九州大学学術情報リポジトリ
Kyushu University Institutional Repository

同一学習者によるL2スピーキングとライティングの言語特徴の比較 ： ８指標に基づくコーパス分析

前田, 裕司
九州大学大学院地球社会統合科学府

https://doi.org/10.15017/7403516

出版情報：地球社会統合科学. 32 (2), pp.1-15, 2026-02-15. Graduate School of Integrated Sciences for Global Society, Kyushu University
バージョン：
権利関係：© 2026 MAEDA Yuji

KYUSHU UNIVERSITY

『地球社会統合科学』第32巻２号（2025）１～15
Bulletin of the Graduate School of Integrated Sciences for Global Society, Kyusyu University
Vol.32, No.2 (2025), pp.1～15

同一学習者によるL2スピーキングとライティングの言語特徴の比較
― ８指標に基づくコーパス分析 ―

Differences in Linguistic Features Between Japanese High School Students' Spoken and Written English: A Corpus-Based Analysis Using Eight Metrics

2025年10月7日提出，2025年11月11日受理

前田　裕司[1]
MAEDA Yuji[1]

キーワード：L2スピーキング，L2ライティング，学習者コーパス

要　旨

本研究は，日本の高校生同一学習者集団による英語スピーキングおよびライティングの産出の違いを明らかにすることを目的とし，８つのテキスト指標（ARI，VperSent，AvrDiff，BperA，CEFR，STTR，C/F，MSL）を用いた包括的な分析を行ったものである。分析には，31名の学習者から収集したスピーキングおよびライティングデータによる小規模学習者コーパスを使用した。統計分析の結果，８指標のうち６指標（ARI，AvrDiff，BperA，CEFR，STTR，C/F）においてスピーキングとライティングの間に有意差が認められた一方で，VperSentおよびMSLには有意差は見られなかった。さらに，相関分析の結果，いずれの指標においても強い正の相関は確認されず，弱い正または負の関係にとどまった。このことは，スピーキングとライティングのパフォーマンスが，これらの言語的側面において強く関連していないことを示唆している。総じて，学習者はスピーキングにおいてライティングよりも単純な語彙に依存し，語彙の多様性がより限定的である傾向が示された。

1．はじめに

日本人英語学習者のスピーキングおよびライティングという２技能の産出結果に関して，それぞれの技能に関連するコーパス研究は盛んに行われている。例えば，スピーキングに関する代表的なコーパスとして，The National Institute of Information and Communications Technology Japanese Learner English（NICT JLE）Corpus（和泉･内元･井原, 2004）が挙げられる。これは，日本人英語学習者約1,200人のインタビューテストの発話データをコーパス化したものである。また，Japanese EFL Learner（JEFLL）Corpus（投野, 2007）は，中学・高校の日本人英語学習者約10,000人の自由英作文をコーパス化したものである。これらの学習者コーパスは，学習者の言語産出を多角的に捉えるために有用であり（阿部・近藤, 2024），これらを用いて学習者の誤りから中間言語の特徴を明らかにするエラー分析（和泉・内元・井原, 2005）や共起しやすい語の組み合わせを分析するコロケーション研究（佐竹, 2015），スピーキングとライティングの産出結果の比較研究（Abe, 2007; Nomura, 2012）などが行われている。さらに近年では，学習者の発達過程を観察可能とする縦断的コーパスも構築されるようになった。阿部・近藤（2024）は，日本語を母語とする高校生104名を対象に，英語スピーキング能力の発達を23カ月にわたって追跡調査した縦断的コーパス，Longitudinal Corpus of Spoken English（LOCSE）を構築し，スピーキング能力の経年変化を分析している。

このように，多様な学習者コーパスが構築されてい

[1] 九州大学地球社会統合科学府
Graduate School of Integrated Sciences for Global Society

る一方で，多くのコーパス研究は，単一技能を対象としている（e.g., 阿部・近藤, 2024; 和泉他, 2005; Kobayashi, Abe, & Kondo, 2022; 成田, 2024）。２技能を対象とした研究には，産出モードの違いによる言語的特徴を n-gram（単語および品詞連鎖）の観点から分析した研究（野村, 2009）や，品詞使用の傾向に着目した研究（Nomura, 2012），Ｌ２学習者の語彙が話し言葉と書き言葉でどのように発達しているかを分析した研究（Ishikawa, 2015）が存在するものの，単一技能コーパスを用いた研究と比較するとその数は非常に少ない。

その要因の一つとしては，同一学習者によるスピーキングとライティングの２技能の産出データをコーパス化したものは非常に限られていることが挙げられる。先行研究の中には，研究者が独自に収集･構築した２技能コーパスを用いた事例も少数ながら報告されている（e.g., 野村, 2009; Nomura, 2012; Vibar, 2022）。しかし，同一の日本人英語学習者によるスピーキングとライティングの産出データを公開しているコーパスは，アジア圏学生による英語産出を集めた International Corpus Network of Asian Learners of English（ICNALE; Ishikawa, 2023）を除いて，確認した限りでは存在しない。

昨今の日本の英語教育においては，より実践的な英語運用能力の育成を目指し，「話すこと（やり取り）」「話すこと（発表）」「書くこと」の２技能３領域について，パフォーマンステストを通じた評価の導入が推奨されている（国立教育政策研究所教育課程研究センター , 2021）。このような教育的背景を踏まえると，スピーキングとライティングという異なるモードの産出にどのような差異があるのかを明らかにすることは，英語教師が中間言語変異（Selinker, 1972）を理解し，適切な指導方針を立てる上で有益な知見を提供すると考えられる。

そこで本研究では，日本の高校１年生31名からスピーキングおよびライティングの産出データを収集し，独自にコーパス化を行った上で，同一学習者集団による2技能間の産出結果を比較・分析し，その言語的特徴の違いや関係性を明らかにすることを目的とする。

2.　先行研究

スピーキングとライティングの違いは，言語処理上の時間的制約や計画性の差に加え，モードやレジスターの違いとも密接に関連している。例えばHalliday（1989, 2004）は，話し言葉の複雑性がアカデミックライティングの複雑性とは劇的に異なると主張している。具体的には，話し言葉における主要な文法的複雑性は従属節に関わるものであり，一方で書き言葉は名詞や名詞化に依存していると報告している。またBiber（1988）は，主語と時制を持つ動詞が含まれる定形従属節（that 節, WH 節，因果関係を表す副詞節，条件を表す副詞節など）が対人的な話し言葉レジスターの特徴であることを明らかにしている。これに対して，名詞を修飾する句レベルの特徴（例：限定用法の形容詞や前置詞句）は，形式的な書き言葉レジスターに典型的であると述べている。さらにBiber, Gray, & Poonpon（2011）は，Ｌ２英語の文法的複雑性を評価する際に，spoken register の特徴をそのまま指標として用いることの妥当性を再検討し，レジスター差が指標の解釈に大きく影響することを指摘している。したがって，スピーキングとライティングといった異なるモードを比較する際には，それぞれのモードに適した複雑性指標を設定する必要があると述べている。このようなモード差は，中間言語の可変性（Selinker, 1972）を理解する上でも重要な要素である。

