

第36回 研究助成

A 研究部門・報告 II・英語能力テストに関する研究

結束性指標に基づく日本人英語学習者の エッセイライティングスコアの予測

研究者:久保 佑輔 茨城県／筑波大学大学院 在籍

《研究助言者:寺内 一》

概要

ライティングタスクでは、語彙や文法の正確性に加えて、テキストの一貫性や論理性の観点からの指導や評価も求められる。このようなテキストのまとまりを測定する指標の1つに結束性が挙げられる。しかし、どのような結束性の特徴がライティングスコアに寄与するのかは十分に明らかになっていない。そこで本研究では、日本人英語学習者の作成したライティングスコアの予測に寄与する結束性指標を調査した。具体的には、テキストの言語的特徴に基づいて結束性指標を算出する Tool for the Automatic Analysis of Cohesion(TAACO)を用いて、テキストの一貫性とエッセイの質を予測する変数の特定を行った。その結果、テキストの一貫性とエッセイの質の両方において、内容語の重複がマイナスの影響を与えることが示された。また、3人称代名詞を使用して同じ名詞の過剰使用を避けることで、エッセイの質が高まることが確認された。以上のことから、テキストの一貫性やエッセイの質を高めるためには、隣接文間で同じ内容語の繰り返しの使用を防ぐための指導やフィードバックをする必要性などが示唆された。

1

はじめに

グローバル化が急速に進展している現在では、外国語によるコミュニケーション能力は様々な場面で益々必要となることが想定される。このような背景から、英語による発信能力の強化を目指して、高等学校では新たに「論理・表現」が新設された。この科目では、情報や自分の考えについて論理の構成や展開を工夫しながら話したり書いたりする能力の育成を目標としている(文部科学省, 2018)。つまり、産出した英文の正確性の向上だけではなく、内容を効果的に伝えるためにまとまりのあるプロダクトを産出する能力が求められていると言える。また、英語外部試験(e.g., 英検, IELTS)においても、語彙や文法の正確性に加えて、英文のつながりやまとまりの観点が評価項目に含まれている。以上のことから、プロダクトの論理性や一貫性に焦点を当てた指導や評価も取り入れていく必要がある。

このようなプロダクトの内容や構成のつながりを測定する指標の1つに結束性が挙げられる(Crossley et al., 2016a)。一般的に、質の高いプロダクトにとって結束性は重要な要素とされる(Crossley, 2020)。そのため英語熟達度の高い学習者ほど、分かりやすく伝えるために論理的で一貫したプロダクトの産出が可能になると考えられる。しかし、どのような結束性の特徴が質の高いプロダクトと判断されるのか十分に明らかになっていない(e.g., Crossley et al., 2016b; Tsunemoto & Trofimovich, 2024)。

そこで本研究では、日本人英語学習者の作成したエッセイライティングに焦点を当て、テキストの一貫性やエッセイの質の評価に寄与する結束性指標を明らかにする。結束性指標の算出には、英文自動解析ツー

ルであるTool for the Automatic Analysis of Cohesion 2.0(TAACO;Crossley et al., 2019)を使用する。本研究の結果から、ライティング活動における結束性の観点からの指導や評価の方法について示唆を得ることを目指す。

2

先行研究

2.1 結束性

結束性とは、読み手がテキスト内のアイデア同士を結びつけるための明示的な手がかりと定義される(Crossley et al., 2016a)。具体的な明示的手がかりとして、隣接文間での単語や概念の重複が挙げられる。また、連結詞(e.g., because, therefore)を使用することで、内容のつながりを読み手に明示的に伝えることもできる(Halliday & Hasan, 1976)。一貫性もテキストのまとまりを示す言語的特性であるが、これは読み手によるテキスト内容の理解度を指す(O'Reilly & McNamara, 2007)。そのため、背景知識や言語熟達度などの読み手の要因によって一貫性は変化する(McNamara & Kintsch, 1996)。

結束性も一貫性もテキスト内のつながりやまとまりの程度を示す要素である。しかし、一貫性は読み手の主観的な判断を含むため、客観的な指標の算出は困難である。その一方で結束性は、代名詞や連結詞などの表面的な言語的特徴の有無によって変化する(保田, 2024)。したがって、テキスト内のレマの重複の程度や接続語の使用頻度などに基づいて客観的に数値化することができる(Crossley et al., 2016a)。そのため、テキストの結束性を測定するCoh-Metrix(Graesser et al., 2004)やTAACO(Crossley et al., 2016a)などの自動解析ツールがこれまでに開発されている。

結束性の特徴が確認できる英文例を表1にまとめた。(例1)では、前文のcollege studentsを代名詞のtheyで置き換えることによって、隣接文間での結束性を強めている。(例2)では、moneyという同じ内容語を繰り返し使用することによって、隣接文間での結束性を強めている。(例3)では、副詞のlastlyを使用することによって、テキスト全体での結束性を強めている。英文自動解析ツールでは、このように結束性指標を数値化している。より詳細な指標の算出方法についてはCrossley et al.(2019)を参照されたい。

■表1: 結束性の特徴が確認できる英文例

結束性の特徴のある英文の抜粋
(例1) And the college students have much time. So if they want
(例2) they can make money by doing part-time job for their own. Planning how to use money is helpful
(例3) Lastly , they have responsibility for their job.

