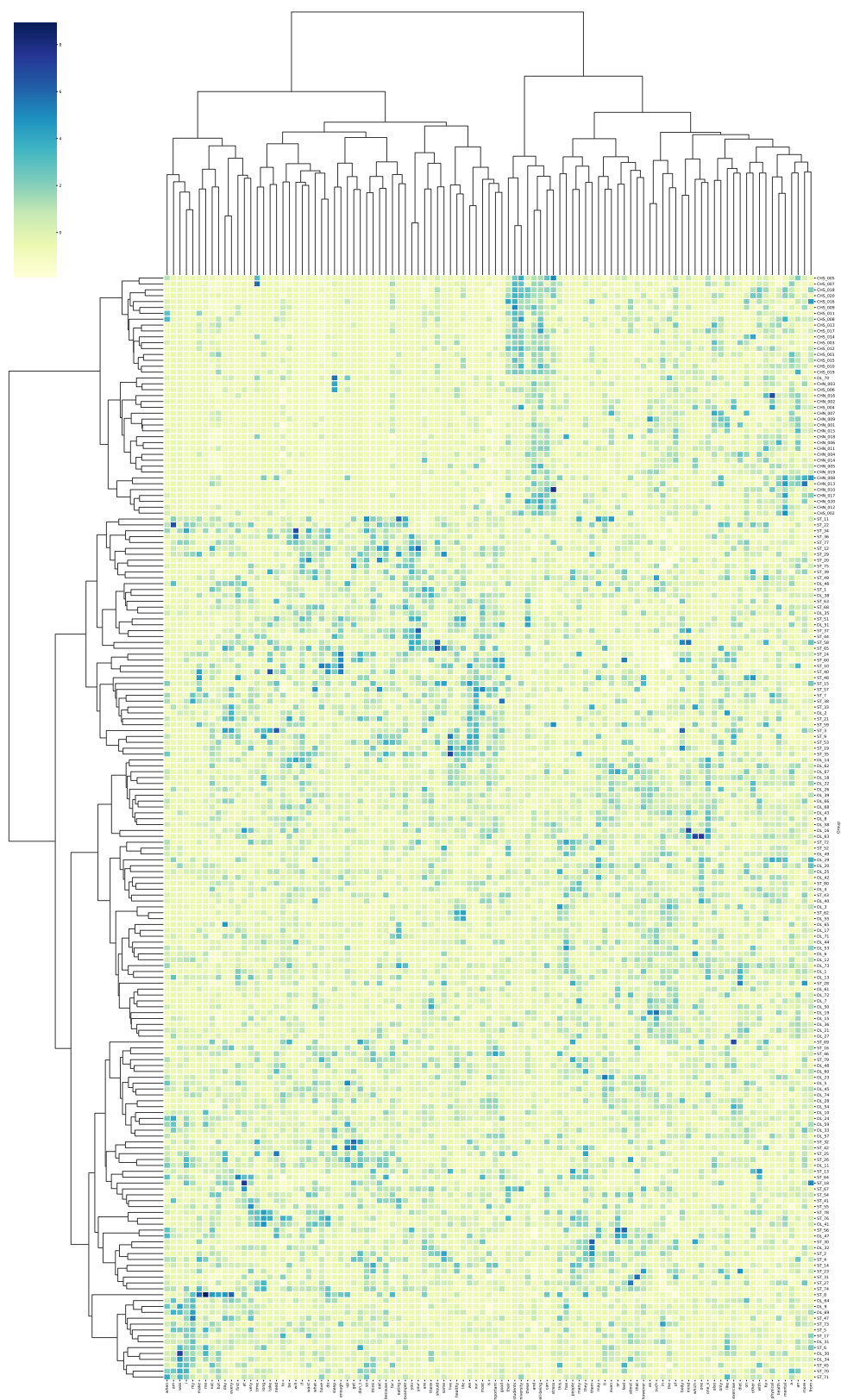


Fig. 4: 3 コーパス上位 100 語の HM 図 (Euclidian 距離, Ward 法)