野村（2009）は，中高生それぞれ20名，計40名から，同一テーマによるスピーキングデータとライティングデータを収集，コーパス化し，単語および品詞n-gram（n個の連続した単語または文字から成る語列の単位）から分析を行った。その結果，産出モードの違いによる言語産出の差異が明らかになった。例えば，スピーキングとライティングにおける頻度３以上のtrigramパターン数はスピーキングが74に対し，ライティングでは46と，スピーキングの方が同じtrigramを使用している割合が高いことが示された。また，品詞n-gramの分析では，スピーキングでは代名詞を主語とした短い構造が頻繁に使用されており，簡潔な言語処理が行われていることが示唆された。一方，ライティングでは，産出までに時間的な余裕があることで，より多様な名詞や修飾語を用いた表現が選択される傾向が見られた。これらの結果から，スピーキングとライティングの違いは，言語処理上の時間的制約と密接に関連していると考察している。

さらにNomura（2012）は，日本の中学３年生から高校３年生，計324名から，同一テーマに基づく，スピーキングとライティングデータを収集し，冠詞使用の変異について分析した。調査結果によると，産出モードの違いは冠詞使用の変異に影響を与えないという結果が得られており，その理由として，日本人初学者のライティングがスピーキングに近い特徴を持っており，いずれのモードにおいても冠詞に注意を払っていない可能性を指摘している。

またVibar（2022）は，フィリピンの女子高校生1名のスピーキングとライティングの語彙密度や文法的特徴の違いを分析することを目的として，事例研究を行っている。共通テーマについてエッセイライティングとインタ

ビューを実施し，その産出結果を比較したところ，スピーキングの方が産出語数が多かったこと，また機能語の使用割合が多かったことなどが明らかになった。ライティングの方では，従属節はスピーキングよりも少なかったのに対し，独立節の数はスピーキングよりも多かったことが観察されている。また語彙密度（lexical density）に関しては，語彙的内容語の割合から情報の密度を示す指標であるが，スピーキングとライティングではほぼ同じ数値となり，モードによる差は見られないという結果が明らかになった。

さらにAbe（2007）は，中学１年生から高校３年生までの日本人英語学習者によるスピーキングとライティングにおけるエラーの差を分析した。その結果，発達段階が進むに応じて，両方の産出モードにおいて動詞に関連する前置詞の正確性が向上することが明らかになった。また，代名詞，形容詞，副詞，名詞，動詞の正確性は，スピーキングでは向上する一方，ライティングではほぼ変わらない，もしくは不規則な変化を示した。

Ishikawa（2015）は，ICNALE を用いて，スピーキング（60秒の即興スピーチ）とライティング（200～300語のエッセイ）の２モードにおける語彙使用を比較分析した。分析指標として，語彙多様性（Lexical Diversity），語彙密度（Lexical Density），語彙複雑性（Lexical Complexity），高頻度語の使用割合を示す語彙基本度（Lexical Fundamentality），およびテキストの名詞の使用割合を示す名詞志向度（Noun Orientation）の５項目を設定し，各指標の変化をCEFR準拠の習熟度別に検証している。その結果，すべての指標においてモード間で明確な違いが確認された。具体的には，ライティングではスピーキングよりも語彙多様性・密度・複雑性が高く，より情報指向的で形式的な語彙使用が見られたのに対し，スピーキングでは高頻度語や日常語の使用率が高く，基本語への依存傾向が示された。また，主成分分析の結果から，学習者の語彙使用を最も強く規定する要因は「英語習熟度」ではなく「産出モード（スピーキング／ライティング）」であることが明らかにされている。

以上のように，Ｌ２英語学習者のスピーキングおよびライティングの産出を比較した研究は一定数存在するが，単一技能コーパスを対象とした研究と比べると依然として数は限られている。 特に，同一学習者集団による２技能の比較研究は非常に少なく，その多くは個別の文法項目や語彙的特徴に焦点を当てたものであり，両モードを包括的かつ定量的に比較した研究はほとんど見られない。さらに，先行研究ではレジスターやモードによる言語使用の差異を指摘するものの，スピーキングとライティングにおける複数の言語的指標を同一の学習者データに基づいて統合的に検討した研究は限られている。

したがって，本研究は，日本人高校生31名によるスピーキングおよびライティングの産出データを対象に，モード差に基づく言語使用の特徴を統計的に検証し，学習者の英語産出におけるスピーキングとライティングの関係をより多面的に捉える試みである。

３．リサーチデザインと手法

３．１　研究目的とRQ

本研究の目的は，日本の高校生を対象に，同一学習者集団から得られた英語スピーキング（発表）およびライティングという２技能の産出結果を比較し，その言語的特徴の違いや関係性を明らかにすることである。ここでスピーキング（発表）は，モノローグ型スピーチタスクを指し，学習者が与えられたテーマについて約１分間で意見を述べる一方向的な発話を対象とする。本研究では，以下のリサーチクエスチョン（RQ）を設定した。

RQ 1: 同一学習者による英語スピーキングとライティングの間には，言語的特徴に統計的に有意な差がみられるか。

RQ 2: スピーキングとライティングにおける各言語的特徴の間には，どの程度の相関関係がみられるか。

３．２　協力者

本研究では，日本の私立高校の英語系の学科に通う高校１年生52名に研究協力を依頼し，そのうちスピーキングおよびライティングの両方の産出データが取得できた31名のデータを最終的な研究対象とした。研究に参加する生徒には，研究における提供データの範囲，個人が特定されないよう匿名化した上での利用などについて文書で説明を行い，同意書を取得した。今回研究協力を依頼した学科には，英語が得意または好きだという生徒が多く，英語力を強みとして主に私立大学へ進学する傾向がある。

３．３　取得データのタスクと手順

スピーキングデータに関しては，１つのスピーチテーマを設定し，そのテーマに対するスピーキングデータを収集した。データ収集時の条件や方法について，スピーキングタスクは，授業外に研究協力者を対象として教室内で実施されたものであり，発話データは個別に録音・提出された。教室は100名程度収容可能な通常の教室よりも広い教室で，その場に研究協力者54名を集め，一斉にデータ収集を実施した。生徒にスピーチテーマを提示し，テーマについて話す内容を考えるための１分間の準

表1　取得データに関する概要

データの種類	テーマ（問い）	データ収集時期
スピーキング	Choose one of the prefectures in Japan you want to visit and make a 1-minute speech. Please include why you want to visit the prefecture.	2024 年 1 月
ライティング 1	次の問いについて，あなたの意見とその理由を 2 つ挙げ，30 語～50 語程度の英文で書きなさい。 What is the best season to visit your hometown?	2023 年 12 月～ 2024 年 1 月
ライティング 2	次の問いについて，あなたの意見とその理由を 2 つ挙げ，30 語～50 語程度の英文で書きなさい。 What is your favorite school subject and why?	2023 年 12 月～ 2024 年 1 月
ライティング 3	次の問いについて，あなたの意見とその理由を 2 つ挙げ，30 語～50 語程度の英文で書きなさい。 Do you prefer to eat Japanese food or Western food?	2023 年 12 月～ 2024 年 1 月