注: ICNALE Edited Essay(Ishikawa, 2018a)のW_JPN_PTJ0_027_A2_0が作成したエッセイから抽出。結束性の特徴の見られる単語は太字で表示。

2.2

ライティング評価と結束性の関係

評価者によるライティングの質と一貫性の判断との間には関連があることが報告されている(Bae et al., 2016; Crossley & McNamara, 2010; Plakans & Gebril, 2017)。したがってライティング熟達度の向上には、まとまりやつながりのあるテキストを産出する能力の育成が必要と言える。そのため、ライティング熟達度の予測因子の1つとして、結束性が挙げられている(Crossley, 2020; Kojima & Kaneta, 2022; McNamara et al., 2010)。しかし、どのような結束性指標がライティング評価に影響を及ぼすのかについて、L1とL2の学習者を対象とした両先行研究の間で結果は一致していない(e.g., Crossley et al., 2016b)。

2.2.1 L1ライティングにおける結束性の影響

英語母語話者のライティング熟達度に対する結束性の影響は、学年によって異なる傾向が見られる。初期段階では、隣接文間のつながりを示す局所的結束性がライティングの質に寄与する結果が支持されている(Crossley, 2020)。例えば、小学生が産出する質の高いライティングには、因果関係や対立関係を示す接続語や、代名詞や指示詞が多く含まれることが報告されている(Cameron et al., 1995; Myhill, 2008)。その後、学年が上がるにつれて明示的な結束性マーカーの使用頻度が減っていき、局所的結束性指標はライティングスコアの予測に寄与しないことが確認されている(e.g., McNamara et al., 2010)。その一方で、指示詞の使用がライティングスコアを予測する結果も報告されている(MacArthur et al., 2019)。大局的結束性については、段落間の意味的重複や語彙的重複の指標はライティングスコアの予測に寄与することが確認されている(e.g., McNamara et al., 2013)。

まとめると、ライティングの発達初期段階では、局所的結束性の使用はライティングの質に影響を与える。しかし、ある程度のライティングスキルが身に付くにつれて、局所的結束性の影響はライティングトピックや熟達度によって異なることが推察される。しかし、段落間やテキスト全体での結束性(大局的結束性、テキスト全体の結束性)はライティングスコアと関連する傾向にある。

2.2.2 L2ライティングにおける結束性の影響

L1と同様にL2ライティングにおいても、局所的結束性は一般的にライティングの質にマイナスの影響を与える傾向にある(e.g., Crossley, 2020)。例えば、高校生のL2学習者を対象に検証したCrossley and McNamara(2012)は、内容語の重複が評価にマイナスの影響を与えることを報告している。語彙の重複とライティングの質やテキスト一貫性の判断との間には負の相関があることは、その他の先行研究においても確認されている(e.g., Grant & Ginther, 2000; Guo et al., 2013)。同様に、代名詞の使用もライティング能力との間に負の相関が確認されている(Crossley et al., 2016b)。したがって、ライティング能力の高いL2学習者は、より多様な語彙を用いてエッセイを作成する傾向にあると言える。しかし、多様な語彙の使用は評価結果と負の相関を確認した結果も報告されている(Tywoniw & Crossley, 2019)。

連結詞について、質の高いライティングほど多く使用されることが明らかになっている(e.g., Connor, 1990; Duggleby et al., 2016; Tywoniw & Crossley, 2019)。しかし、等位接続詞の使用は負の予測因子として確認されている(Crossley et al., 2016b)。また、連結詞の使用がスコアの予測に寄与する変数として抽出されていない先行研究もある(e.g., Crossley & McNamara, 2012)。

このように、L2ライティング評価における結束性の影響は十分に明らかになっていない。これは、ライティングにおける言語的特徴は、学習者の要因だけでなく、ライティングタスクによっても異なるためであると推察される(Crossley, 2020; Plakans, 2008; Plakans & Gebril, 2013)。したがって、先行研究から得られた知見を一般化するのは難しく、それぞれの文脈に合わせた検証をすることで、結束性がライティング評価に与える影響を解明していく必要があるだろう。

3

本研究

3.1 本研究の目的

結束性は、ライティングの質と一貫性の判断に影響を及ぼす言語的特徴とされる(Crossley, 2020)。そのため、英文自動解析ツールによって、ライティング評価における結束性の影響について包括的検証がこれまでに実施されている(e.g., Crossley et al., 2016b; Kim, 2022; Tywoniw & Crossley, 2019)。しかし、学習者やタスクの要因が影響することから、先行研究間で報告されている結果は一致していないため、特定の文脈ごとに検証する必要がある。また、日本人英語学習者が産出したライティングプロダクトにお

ける結束性の影響を調査した先行研究はいくつかあるものの(e.g., 福田, 2020; 倉橋, 2021; 沢谷・鈴木, 2016), 特定の結束性マーカーに焦点を当てた調査であり, 英文自動解析ツールを用いた結束性指標の包括的な検証は未開拓である。以上のことから, 日本人英語学習者の産出したエッセイにおける結束性指標の影響を調査することで, 先行研究のギャップの解消に取り組む必要がある。

そこで本研究では, どのような結束性指標が日本人英語学習者の作成したエッセイに付与されたスコアの予測に寄与するかを明らかにすることを目的とした。なお, ここでのエッセイスコアとは, テキスト一貫性の判断と全体的なエッセイの質のことを指す。結束性指標の算出では, 幅広く詳細な結束性指標の算出が可能なTAACO 2.0(Crossley et al., 2019)を用いた。本研究の分析結果から, エッセイの一貫性や質を評価するためのルーブリック開発やフィードバックの検討に寄与する内容を提示することを目指す。また, 質の高いエッセいや, まとまりのあるテキスト作成を促す指導方法への示唆を得ることも試みる。上記の目的の達成のため, 以下の検証課題(Research Questions: RQs)を設定した。