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。

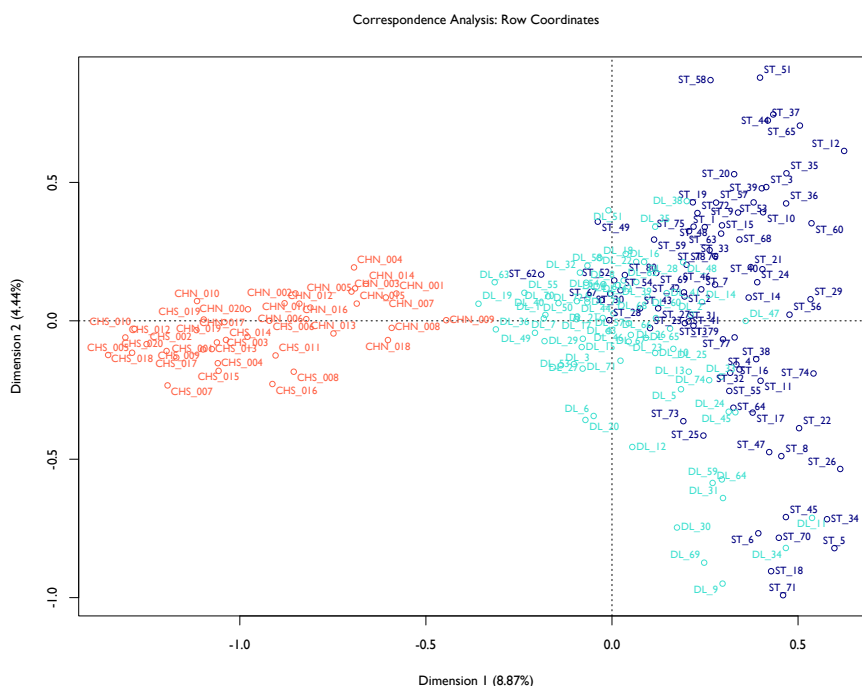


Fig. 5: 3 コーパス上位 500 語の CA (Row)

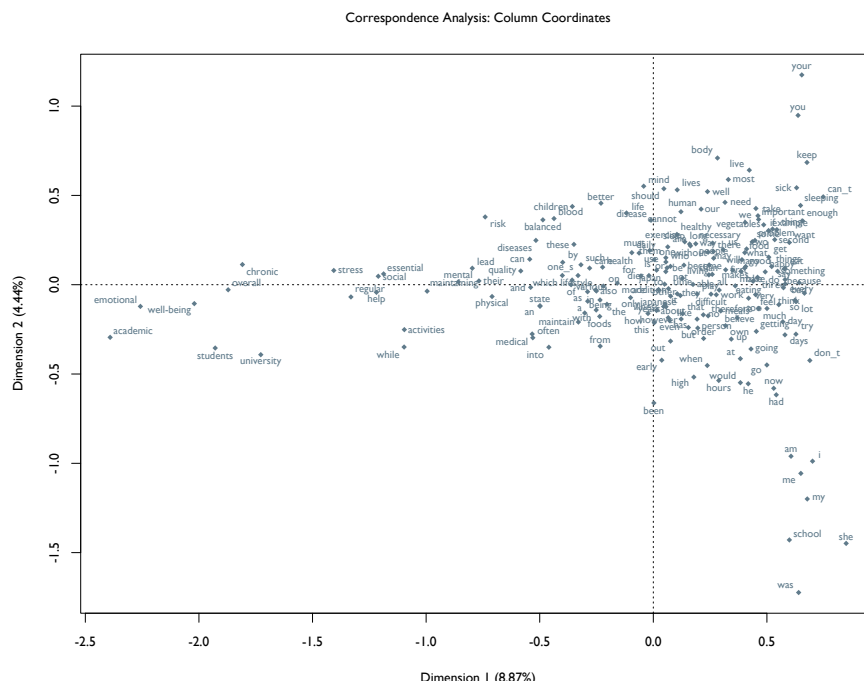


Fig. 6: 3 コーパス上位 500 語の CA (Column)

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ネ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。

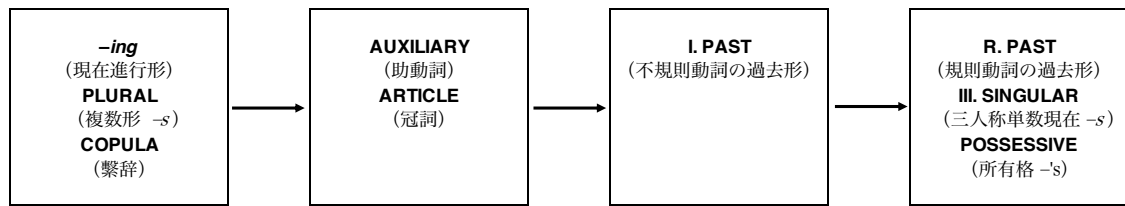


Fig. 7: 形態素の習得順序 (Krashen, 1997: 149, Fig. 1 改)