備時間を設けた。その間，生徒がメモを取ることは許可した。準備時間の 1 分が経過した後，筆者の合図で一斉に発話を開始させた。録音には各生徒のスマートフォンに搭載されたボイスレコーディング機能を使用し，各自が端末に向かって発話を録音した。1 分間の発話後，筆者の合図で録音を終了し，録音データに名前をつけてオンライン経由（AirdropまたはGoogle Drive）で筆者に提出をさせた。

比較するライティングデータについては，協力校が利用する「スマートレクチャーコレクション」というオンライン英語添削サービスを通じて取得した （https://slc.tbshare.net/info/index.html）。このサービスは株式会社新興出版社啓林館が学校向けに提供しているもので，生徒は専用のウェブサイトで，与えられた問いに対する英作文を提出することで，海外に住む外国人講師から添削とコメントを付与されるサービスである。ライティングデータの取得において，スピーキングデータの 1 人あたりの平均語数が45.1語であったことから，総語数に大きな差が出ないよう，1 つのテーマに対して30語～50語程度で書くタスクから取得することとした。また，今回取得できたスピーキングデータは，協力校の授業時数やスピーキングデータ取得にかかる生徒への負担等から 1 つのタスク分のみであったが，ライティングデータに関しては，対象の学習者集団の一般的なライティング産出結果として扱うため複数のライティングテーマから収集することとし，結果として 3 つの英作文のライティングデータを収集した。収集したスピーキングデータとライティングデータの収集（産出）時期やテーマについては表 1 にまとめて示している。

協力校の英語科教諭に研究対象者のライティング産出時の実施条件や実施環境等についてヒアリングを行ったところ，以下のことが明らかになった。まず，今回収集したライティングデータに関しては，生徒は冬季休暇の課題として取り組んでいた。実際に英作文を入力する際には，ネット辞書やネット情報の使用は許可されていたが，日本語の文章をそのまま翻訳することは禁止されていた。また生徒は付属の教材に沿って英作文を作成した。教材には，トピックに関連する語彙や表現を学ぶための演習問題が付属していたが，モデル文は提示されておらず，学習者はこれらの語彙を参考にしながら自由に英作文を作成した。生徒によって英語力に差があるとの理由から，文章作成の時間制限は設けていない。提出する際にはシステム上で語数制限が設定されており，今回ライティングデータを収集した問いでは，提出に必要な最低語数は30語，最大語数は60語となっている。研究対象となった学習者は同一校内の複数クラスに所属していたが，すべて同一条件下でライティングタスクを実施した。

3．4　分析

3．4．1　データの事前処理

スピーキングの産出データに関しては，まず学習者が自身の録音データを基に書き起こしを行った。その後，筆者が録音データを聞きながら書き起こしデータを確認し，必要に応じて加筆修正を行った。また，「un」や「hmm」

表2 作成したコーパスの概要

	N	総語数	平均語数	最大語数	最小語数
スピーキング	31 (31)	1,398 (1269)	45.1 (40.1)	129 (128)	14 (14)
ライティング(平均)	31	1,236	39.9	57.7	30
ライティング 1	31	1,314	42.4	56	30
ライティング 2	31	1,068	34.5	59	30
ライティング 3	31	1,325	42.7	58	30

注：スピーキングについてカッコ内の数字は pruned token の数字を示す。

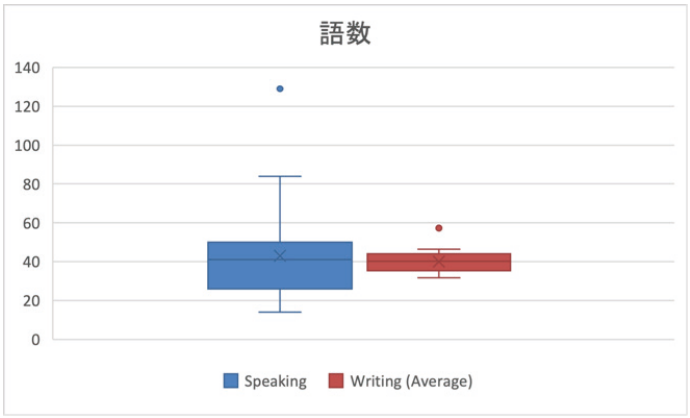

図1　スピーキングとライティング（平均）の語数のボックスプロット

といった単独では意味を持たないフィラーを除去したが，繰り返しや自己訂正についてはそのままとしたデータと繰り返しや自己訂正も取り除いたpruned token（三谷, 2019）の2種類のデータを作成した。Pruned tokenについては後節で詳述するSTTR，C/F，MSLでの分析にのみ用いている。これらの指標は，発話の表層的特徴よりも文法的・語彙的構造を精緻に測定することを目的としており，非流暢性要素を含めると指標値が不当に低下または変動する可能性があるためである。一方，その他の指標では，学習者の自然な発話特性を反映させることを重視し，非処理データを分析対象とした。文の区切りについては，学習者による書き起こしを基に筆者が音声データを確認し，ポーズの長さや文構造を参考にしてピリオドの位置を設定した。

ライティングの産出データに関しては，各指標の数値算出に影響が出ることを防ぐため，スペルミスのみ，Googleスプレッドシートのスペルチェック機能を使用して訂正を行った。訂正が適切かどうかの判断については，筆者が提示されたスペルミスを全て確認した上で行った。一方で，文法的なエラーや句読点については学習者が入力したデータそのままとし，一切修正を加えていない。

3.4.2　作成したコーパスの概要

取得したスピーキングデータおよびライティングデータは，エクセルファイルに格納し，本研究のための小規模の学習者コーパスを構築した。コーパスの概要は表2の通りである。総語数，平均語数，最大語数，最小語数の算出には，Python（version 3.10.12）のpandasライブラリ（version 2.2.2）を使用した。また図1はスピーキングとライティング（平均）の語数をボックスプロット化したものである。

3.4.3　分析ツール・指標

本分析には，Uchida and Negishi（2018）によって公開されたCEFR-based Vocabulary Level Analyzer（CVLA version 2.0）というウェブアプリケーションを使用した（https://cvla.langedu.jp/）。CVLAは入力されたテキストに対して，4つの言語的特徴に基づき，CEFRを日本の英語教育の実態に即しより細分化した（投野・根岸, 2020）CEFR-Jレベルを付与するシステムである

表3　CVLAで得られる指標の説明

指標	指標の説明
ARI	英語テキストの読みやすさを数値化したもの
VperSent	一文あたりの動詞の平均数
AvrDiff	語彙難易度の平均値（A1，A2，B1，B2 はそれぞれ 1，2，3，4 としてカウントされる。内容語のみ，機能語は除外）
BperA	A レベルの内容語に対する B レベルの内容語の割合
CEFR	上記 4 つの指標をもとに算出されるテキスト全体のレベルを表す数値