RQ1 どのような結束性指標が, テキストの一貫性の判断に影響を及ぼすか

RQ2 どのような結束性指標が, 全体的なエッセイの質に影響を及ぼすか

3.2 データ

本研究では, アジアの国と地域の英語学習者によるスピーチやエッセイを収録した学習者コーパスであるInternational Corpus Network of Asian Learners of English(ICNALE; <https://language.sakura.ne.jp/icnale/>)から, 学習者のCEFRレベルに加えてエッセイスコアが付与されたEdited Essay(Ishikawa, 2018a)を使用した。日本の英語教育の文脈に焦点を当てているため, Edited Essayの中でも日本の大学に通う英語学習者が作成した80本(40名)のエッセイを分析対象とした。ICNALEプロジェクトに参加した学習者全員が, 以下2つのトピックについて論述型エッセイを作成した:(1) It is important for college students to have a part-time job (PTJ), (2) Smoking should be completely banned at all the restaurants in the country(SMK). エッセイ作成時の条件は統制されており, 1つのエッセイにつき40分以内で作成するように指示を受けている。なお, 辞書の使用は禁止されているが, スペルチェックの使用は許可されていた(Ishikawa, 2013)。

ICNALEでは, 語彙サイズテスト(Nation & Beglar, 2007)と英語外部試験(e.g., TOEIC, TOEFL)のスコアに基づいて, 各学習者に4つのCEFRレベル(i.e., A2, B1_1, B1_2, B2)が付与されている。したがって, 本研究ではCEFRのA2からB2レベルの英語熟達度を有する日本人が作成したエッセイが分析対象である。なおEdited Essayでは, それぞれのCEFRレベルに10名の学習者が作成したエッセイが収録されているため, 4つのCEFRレベルごとに20本のエッセイの抽出が可能である(Ishikawa, 2018a)。

ICNALE Edited Essayでは, 英語母語話者で学術論文の校正経験が豊富なプロのエディター5名がESL Composition Profile(Jacobs et al., 1981)に基づいてエッセイを評価した。評価観点の情報を表2に示す。実用性や信頼性の観点からオリジナルのルーブリックを多少改良し, 5つの観点について1~12点の間で採点した。採点の信頼性のため, 8本のエッセイを抽出して各評価者が付与したスコアを共有する小規模の評価者トレーニングを実施した後に, 評価者は担当するエッセイを独立して採点した(Ishikawa, 2018a)。ルーブリックのOrganizationの観点では, テキストの論理的な流れに基づいて評価されるため, 一貫性の判断と類似している。したがって, Organizationの観点のスコアをテキスト一貫性の判断として分析した。また, 5つの観点の合計スコアを全体的なエッセイの質として分析を進めた。

■表2: ICNALE Edited Essayで使用されたループリックの概要(Ishikawa, 2018b)

観点	説明
① Content	知識が豊富で、実質的で、徹底的に練り上げられ、トピックに関連している
② Organization	十分な裏付けに基づいて、明確且つ流暢にアイデアを説明している エッセイは論理的な流れで構成されている
③ Vocabulary	語彙は洗練され、効果的に選択され、適切な範囲で使用されている
④ Language Use	複雑な構文を含み、文法的に正しい文を使用している
⑤ Mechanics	スペル、句読点、大文字、段落分けの正しい規則に基づいている

日本人英語学習者が作成したエッセイの基本情報として、Organizationスコア(テキスト一貫性の判断)とTotalスコア(エッセイの質)の記述統計を表3に示す。各スコアの記述統計をトピックごとに算出したところ、両スコア共にトピックAよりもトピックBの平均点が低いことが確認された。この結果から、トピックの要因の影響が多少表れていると言える。

■表3: 分析対象のエッセイスコアの記述統計

	エッセイ数	M	SD	Min	Max
Organizationスコア	80本	7.513 [7.210, 7.815]	1.359	4	11
	トピックA (PTJ)	7.950 [7.469, 8.430]	1.501	6	11
	トピックB (SMK)	7.075 [6.740, 7.409]	1.047	4	9
Totalスコア	80本	37.212 [36.015, 38.410]	5.383	21	52
	トピックA (PTJ)	39.225 [37.835, 41.066]	5.757	30	52
	トピックB (SMK)	35.200 [33.872, 36.527]	4.152	21	43

3.3 分析に使用する指標

エッセイの結束性指標を算出するため、英文自動解析ツールであるTAACO 2.0(Crossley et al., 2019)を使用した。TAACOは局所的結束性(隣接文間のつながり)、大局的結束性(段落間のつながり)、テキスト全体の結束性における幅広く詳細な150以上の指標を報告する点が、既存のツールとは異なる点である(Crossley et al., 2016a)。TAACOでは表4に示す4つの結束性の特徴を報告する。なお本研究では、複数段落で構成されるエッセイを分析対象としていないため、局所的結束性とテキスト全体の結束性の指標のみ算出する。以下では、4つの結束性の特徴について説明する。各結束性指標の詳細はCrossley et al.(2016a, 2019)を参照されたい。

■表4: TAACOで報告される結束性の特徴

特徴	説明	例
語彙的重複	各品詞のレマのタイプ数を測定	The sun was bright. The day was sunny.
意味的重複	同義語の重複度を測定	The animal was huge. It was a dog.
連結詞	接続詞や語彙的従属詞などの使用を測定	Firstly, she was happy and excited.
既知性	代名詞や指示詞の出現率を測定	The girl was satisfied with what she had.