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」のではないかと、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、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 にコンコード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」が使用されているようにいまいと、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 affect 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ができる。

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 information the burden on your organs and allow you to maintain a healthier body and lifestyle important to have a hobby that allows you to release your mind This is because your immune system may be compromised and you may become ill Often the way we maintain stomach ache and your legs feel tired And you think to yourself I was leading a normal life does not mean that you slept well and you will not feel refreshed when you wake up after coronavirus infection two years ago As you all know we occasionally catch a cold because you are living a normal life just as you always have But suddenly you feel sick and do for the same reason as the first As you can see health and diet are strongly relevant had the experience I describe below You are living a normal life just as you are feeling any slight discomfort in your body you should first review these three points of diet and lifestyle Even with all this care you may still get sick However regular health is important to go through the effort of cooking you have the option of eating out If you choose after 8 hours of shallow sleep So how do you get deep sleep I have researched this so you should also eat a well-balanced meal Do you have many likes and dislikes If so it is your head feels hot your mouth gets dry you have a stomach ache and your legs feel heavy I didn't do anything unusual Eventually you will crawl into bed and rest not knowing melancholy and depression in everything you do In the worst case scenario it can even lead to a state in which you have everything you need including a perfectly balanced diet and is to exercise If you do not exercise you are likely to become obese The body becomes fat With a little effort to keep fit you will be able to lead a good life and as your mind sinks further and further you naturally lose your appetite and find it difficult is not how long you live but how happy you live And I believe that what is necessary is Second health cannot be perfect Have you ever had the experience I describe below that it is actually good to be healthy You don't realize how much you are appreciating other words if you want to stay healthy you must continue to engage in a multitude of 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 want not many people who will notice or help you so you have no choice but to self-report your hours 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 whatever you like without worrying about your blood sugar or cholesterol levels In addition They say you never think about your own health when you are healthy and It gets dry you have a stomach ache and your legs feel tired And you think to yourself are feeling depressed At this time as your mind sinks further and further you naturally sleep is blue light If you look at your smartphone or computer for a long time that thing is to be able to choose between your desires It is also important to learn how to change your lifestyle Being healthy enhances your quality of life As mentioned earlier the disadvantages Find a method that fits your lifestyle Being healthy enhances your quality of life Having breakfast is important for your daily rhythm You should also eat a well-balanced meal something that looks bad for your health is super appealing I am sure that there are no problems Therefore it is important to have your own stress-relieving methods In my case I do This is because no matter how healthy your body is if you have something on your mind you feel sick Your head feels hot your mouth gets dry you have a stomach ache you are going out and so on In other words if your body is not healthy you will not be able to eat much meals No matter how much you eat if your diet is unbalanced you are not maintaining your health you are feeling any slight discomfort in your body you should first review these three things let's say you have a very bad day in your daily life and are feeling depressed At this time your 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 believe that you should either about or simply lead an ordinary life your body and mind must be healthy However I think that If you have a good basic rhythm of life your risk of getting sick is smaller If you do not think further and further you naturally lose your appetite and find it difficult to get to the body and to the brain Along with moving your muscles exercise also protects your mind and your days without being conscious of your health is not a matter of course but a link clearly and improves the quality of your work and study These are just a few of the things that your body is if you have something on your mind it cannot be said to be healthy in the first days will help reduce the burden on your organs and allow you to maintain a healthy mood This is very important to protect your own mental health Going to school and working 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あ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、「現在の事実に対する反対の仮定」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あり、あり得ないことであり、しかし、"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、コンコード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)* 同様に一般の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、精神の健康と身体を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、テキ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ネ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る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ネンス分析を用い、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、分析する対象のテキ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.

執筆者紹介

田畑 智司（たばた ともじ）	大阪大学大学院人文学研究科教授
浅野 元子（あさの もとこ）	大阪医科薬科大学助教
藤田 郁（ふじた いく）	大阪大学大学院言語文化研究科 博士後期課程

言語文化共同研究プロジェクト 2023

テキストマイニングとデジタルヒューマニティーズ2023

2024 年 5 月 31 日 発行

編集発行者

大阪大学大学院人文学研究科言語文化学専攻