（Uchida & Negishi, 2018）。本分析ではARI（Automated Readability Index），VperSent（Verbs per Sentence），AvrDiff（Average words difficulty），そしてBperA（B level content words per A level content words）という４指標が算出される。これらは，文レベルの特徴（ARI・VperSent）および語彙レベルの特徴（AvrDiff・BperA）を捉えるものであり，テキストの総合的な難易度を推定するための基礎データとして活用される。さらに，CVLAではこれらの指標に基づいてCEFR（CEFR-J）レベルを自動算出するが，本研究ではウェブアプリケーション上に表示されないCEFR値について，Uchida and Negishi（2018）に提示された算出式を用いて算出した。各指標の簡易的な説明は表３の通りである。これらの指標において，同一学習者集団によるスピーキングとライティングの産出結果の違いを分析した。なおCVLAは本来，リーディングとリスニングというインプットデータのためのテキストレベル判定システムである。本研究の目的は，スピーキングとライティングという異なるモード間の語彙的・統語的特徴の差異を定量的に明らかにすることであり，CVLAが提供する指標群は，その目的に適合すると判断した。

CVLAで取得できる上記の５指標に加え，スピーキングおよびライティングの産出結果の差をより多角的に比較するため，Type-Token Ratio（TTR）を50語で調整したStandardized Type-Token Ratio（STTR；水本, 2008），機能語に対する内容語の割合を示すContent words per Function words ratio（C/F），そして一文あたりの平均単語数を示すMean Sentence Length（MSL）についても算出し，分析を行った。TTRはテキストが短いほど数値が高くなるという問題点が指摘されている（Baayen, 2001）。本研究で取得したスピーキングデータでは発話語数の少ない学習者が多かったこと，学習者間で語数にばらつきが見られたことから，TTRではなくSTTRを分析指標として採用した。

ライティングについては，英作文のテーマによる影響を減らし，本研究で対象となった学習者集団の一般的なライティング産出結果としての数値として扱うため，各指標の平均値を統計検定に使用している。

３．４．４　統計分析

本分析では，まず学習者から取得したスピーキングおよびライティングの産出データをエクセルファイルに集約し，小規模の学習者コーパス（スピーキング1,398語，ライティング3,707語，総語数5,105語）を構築した。その後，分析対象となる指標の数値データを取得するためのPythonコードと，その数値データに対して統計的分析を行うためのPythonコードを作成した。これらのコードを使用して，学習者コーパスに格納されたスピーキングおよびライティングの産出データを自動的に分析し，得られた数値データに対して統計的な分析を実施した。

３．５　データの妥当性

今回取得したスピーキングとライティングの産出データは，それぞれのレジスター的特徴が反映されたデータであり，研究データとして一定の妥当性を有すると考える。レジスターとは一般的に「言語使用域」と訳され，使用場面や状況によって使い分けられる言語変種のことである。今回取得したスピーキングデータは，日常的な場面に近い即時性や即興性が求められる制約のもとで取得されたものであり，一方ライティングデータについても辞書等の使用が許可されていたが，外部ソースの参照が可能であった点は，現実的な学習場面を反映している。

スピーキングとライティングを比較する際のタスク統制については，先行研究においても課題とされている。投野（2008）は，同一学習者によるスピーキングとライティングを比較する場合，同一タスクの言語使用データを取得する必要があると指摘しているが，これまでの比較研究においても，タスクや条件の違いが結果に影響を与える点は共通の課題である（Abe, 2007; Lintunen & Mäkilä, 2014）。また，同一学習者による同一タスクの取

得においても，どちらの技能を先に実施するかが結果に影響を与える可能性がある。例えば，スピーキングとライティングのどちらを先に取得するかが，結果に影響を与える。先にライティングデータを取得し，その後同一テーマでスピーキングデータを取得しようとすれば，学習者はライティング産出時の記憶を頼ってスピーキングを行うであろう。仮にライティングデータの取得から一定の期間をおいて，スピーキングデータを取得したとしても，その間の英語学習の影響が全くないとは言えない。同一学習者によるスピーキングとライティングを，完全に統制された同一条件で取得することは現実的に困難である。

この課題を踏まえ本研究では，統制されたテスト環境ではなく，学習者が日常的に使用する教室や自宅などの自然な状況下で産出したデータを用いた。本研究で用いるデータは，学習者の実際の言語使用を反映しており，より実践的な視点からスピーキングとライティングの違いを分析することが可能である。このようなデータを分析することは，実際の学習環境におけるスピーキングとライティングの使用状況をより正確に捉えることができるため，教育的な応用可能性が高いと考えられる。

もっとも，本研究のデータにも，タスク条件の統制やモード間影響といった課題は残る。今後は，同一タスク条件下でのデータ収集や，モード間の相互影響を最小限に抑える方法を検討する必要がある。しかしながら，本研究で収集したデータからも，スピーキングとライティングの異なる特徴を明らかにできることから，比較研究として一定の意義を有すると言える。

4．結果と考察

4．1　言語的指標の統計的分析（RQ1）

4．1．1　CVLAで得られる言語的指標

CVLAで得られた指標ごとのスピーキングおよびライティングの平均値に有意差があるかを検証するため，統計的な検定を行った。まず，各指標について得られた数値が正規分布しているかを確認するため，シャピロ・ウィルク検定を実施した。正規分布に従っている可能性が高いと判断された場合には対応のあるt検定を実施し，いずれかが正規分布に従っていない可能性が高いと判断された場合にはウィルコクソン符号付順位検定を適用した。これらの統計的検定にはPythonのSciPyライブラリ（version 1.13.1）のstatsモジュールを使用した。表4には，CVLAで得られた指標ごとのスピーキングとライティングの平均値および有意差検定の結果をまとめている。また図２はCVLAで得られた５つの指標の数値をボックスプロット化したものである。

検定の結果，VperSentを除くすべての指標において，５％水準でスピーキングとライティングの間に統計的有意差が確認された。

まず，ARIでは５％水準で統計的有意差が確認され，効果量に関しても）d = -1.31という非常に大きな効果量が確認されたことから，スピーキングとライティングの間には実質的にも大きな差があることがわかった。ARIは文長や単語の長さに影響を受けやすい指標である。収集したライティングデータでは，提出時に語数の下限制限が課されていたことから，一定量のテキストが産出されていた。一方で，スピーキングデータでは，発話語数が極端に少ない学習者が多かったこと，また繰り返しや自己訂正をそのまま保持していたこと等が結果に影響を与えたと考えられる。

次に，テキストで使用された語彙の難易度をCEFR-J Wordlistに基づいて算出するAvrDiffについては，スピーキングとライティングの間で５％水準で統計的に有意差が確認された。同様に語彙のレベルを表す指標であるBperAでも５％水準で有意差が認められ，効果量はそれ