注: Crossley et al.(2016a)とKim(2022)に基づき作成。結束性の特徴に該当する英単語は太字で表示。

(1)語彙的重複

隣接する2つの文、もしくは隣接する3つの文におけるレマの重複を測定する。テキストの総レマ数や総文数で重複したレマ数を割った値や、隣接文間で共通して使用されているレマがあるかを2値で判断した値によって、この結束性指標は数値化される(Crossley et al., 2019)。内容語と機能語のレマに加えて、名詞、動詞、形容詞、副詞、代名詞の品詞ごとにも報告される。

(2)意味的重複

Corpus of Contemporary American English(COCA)に基づいて、テキスト間の名詞と動詞の同義語の重複を測定する。例えばjumpの同義語には、leap, bound, springなどの関連する単語が含まれる(Crossley et al., 2016a)。意味的重複には次の3つのモデルが含まれる：(a) テキスト内の潜在的な意味のつながりを推定するLatent Semantic Analysis(LSA), (b) テキスト内の潜在的なトピックを推定するLatent Dirichlet Allocation(LDA), (c) テキスト内の語句の意味的類似性を推定するWord2vec(Crossley et al., 2019)。

(3)連結詞

主に2つの側面に基づいて、分析の対象となる連結詞のリストが作成されている：(a) 否定的つながり(e.g., although, unless)と肯定的つながり(e.g., actually, next)に関連する語句、(b) Halliday and Hasan(1976)の提案した文法的結束性の中でも、接続(conjunction)に分類される結束性マーカー(Crossley et al., 2016a)。TAACOではこれらの連結詞の出現率が測定される。

(4)既知性

1人称代名詞(e.g., I, us), 2人称代名詞(e.g., you), 3人称代名詞(e.g., he, them), 主語代名詞(e.g., I, you, she), 数量代名詞(e.g., many)など、さまざまなタイプの代名詞と名詞の比率が報告される。また、定冠詞(e.g., the)や指示詞(e.g., this, those)の出現率も算出される(Crossley et al., 2016a)。

3.4 分析方法

日本人英語学習者の作成したエッセイに付与されたスコア(i.e., Organizationスコア, Totalスコア)を予測するため、本研究では決定木分析と重回帰分析を実施した。分析にはR ver. 4.2.2(R Core Team, 2022)を用いた。分析手順は以下の通りである。

1. ICNALE Edited Essay(Ishikawa, 2018a)から日本人英語学習者の作成したエッセイ80本の抽出
2. TAACO 2.0(Crossley et al., 2019)によって各エッセイの結束性指標の算出
3. 結束性指標を説明変数、スコアを目的変数とする決定木分析の実施
4. 重要度が高いと判断された結束性指標を説明変数、スコアを目的変数とする重回帰分析

TAACOでは150以上の結束性指標が算出されるため、指標を選定するために決定木分析を実施した。決定木分析ではrpartパッケージ(Therneau & Atkinson, 2022)を使用し、最も精度の高いノード数で決定木を作成した。なお、決定木を複数作成することで精度を高めるランダムフォレストも実施したもの、予測精度は2%程度と信頼できる結果が得られなかった。

続いて、決定木分析によって重要度が高い変数として抽出された結束性指標を説明変数に設定してエッセイスコアを予測するため、ステップワイズ法による重回帰分析を実施した。ステップワイズ法は近年批判されている手法ではあるものの(e.g., Maie et al., 2023)、結束性がライティング評価に与える影響は十分に明らかになっていない(e.g., Crossley, 2020)。したがって、本研究は探索的な検証の性質をもっていると言えるため、この手法を採用した(Eguchi & Kyle, 2020)。

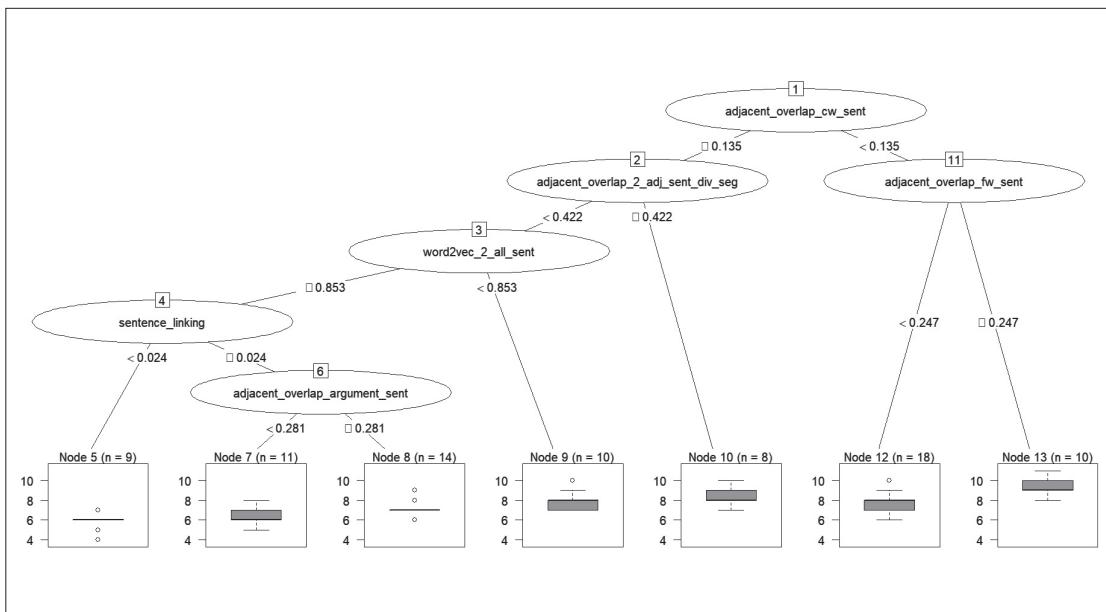
本研究で用いたデータと分析コードの詳細は資料1(補足データ <https://osf.io/h5vpr>)を参照されたい。分析コードは平井他(2022)や小林他(2020)に基づいて作成した。

4 結果と考察

4.1 テキスト一貫性の判断(Organizationスコア)を予測する結束性指標

テキスト一貫性の判断を予測する決定木を図1、分析の結果抽出された結束性指標の概要を表5に示す。決定木分析の予測精度は56.884%で、語彙的重複の指標が重要度の高い変数として多く抽出された。

表5に示す6つの指標間の相関係数は.80以下で許容度の逆数を示すVIF(Variance Inflation Factor)の値も10以下であったため(資料1と2参照)、多重共線性の問題は確認されなかった(平井他, 2022)。重回帰分析の結果を表6に示す。2つの指標が削除され、残った4つの指標でスコアを約17%予測することが確認された。以下では、有意差が確認された2つの結束性指標に焦点を当てて、考察を行う。



■図1: テキスト一貫性の判断(Organizationスコア)を予測する決定木

■表5: テキスト一貫性の判断を予測する決定木で抽出された結束性指標の概要

結束性指標		結束性の特徴	説明
1	adjacent sentence overlap content lemmas	語彙的重複	隣接文間で少なくとも1回出現する内容語のレマタイプの数
2	adjacent sentence overlap function lemmas	語彙的重複	隣接文間で少なくとも1回出現する機能語のレマタイプの数
3	adjacent two-sentence overlap adjective lemmas (sentence normed)	語彙的重複	次の2文に少なくとも1回出現する形容詞のレマタイプの数
4	word2vec similarity (two adjacent sentences)	意味的重複	すべての隣接文(2文スパン)間の意味的類似度スコアの平均
5	sentence linking	連結詞	文を連結する機能をもつ単語の数 EX) nonetheless, therefore
6	adjacent sentence overlap noun and pronoun lemmas	語彙的重複	隣接文間で少なくとも1回出現する名詞と代名詞のレマタイプの数

■表6: 結束性指標によるテキスト一貫性の判断(Organizationスコア)を予測する重回帰分析の結果

結束性指標	B	95% CI	SE B	β	t	p
adjacent sentence overlap content lemmas	-11.665	[-18.481, -4.747]	3.374	-0.586	-3.457	.001
adjacent two-sentence overlap adjective lemmas	1.954	[0.195, 3.450]	0.742	0.271	2.633	.010
sentence linking	17.601	[-5.111, 38.769]	10.868	0.162	1.629	.108
adjacent sentence overlap noun and pronoun lemmas	5.720	[-1.694, 11.884]	2.934	0.288	1.950	.055

注. $R^2 = .208$, adjusted $R^2 = .166$.

表6から、隣接文間で内容語が多く重複しているエッセイほど、テキスト一貫性が低いと判断される傾向にあることが示された($t = -3.457$, $p = .001$)。このような特徴を含むエッセイの抜粋を以下に示す(Excerpt 1, Excerpt 2)。隣接文間で繰り返し使用している内容語レマは、太字で表記している。

高いスコアが付与されたエッセイ(Excerpt 1)では、隣接文間で重複している内容語レマがほとんどないことが確認できる。また、低いスコアが付与されているエッセイ(Excerpt 2)では、プロンプト記載の単語(part-time job)を繰り返し使用している傾向にあることも示された。以下ではトピックA(PTJ)のみを抜粋しているが、トピックB(SMK)についても同様の傾向が確認された。

Excerpt 1: 内容語レマの重複の指標が低くてスコアが高いエッセイの一部

- I agree that it is important for college students to have a part-time job. I have three reasons to support my stance. Firstly, a **part-time job** is an important chance to experience the society. Because a college is the place to prepare before going to the real society, **part-time job** is very necessary to know

(ID = W_JPN_PTJO_015_B1_2, スコア = 10, トピック = PTJ)

Excerpt 2: 内容語レマの重複の指標が高くてスコアが低いエッセイの一部

- I agree with the statement because **part-time job** gave me lots of experiences. For example, I understood following things through **part-time job**. It is hard to **make** money because I **have** to go to **part-time job** when I promised day and time, even if I don't want it. I **have** to keep time because if I delayed, I interrupted colleague's **jobs** and **make** bad relationships. I can't stay I'm center of the world because guests choice restaurant which they want to go, so I need to **make** them comfortable for they choose my shop. I **learned** from part-time **jobs** how important money and to **make** money, to **make** relationship. But I **learned**....

(ID = W_JPN_PTJO_024_B1_1, スコア = 6, トピック = PTJ)

また、内容語でも形容詞が隣接文間で重複しているエッセイでは、テキスト一貫性が高いと判断される傾向にあることが示された($t = 2.633$, $p = .010$)。このような特徴を含むエッセイの抜粋を以下に示す(Excerpt 3)。隣接文間で繰り返し使用している形容詞レマは太字で表記している。Excerpt 3では、主張を支える理由部分でトピックについての利点や欠点を説明するために形容詞(e.g., important, good, better, less)を使用する傾向にあることが確認された。このような特徴は両トピックのエッセイで見られた。

しかし、形容詞が隣接文間で重複していてもテキスト一貫性が低いと判断されるエッセイ(Excerpt 4)も確認された。このようなエッセイでは、他の形容詞(e.g., college)を繰り返し使用していたり、利点や欠点の説明で形容詞を使用していても同じ単語(e.g., bad)を繰り返し使用していたりする特徴が見られた。これは、内容語レマを繰り返し使用すると、テキスト一貫性の判断にマイナスの影響を与えるという上記の結果と一致するものである。

Excerpt 3:形容詞レマの重複の指標が高くてスコアが高いエッセイの一部

- I think that it is **important** for college student to have a part-time job. This is because that they can learn a lot of things through their part-time jobs. For example, they work in a store, they learn that to smile is very **important**, and And this is also because that to do a lot of part-time jobs when they were college students will become a **good** memory in the future. They will enjoy their lives **better** by