表４　スピーキングとライティングにおける指標ごとの平均値と有意差検定結果（N=31）

	スピーキング M （SD）	ライティング M （SD）	有意差検定結果	効果量
ARI	0.06 （1.99）	3.73 （1.42）	$t(30) = -7.30, p < .05$	$d = -1.31$
VperSent	1.87 （0.75）	1.67 （0.34）	$W = 165.5, p = .11$	$r = 0.29$
AvrDiff	1.17 （0.10）	1.45 （0.12）	$t(30) = -10.39, p < .05$	$d = -1.87$
BperA	0.04 （0.04）	0.18 （0.06）	$W = 2.0, p < .05$	$r = -0.86$
CEFR	0.31 （0.62）	1.73 （0.56）	$t(30) = -8.49, p < .05$	$d = -1.52$

注：tはt検定，Wはウィルコクソン符号付順位検定の検定統計量を示す。tについてカッコ内の数字は自由度を示す。

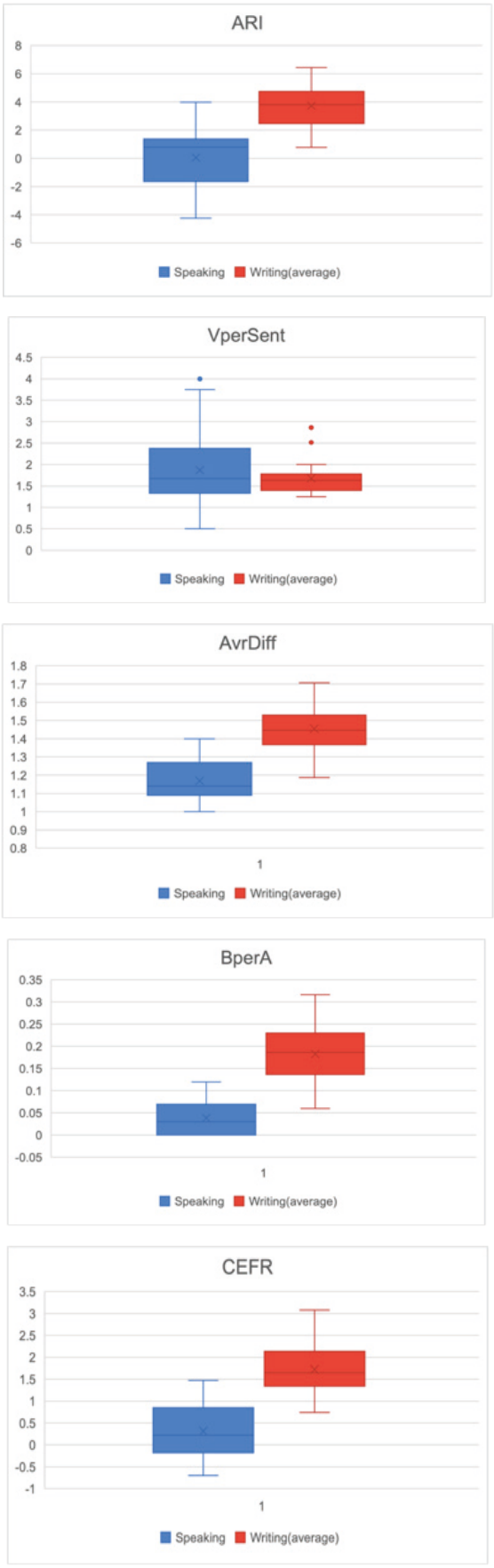

図2　スピーキングとライティングの指標ごとの分布

ぞれ，d = -1.87，r = -0.86という非常に大きな効果量が確認され，スピーキングの方がライティングよりも産出語彙のレベルが低かったことが示された。これらの結果には，ライティングの言語産出時に辞書の利用が許可され時間制限がなかったことが影響した可能性がある。また，スピーキングの言語産出時には，準備時間と発話時間がそれぞれ１分間に制限されていたため，使用語彙が限定的になった可能性が考えられる。

一方，一文あたりの動詞の平均を示すVperSentについては，統計的有意差が認められなかった。ライティングの方が接続詞や関係詞を用いることで一文あたりの動詞数が多くなると予想していたが，結果は異なった。この結果には，両タスクのテーマ設定が影響した可能性がある。いずれのタスクでも理由を含めて説明することが求められており，多くの学習者データでbecauseを用いた複文が産出された。こうした構文的特徴により，スピーキング・ライティング双方における文あたりの動詞数が増加し，両技能間で統計的な差が見られなかったものと考えられる。さらに，スピーキング音声データの書き起こしの際に，一文の区切り方に筆者の主観的判断が影響した可能性も否定できない。

CEFRは言語的特徴を示す４つの指標（ARI, VperSent，AvrDiff，BperA）に基づいて算出される総合的なテキストレベル指標である。本研究においては，スピーキングとライティングの間でこのCEFR指標に５%水準で統計的有意差が認められた。語彙難易度や語彙レベルがCEFR指標の構成要素であることを踏まえると，スピーキングにおける語彙使用の特徴が，CEFRスコアの低さに寄与した可能性を示唆している。

４．１．２　STTR（標準異なり語比率）とC/F（内容語・機能語比率）

次にスピーキングとライティングのSTTR，C/Fを算出した。その結果を表５に示す。STTR，内容語，機能語のカウント，C/Fの算出には，Pythonのpandasライブラリ，NumPyライブラリ（version 1.26.4），NLTKライブラリ（version 3.8.1）を使用した。

STTRについて，スピーキングでは平均0.55（0.61）であったのに対し，ライティングでは平均0.70であった。またC/Fについては，スピーキングが平均1.13（1.19）であったのに対し，ライティングでは平均1.43という数値が得られた。

これらの結果を基に，スピーキングとライティング間のSTTRおよびC/Fに有意差があるか確認するため，統計的な検定を実施した。まず指標ごとに得られた数値が

表5　タスクごとのSTTR，C/Fに関する基本統計量（N=31）

	STTR	平均内容語数	平均機能語数	C/F
スピーキング	0.55 (0.61)	23.6 (21.4)	22.4 (19.5)	1.13 (1.19)
ライティング	0.70	23.3	16.9	1.43

注：カッコ内の数字は pruned token の数値を示す。

表6　STTR，C/Fの数値と検定結果

	スピーキング	ライティング	対応のある t 検定	効果量
STTR	0.55	0.70	$t(30) = -9.81,\ p < .05$	$d = -1.95$
	(0.61)		$(t(30) = -5.48,\ p < .05)$	$(d = -1.05)$
C/F	1.13	1.43	$t(30) = -4.84,\ p < .05$	$d = -0.87$
	(1.19)		$(t(30) = -3.84,\ p < .05)$	$(d = -0.69)$

注：カッコ内は pruned token の数値を表す。t についてカッコ内の数字は自由度を示す。

正規分布しているか確認するため，シャピロ・ウィルク検定を行った。その結果，どちらも正規分布に従っている可能性が高いと判断されたため，対応のあるt検定を実施した。これらの統計的検定にはPythonのSciPyライブラリのstatsモジュールを使用した。検定結果を表６に示す。