(ID = W_JPN_PTJ0_022_B1_1, スコア = 10, トピック = PTJ)

- I think that smoking should be As a result of this, in the case where a restaurant does not ban smoking completely, it gathers **less** customers than if it banned smoking. This leads to the **less** benefit of

(ID = W_JPN_SMK0_009_B2_0, スコア = 8, トピック = SMK)

Excerpt 4:形容詞レマの重複の指標が高いのにスコアが低いエッセイの一部

- I agree with this statement The second reason is that the wage for the **college** students is cheaper than other workers. So if **college** students work part time, many companies and shops do without paying higher wage to employments. But now the condition of business is very bad in Japan. Some people think because of the **college** students part-time, many unemployed people miss their opportunities to work and the unemployed rate is high. Certainly the population of workforce doesn't contain the number of **college** students and the part-time by **college** students causes ...

(ID = W_JPN_PTJ0_019_B1_2, スコア = 6, トピック = PTJ)

- I agree the idea Smoking in restaurants has two **bad** things. One is that smell of smoking is **bad** for Smell can make foods good, or otherwise it can make foods **bad**. And the other is that smoke is **bad** for health. In restaurants, there are a lot of people, babies, children, teens, adults, and olds. Smoking has **bad** influence in them all. Smoking is **bad** for health of all people, especially **bad** for

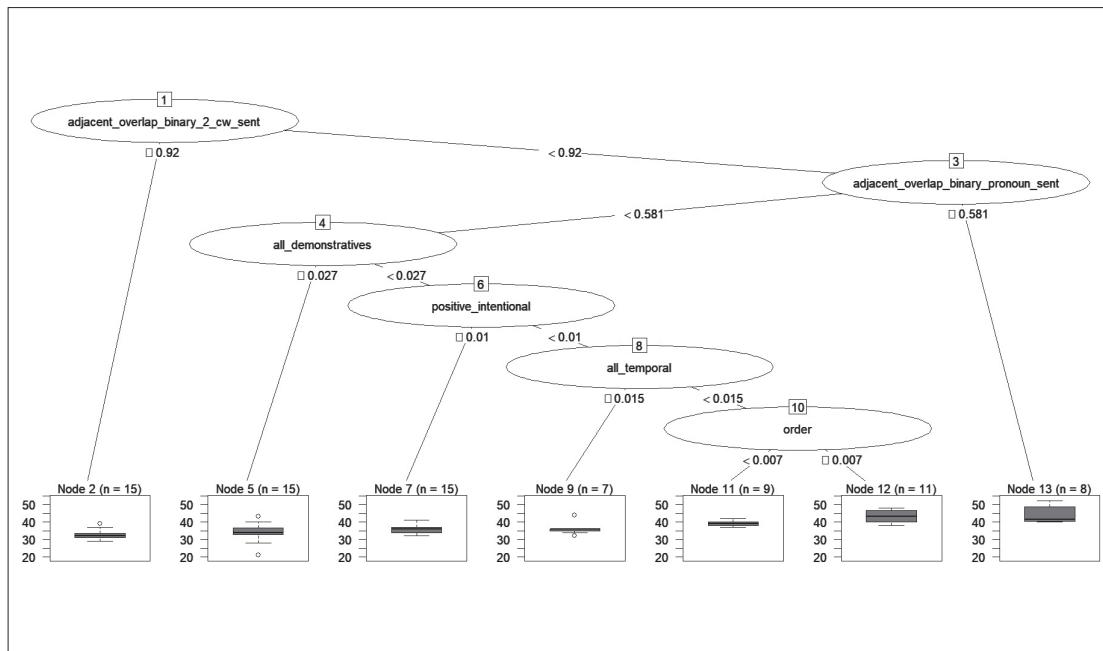
(ID = W_JPN_SMK0_018_B1_1, スコア = 6, トピック = SMK)

形容詞レマの重複の結果について、利点や欠点を明示的に表現して主張と理由の間のつながりを明確にさせることによって、テキスト一貫性の判断に影響を与えた可能性が示唆された。しかし、形容詞を使用せずとも、トピック内容の利点や欠点を説明することは容易である。そのため先行研究では、形容詞レマの繰り返し使用がエッセイの一貫性に与える影響について報告しているものは見当たらない。また、今回のデータにおいて高いスコアが付与されたエッセイで確認されることの多かった形容詞(e.g., good, bad)は、どれも洗練性の低い語彙であった。一般的に、語彙の洗練性はライティング熟達度を予測する因子として考えられているため(e.g., Crossley, 2020; Guo et al., 2013; Kojima & Kaneta, 2022), 語彙研究の知見とは矛盾する結果である。以上のことから、形容詞レマの重複の観点のみではこの結果の説明は難しく、主張を支える理由の内容や質など、他の要素が影響している可能性もあり得るだろう。実際に、今回使用したICNALE Edited Essayでの採点では、Organizationは5つの観点の合計スコアと最も強い正の相関があり($r = .889$)、アイデアの明確性や流暢性などの多様な要素が含まれていることが示唆されている(Ishikawa, 2018b)。つまり、テキストのつながりやまとまり以外の要素もOrganizationの評価に関連していた可能性がある。

4.2 エッセイの質(Totalスコア)を予測する結束性指標

エッセイの質を予測する決定木を図2、分析の結果抽出された結束性指標の概要を表7に示す。決定木分析の予測精度は57.100%で、連結詞の指標が多く抽出されたが、重要度の高い変数(決定木の上のノードにある変数)は語彙的重複の指標であった。

表7に示す6つの指標間の相関係数は.80以下でVIFの値も10以下であったため(資料3参照)、多重共線性の問題は確認されなかった(平井他, 2022)。重回帰分析の結果を表8に示す。3つの指標が削除され、残った3つの指標でスコアを約18%予測することが確認された。 $(t = -2.105, p = .039)$



■図2: エッセイの質(Totalスコア)を予測する決定木

■表7: エッセイの質を予測する決定木で抽出された結束性指標の概要

結束性指標		結束性の特徴	説明
1	binary adjacent two-sentence overlap content lemmas	語彙的重複	次の2文に何らかの内容語レマが重複する文の数
2	binary adjacent sentence overlap pronoun lemmas	語彙的重複	隣接文間で何らかの代名詞レマが重複する文の数
3	demonstratives	連結詞	指示詞の使用頻度 EX) this, that, these
4	positive intentional connectives	連結詞	目的・願望・手段などの肯定的なつながり示す単語の数 EX) by, so, want
5	temporal connectives	連結詞	時間経過や順序を示す単語の数 EX) a consequence of, after, again
6	order	連結詞	順序を示す機能をもつ単語の数 EX) to begin with, next, first

■表8: 結束性指標によるエッセイの質(Totalスコア)を予測する重回帰分析の結果

結束性指標	B	95% CI	SE B	β	t	p
binary adjacent two-sentence overlap content lemmas	-8.012	[-15.555, 0.180]	3.806	-0.209	-2.105	.039
binary adjacent sentence overlap pronoun lemmas	7.847	[1.638, 12.531]	2.565	0.283	3.059	.003
positive intentional connectives	.182.983	[-324.696, -15.090]	74.792	-0.237	-2.447	.017

注. $R^2 = .214$, adjusted $R^2 = .183$.

以下では、有意性が確認された3つの結束性指標について考察を行う。

まず、隣接文間で内容語が多く重複しているエッセイほど、質が低いと判断される傾向にあることが示された($t = -2.105$, $p = .039$)。つまり、質の高いエッセイでは同じ内容語レマを隣接文間で繰り返し使用しない特徴が見られる。この結果は、テキスト一貫性の判断を予測するRQ1の結果や、L2ライティングにおける結束性の影響を調査した多くの先行研究(e.g., Crossley & McNamara, 2012)の結果とも一致する。したがって、内容語を繰り返し使用して局所的結束性を高めているエッセイは、テキスト一貫性やエッセイの質の判断の両方において、マイナスの影響を及ぼすことが明らかになった。

次に、内容語でも代名詞が隣接文間で重複しているエッセイでは、質が高いと判断される傾向にあることが示された($t = 3.059$, $p = .003$)。このような特徴を含むエッセイの抜粋を以下に示す(Excerpt 5)。隣接文間で繰り返し使用している代名詞レマは太字で表記している。

高いスコアが付与されたエッセイ(Excerpt 5)では、3人称代名詞theyのレマを使用して、college studentsやpeopleの繰り返し使用を避けながら理由を説明している特徴が見られた。但し、この特徴はトピックA(PTJ)のみでトピックB(SMK)では確認されなかった。一方で、代名詞が隣接文間で重複していてもエッセイの質が低いと判断されるエッセイ(Excerpt 6)も確認された。このようなエッセイでは、3人称代名詞theyのレマではなく、1人称代名詞Iのレマを過剰使用している特徴が見られた。この傾向は両トピックにおいて発見された。これは日本人英語学習者が英語母語話者と比較してIを多用することを報告している先行研究(e.g., Natsukari, 2012)と一致する結果であり、エッセイの質が低いほどこの特徴は顕著であることが示唆された。

Excerpt 5:代名詞レマの重複の指標が高くてスコアが高いエッセイの一部

- I think that it is important for college student to have a part-time job. This is because that **they** can learn a lot of things through their part-time jobs. For example, **they** work in a store, **they** learn that to smile is very important, and that there are a lot of types of people in the society. This is also because that **their** experiences which were got by **their** part-time jobs are very useful, when **they** look for what **they** want to be in the future. **They** can do various types of part-time jobs, so it helps for **them** to look for what **they** want to be in the future. And this is because that **they** cannot do a part-time job when **they** start to work. When **they** are college students, **they** have a lot of time to do a part-time job. But when **they** start to work, **they** don't have. And this is also because that to do a lot of part-time jobs when **they** were college students will become a good memory in the future. **They** will enjoy **their** lives better by

(ID = W_JPN_PTJ0_022_B1_1, スコア = 48, トピック = PTJ)

Excerpt 6:代名詞レマの重複の指標が高いのにスコアが低いエッセイの一部

- I agree with the statement because part-time job gave **me** lots of experiences. For example, **I** understood following things through part-time jobs. It is hard to make money because **I** have to go to part-time job when **I** promised day and time, even if I don't want it. **I** have to keep time because if **I** delayed, I interrupted colleague's jobs and make bad relationships. **I** can't stay I'm center of the world because guests choice restaurant which they want to go, so **I** need to make them comfortable for they choose my shop. **I** learned from part-time jobs how important money and to make money, to make relationship. But **I** learned it is makes **me** tired to have job at the same time. **I** need time to have a rest for recovery. So **it** is interrupt **my** study to....