STTRの分析においては，スピーキングデータのうち，繰り返しや自己訂正を含むデータ，およびそれらを除去したpruned tokenデータの双方において，ライティングとの間で５%水準の有意差がt検定により確認された。この結果から，ライティングよりもスピーキングの方が同じ語を繰り返し使用する傾向がある，または語彙の使用が限定的であるということが考察される。Ishikawa (2015)でも同様の傾向が見られている。さらに，ライティングでは産出までに時間的な余裕があることで，より多様な名詞や修飾語を用いた表現を使う傾向が見られた，野村（2009）の研究結果とも共通点が認められる。C/Fに関しても同様に，５%水準で統計的有意差が認められた。スピーキングの方がライティングよりも機能語の割合が高いという結果は，Ishikawa（2015），Vibar（2022）の研究結果と一致している。Vibarの研究では１人の被験者を対象とした事例研究であり，結果の一般化に課題があった。しかし本研究は，31名の学習者集団を対象としており，Vibarの結果を裏付けるとともに，その結果をより一般化する上で意義のあるものであると考えられる。ただし，Vibarの研究では，本研究では除去した「uh」などの単独で意味を持たないフィラーが分析データに含まれ，それが1語としてカウントされていたため，分析条件には若干の違いがある。より正確に比較するためには完全に同じ条件でのデータ分析が必要であるが，概ね同じ結果を得られたと評価できる。

4．1．3　MSL（一文あたりの平均単語数）

次に，MSLに関して，スピーキングとライティングの間で平均値に差があるかを確認するため，統計的検定を実施した。データの正規性を確認したところ，スピーキングデータの方で正規性が認められなかったため，ウィルコクソン符号付順位検定を実施した。これらの統計的検定にはPythonのSciPyライブラリのstatsモジュールを使用した。表７はその結果をまとめたものである。なおMSL文長における統計的検定には，繰り返しや自己訂正による影響を排除するため，pruned tokenを使用した。

本分析の結果，MSLには，スピーキングとライティングの間で統計的有意差は確認されなかった。前節で統計的有意差が認められなかったCVLAのVperSentは一文あたりの動詞の平均割合を示す指標であり，MSLと類似し

表7　MSLのスピーキング，ライティングの平均値と有意差検定結果

	スピーキング	ライティング	ウィルコクソン符号付順位検定	効果量
MSL	9.89	10.26	$Z = 200.00,\ p = .36$	$r = -0.17$

た指標であると考えられる。具体的には，長い文では動詞の数が増える傾向があるため，両指標が相関する可能性はある。このことから，VperSent と MSL のいずれにも有意な差が見られなかったという結果は，スピーキングとライティングの間で，文構造の基本的なパターンには大きな違いがない可能性を示唆していると考えられる。

4.2　相関分析（RQ2）

本研究で分析対象とした各指標について，スピーキングとライティングの間にどのような相関があるか確認するため，ピアソン相関係数を算出した。表８にその結果を示す。なお，ピアソン相関係数の算出にはPythonのpandasライブラリを使用した。

相関分析の結果，同一学習者集団によるスピーキングとライティングの産出結果には，すべての指標で弱い正の相関または負の相関しか見られなかった。このことから，本研究で対象とした同一学習者集団において，スピーキングとライティングの産出結果は，本分析で用いた各指標において相互に強い相関を示さなかったことが示唆される。

これらの結果は，スピーキングにおけるリアルタイム処理負荷やタスクモードの特性を反映している可能性がある。スピーキングでは，限られた準備時間の中で意味を途切れさせずに伝える必要があるため，汎用的で易しいレベルの語彙を選択しやすい。一方，ライティングでは時間をかけて語彙を選び，より難易度の高い語を使用できるため，両技能間で語彙レベルに逆の傾向が生じたと考えられる。また，スピーキング課題は日常的なトピックであったことから，容易な語彙でも意味が十分伝達できたことも影響している可能性がある。

また図３は，前節までの分析で統計的有意差が確認された指標について，スピーキングとライティングの数値を散布図として視覚化したものである。

表８　各指標におけるスピーキングとライティング（平均）の相関係数

	ピアソン相関係数
ARI	-0.34
VperSent	0.18
AvrDiff	0.03
BperA	0.03
CEFR	-0.24
STTR	0.22 (0.11)
C/F	0.21 (0.16)
MSL	(0.20)

注：カッコ内の数字はpruned tokenの数値を示す。

使用語彙難度に関する指標であるAvrDiffとBperA において特に相関係数が小さかった。AvrDiffについて詳細に確認したところ，ライティングで最も高い数値を示した学習者のスピーキングでは，数値が1.0と最も低い数値となっている。AvrDiffにおいて1.0という数値は，産出された単語の内容語のすべてがCEFR-J WordlistのA１レベルの単語であったことを示している。なお，スピーキング，ライティングいずれのAvrDiffも2.0には達しておらず，両技能間の数値差は１未満であるものの，統計的に有意であったことから，限られた語彙範囲の中でも使用語の分布に技能特性が反映されていると考えられる。

同じく使用語彙レベルに関する指標であるBperAでは，スピーキングにおいて０という数値が複数の学習者で確認された。これは，産出された発話に，CEFR-J WordlistのBレベルの単語がまったく使用されていないことを意味する。発話テーマに影響を受ける可能性はあるものの，準備時間が短く即興に近い条件でスピーキングを実施すると，レベルの高い単語は使用されず，簡単な単語を用いて意見を伝える傾向がある，あるいは難しい語彙を使用することができないということが示唆される。また，話し言葉と書き言葉のレジスター差もこの結果に影響した可能性がある。スピーキングでは，より頻度の高い口語的語彙や汎用的表現が使用される傾向があり，ライティングに比べてレベルの高い語彙が使用されにくいことが先行研究でも指摘されている（e.g. Ishikawa, 2015; Kim & Crossley, 2023）。

負の相関が見られた指標の一つであるARI（Automated Readability Index）は，文の長さや語の平均文字数に基づき文章の難易度を推定するものである。スピーキングの書き起こし文では短い文や文断片が多く、ライティングよりも数値が低くなる傾向がある。一方で，ライティングでは接続詞や従属節を多用した複文が多く，ARIが高く算出されやすい。したがって，ライティングで相対的に高いARIを示す学習者がスピーキングで低い値を示すという傾向が生じ，負の相関として表れた可能性がある。

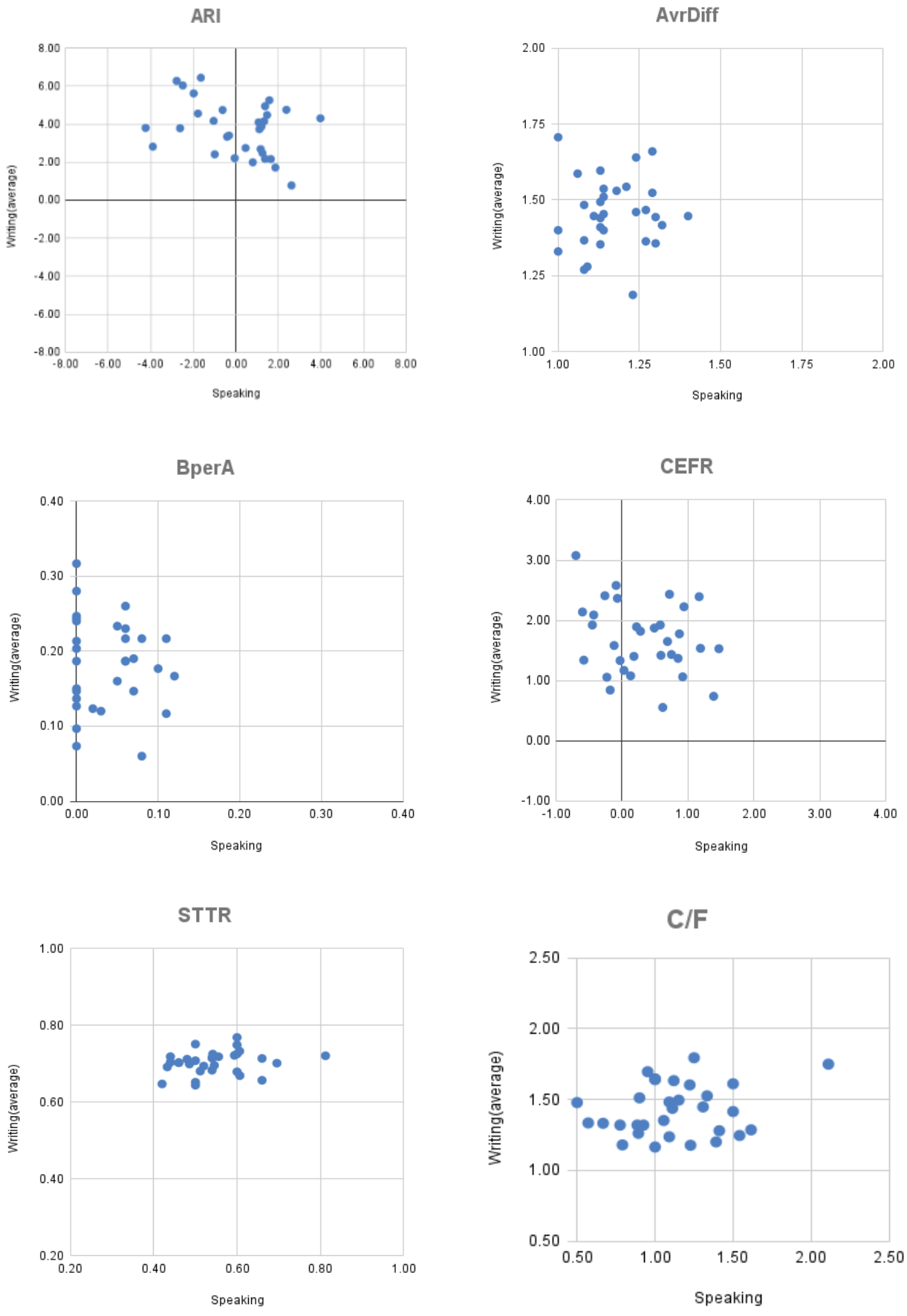

図３　有意差のある指標ごとのスピーキングとライティング（平均）の散布図

5. まとめ

本研究は，日本の高校生同一学習者集団による英語スピーキングとライティングの産出データを分析し，両モード間の言語的特徴の差異を明らかにすることを目的とした。以下では，各リサーチクエスチョン（RQ）に基づいて結果をまとめる。

RQ 1：同一学習者による英語スピーキングとライティングの間には，言語的特徴に統計的に有意な差がみられるか。

分析の結果，ARI，AvrDiff，BperA，CEFR，STTR，C/Fの６指標に統計的有意差が認められた。特に語彙レベルを示すAvrDiffとBperAでは，ライティングがスピーキングよりも高い値を示し，スピーキングでは難度の低い語彙に依存する傾向が明らかになった。この結果は，スピーキングがリアルタイム処理や時間的制約の影響を受けやすいモードであることを示唆する。

RQ 2：スピーキングとライティングにおける各言語的特徴の間には，どの程度の相関関係がみられるか。

相関分析の結果，いずれの指標においても強い正の相関は確認されず，弱い正または負の相関にとどまった。これは，スピーキングとライティングが共通の言語能力に基づいて発達するのではなく，それぞれ独立した処理メカニズムや語彙運用能力に支えられている可能性を示すものである。

これらの結果を総合すると，同一学習者集団においてもスピーキングとライティングの語彙使用は密接に関連しておらず，モードの違いが語彙的特徴に明確な影響を与えることが示された。スピーキングでは，リアルタイム処理や即時性が要求されるため，単語選択がより限定的かつ単純化されやすい傾向があると考えられる。ただし，本研究はコーパス分析に基づくものであり，認知処理負荷を直接的に検証したものではないため，今後はプランニング条件や時間的制約を操作した実験的研究による検証が求められる。

さらに，話し言葉と書き言葉ではレジスター的特性が異なり，スピーキングにおいては必ずしも難度の高い語彙の使用が適切とは限らない。したがって，スピーキング指導では高レベル語彙の使用を一律に促すのではなく，目的や状況に応じて適切な語彙を選択する能力を育成することが重要である。一方で，ライティング指導においては，語彙の多様化や抽象度の高い表現の使用を段階的に指導することが有効であろう。

最後に本研究の制約と課題３点について言及する。１点目は，分析対象としたスピーキングおよびライティングデータの量が限られていたことである。特にスピーキングに関しては１つのテーマしか収集しておらず，一般化するにはデータが非常に少ない点が課題である。本研究のコーパスは，日本の高校１年生31名によるスピーキングおよびライティング産出を対象とし，総語数はスピーキングで1,398語，ライティングで3,707語であった。石川（2021）は，学習者コーパスのような特殊な言語変種を対象としたコーパスにおいては，10万語程度あれば「大規模」と言えると述べている。より一般化された結果を得るためには，10万語を目指したコーパスの構築が必要になるだろう。

２点目は，データの取得方法についてである。今回取得したスピーキングおよびライティングデータは，現実的な状況下で収集されたものであり，それぞれが実際の使用場面に近い形で産出された言語データである。そのため，スピーキングでは口語的な語彙選択や単純構文の使用，ライティングでは文語的で計画的な文構成といった，モード固有のレジスター的特徴が自然に反映されている。このような特性をもつデータを対象とすることで，より現実的な技能間の言語的差異を捉えることが可能となった。一方で，学習者が外部リソースに頼ることなく，自身が持つ語彙知識や文法知識のみを用いて産出するような，統制された条件下で取得したデータではなかった。そのようなデータの収集手法や分析については今後の課題としたい。

そして３点目は，スピーキングデータの処理方法である。フィラーや繰り返し，自己訂正は，学習者の発達段階を示す上でも重要な言語特徴である。本研究では，非流暢性指標を残したものと取り除いたものの２種類のデータを作成したが，取り除く範囲や基準は主観的に判断した部分も一部あった。また，ピリオドやカンマの位置についても主観的な判断に依存した部分があった。今後は，自動音声認識（ASR）を活用することで，書き起こし作業の効率化に加え，研究者の主観的判断を介さず一貫したルールで転記された，より客観的なデータ処理が可能になると考えられる。 もっとも，AIによる自動書き起こし結果にも一定の誤認識や文境界の不自然さが残る可能性があるため，最終的な確認には人間の判断が依然として必要である。

以上の点を踏まえ，今後の研究では，より大規模で多様な学習者データの収集と，スピーキング，ライティングそれぞれのレジスター的特徴を適切に反映した分析手法の精緻化，さらにAI技術を活用した客観的かつ効率的なデータ処理の導入を通して，技能間の言語的差異をよ

り多角的に検証していくことが求められる。

謝辞

本研究は九州大学と新興出版社啓林館の共同研究として助成を受けたものです。また本稿の執筆にあたり，貴重なご意見とご助言をいただいたすべての先生方に深く感謝申し上げます。

引用文献

Abe, M.（2007）Grammatical errors across proficiency levels in L2 spoken and written English. The Economic Journal of Takasaki City University of Economics, 49 (3, 4):117-129.

阿部真理子・近藤悠介（2024）「高校生の縦断的スピーキングコーパスの構築とその活用」『メソドロジー研究部会報告論集』17:53-67. https://doi.org/10.69194/methodologysig.17.4

Baayen, R. H.（2001）Word frequency distributions. Kluwer Academic Publishers.

Biber, D.（1988）Variation across speech and writing. Cambridge, England: Cambridge University Press.

Biber, D., Gray, B., & Poonpon, K.（2011）Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? TESOL Quarterly, 45(1), 5–35. https://doi.org/10.5054/tq.2011.244483

Halliday, M. A. K.（1989）Spoken and written language. Oxford, England: Oxford University Press.

Halliday, M. A. K.（2004）The language of science. London, England: Continuum.

Ishikawa, S.（2015）Lexical development in L2 English learners' speeches and writings. Procedia, Social and Behavioral Sciences, 198, 202–210. https://doi.org/10.1016/j.sbspro.2015.07.437

石川慎一郎（2021）『ベーシックコーパス言語学 第2版』ひつじ書房.

Ishikawa, S.（2023）The ICNALE guide: An introduction to a learner corpus study on Asian learners' L2 English. Routledge. https://doi.org/10.4324/9781003252528

和泉絵美・内元清貴・井佐原均（2004）（編著）『日本人1200人の英語スピーキングコーパス』アルク.

和泉絵美・内元清貴・井佐原均（2005）「エラータグ付き学習者コーパスを用いた日本人英語学習者の主要文法形態素の習得順序に関する分析」『自然言語処理』12(4):211-225. https://doi.org/10.5715/jnlp.12.4_211

Kim, M., & Crossley, S. A.（2023）Lexical and phraseological differences between second language written and spoken opinion responses. Frontiers in Psychology, 14, 1068685. https://doi.org/10.3389/fpsyg.2023.1068685

Kobayashi, Y., Abe, M., & Kondo, Y.（2022）. Exploring L2 spoken developmental measures: Which linguistic features can predict the number of words. English Corpus Studies, 29, 1–18.

国立教育政策研究所教育課程研究センター（2021）「指導と評価の一体化」のための学習評価に関する参考資料:高等学校 外国語. https://www.nier.go.jp/kaihatsu/pdf/hyouka/r030820_hig_gaikokugo.pdf

Lintunen, P., & Mäkilä, M.（2014）Measuring syntactic complexity in spoken and written learner language: Comparing the incomparable? Research in Language, 12(4):377-399. https://doi.org/10.1515/rela-2015-0005

水本篤（2008）「自由英作文における語彙の統計指標と評定者の総合的評価の関係」『統計数理研究所共同研究リポート215 学習者コーパスの解析に基づく客観的作文評価指標の検討』. 15-28.

三谷裕美（2019）「日本人大学生のモノローグ型タスクにおける英語スピーキング能力の発達の評価」『獨協大学外国語教育研究所紀要』7:21-35.

成田眞澄.（2024）英語学習者が産出した英語ライティングの評価に関する考察：評価者はどのように第二言語ライティングを評価しているのか. In Learner Corpus Studies in Asia and the World. 121–132. https://doi.org/10.24546/0100487718

野村真理子（2009）「日本人英語学習者の産出モードの違いによる言語特徴の分析－中学・高校段階の学習者に焦点をあてて－」『四国英語教育学会紀要』29:15-23. https://doi.org/10.32276/seles.29.0_15

Nomura, M.（2012）Variability in articles in Japanese EFL learners' spoken and written production: A learner corpus-based approach. Corpus-based Language Education Research Report, 8:281-302.

佐竹由帆（2015）「日本人英語学習者コーパスにおける動詞-名詞コロケーションとコンビネーション」『情報

学研究』4:118-125.
Selinker, L.（1972）“Interlanguage.” International Review of Applied Linguistics, 10: 209–231.
投野由紀夫（2007）（編著）『日本人中高生一万人の英語コーパス JEFLL Corpus－中高生が書く英文の実態とその分析』小学館.
投野由 紀 夫（2008）「NICT JLE vs. JEFLL: n-gramを用いた語彙・品詞使用の発達」『英語コーパス研究』15:119-133.
投野由紀夫・根岸雅史（2020）（編著）『教材・テスト作成のためのCEFR-Jリソースブック』大修館書店.
Uchida, S., & Negishi, M.（2018）Assigning CEFR-J levels to English texts based on textual features. In Proceedings of Asia Pacific Corpus Linguistics Conference, 4:463-467.
Vibar, J. D.（2022）Syntactic analysis between spoken and written language: A case study on a senior high student’s lexical density. Asian Journal on Perspectives in Education, 3:86-106.

Differences in Linguistic Features Between Japanese High School Students' Spoken and Written English: A Corpus-Based Analysis Using Eight Metrics

MAEDA, Yuji

Abstract

This study aims to elucidate differences in English spoken and written production by the same group of Japanese high school students through a comprehensive analysis of eight textual feature metrics (i.e., ARI, VperSent, AvrDiff, BperA, CEFR, STTR, C/F, and MSL). The analysis was conducted using a small learner corpus consisting of spoken and written data collected from 31 students. The spoken data were elicited through a one-minute monologic speech task, while the written data were collected through three short essay tasks under realistic classroom conditions. Statistical analyses revealed significant differences between spoken and written production for six of the eight metrics, with VperSent and MSL showing no significant differences. Correlation analyses indicated only weak positive or negative relationships between the two modalities, suggesting that learners' performance in speaking and writing is not strongly related across these linguistic dimensions. Overall, the findings suggest that learners tend to rely on simpler vocabulary and exhibit more limited lexical variety when speaking than when writing.

Keywords: L2 speaking, L2 writing, learner corpus