(ID = W_JPN_PTJ0_024_B1_1, スコア = 31, トピック = PTJ)

最後に、連結詞の結束性指標であるpositive intentional connectivesについて、肯定的つながりを示す単語数が多いほどエッセイの質が低いと判断されることが示された($t = -2.447, p = .017$)。このような特徴を含むエッセイの抜粋を以下に示す(Excerpt 7, Excerpt 8)。positive intentional connectivesの機能を持つ単語は太字で表記している。なお、この指標に該当する主な語句は以下の通りである:by, in order, make, purpose, so, want(Crossley et al., 2019)。

低いスコアが付与されたエッセイ(Excerpt 7)では、make, want, soを繰り返し使用している特徴が見られた。一方でpositive intentional connectivesに該当する単語を多く使用していても、質が低いと判断されていないエッセイ(Excerpt 8)も確認された。このようなエッセイでは、上記の3つの単語以外にもin orderやbyも使用している特徴が見られた。この傾向は両トピックで確認された。したがって、質の高いエッセイを作成する際にはwantやmakeなどの洗練性の低い動詞の過剰使用を抑え、so以外の接続語を使用する点に留意する必要があることが示唆された。また、この結果は内容語レマを繰り返し使用するとテキスト一貫性やエッセイの質が低くなるという上記の結果を支持するものもあると言える。

Excerpt 7:positive intentional connectivesに該当する単語数が多くてスコアが低いエッセイの一部

- I agree with the statement It is hard to **make** money because I have to go to part-time job when I promised day and time, even if I don't **want** it. I have to keep time because if I delayed, I interrupted colleague's jobs and **make** bad relationships. I can't stay I'm center of the world because guests choice restaurant which they **want** to go, **so** I need to **make** them comfortable for they choose my shop. I learned from part-time jobs how important money and to **make** money, to **make** relationship. But I learned it is **makes** me tired to have job at the same time. I need time to have a rest for recovery. **So** it is

(ID = W_JPN_PTJ0_024_B1_1, スコア = 31, トピック = PTJ)

Excerpt 8:positive intentional connectivesに該当する単語数が多いがスコアが低くないエッセイの一部

- I agree with Smoking from cigarette **makes** me uncomfortable, especially when I have **So**, Japanese government should **make** smoking banned at all restaurants **in order** to save all citizens' health. From my personal experience, I believe that smoking should be banned in all restaurants in Japan. One of my friends could quite to smoke **by** the number of This **makes** her smoking less cigarettes per a day. Then she doesn't **want** to smoke anymore now. Therefore since I heard this story, I've believed that policy of smoking being banned in all restaurants **makes** it

(ID = W_JPN_SMK0_014_B1_1, スコア = 40, トピック = SMK)

5 結論

5.1 結果のまとめと教育的示唆

本研究では、TAACO 2.0 (Crossley et al., 2019) によって算出した指標を用いて、日本人英語学習者のエッセイスコア (i.e., テキスト一貫性の判断を示す Organization スコア、エッセイの全体的な質を示す Total スコア) を予測する結束性指標を検証した。RQ1(どのような結束性指標が、テキストの一貫性の判断に影響を及ぼすか) では、内容語の重複はテキスト一貫性の判断にマイナスの影響を与えることが示された。また、形容詞の重複はプラスの影響を与える結果が得られたものの、これは ICNALE Edited Essay で使用されたルーブリックにおける Organization の観点の記述子に影響されている可能性が考えられる (Ishikawa, 2018a)。但し、同じ形容詞を繰り返し使用している場合には、スコアにマイナスの影響を与える傾向も見られた。

RQ2(どのような結束性指標が、全体的なエッセイの質に影響を及ぼすか) についても、内容語の重複が多いほど低いスコアが付与されていた。また、代名詞の重複もエッセイの質の予測に寄与することが明らかになったが、1人称代名詞の多用は避け、名詞の繰り返し使用を防ぐために3人称代名詞を用いることで、エッセイの質が高くなる傾向が示された。さらに、連結詞の指標である positive intentional connectives の中でも make, want, so の多用を避けることで、エッセイの質にマイナスの影響を与えないことが示唆された。つまり、同じ単語を繰り返し使用している場合には、エッセイの質が下がる傾向にあることが示唆された。

以上のことから、本研究から得られた主な結果は以下の通りである：(1) テキスト一貫性と質が高いと判断されるエッセイでは、内容語の繰り返し使用が少ない；(2) 質が高いと判断されるエッセイでは、1人称代名詞の繰り返し使用が少なく、名詞の重複を避けるために3人称代名詞を使用している傾向にある。但し、代名詞の使用はトピックに影響を受ける可能性がある点に留意する必要がある。

本研究の結果に基づく、CEFR の A2 レベル以上の日本人英語学習者が作成したエッセイライティングにおける結束性に焦点を当てた指導と評価に対する示唆として、以下の2点が挙げられる。第1に、テキスト一貫性やエッセイの質を高める指導や評価では共通して、隣接文間での内容語の繰り返し使用を避けるように学習者に注意を向けさせることが望ましい。そのため、ルーブリックでは内容語の重複に関する評価項目を設けることで、同じ内容語レマの過剰使用を避けるように促す必要があるだろう。

第2に、エッセイの質を高める指導や評価では、内容語の重複に加えて代名詞の使用にも着目することが推奨される。エッセイの全体的な質の向上に焦点を当てたライティングタスクでは、1人称代名詞の過剰使用を避けて、3人称代名詞を使用して名詞の繰り返し使用を防ぐための明示的な指導や、フィードバックを提示することが望ましいだろう。

本研究では、連結詞の中でもテキスト一貫性の判断やエッセイの質の予測に、because や therefore などの談話標識の使用は影響を与えないという結果が得られた。これは、ほとんどのエッセイで談話標識が確認されたためだと考えられる。しかし、CEFR の A2 レベル以上の日本人英語学習者のエッセイであっても談話標識の出現が確認されているだけで、正確に使用できていない場合も散見された。そのため、代名詞だけではなく、談話標識の適切な使用についての指導や評価も重要事項であると言える。

5.2 本研究の限界点と今後の課題

今後は、主に以下の3点の限界点を調査することで、結束性の観点からのライティング評価や、今回の分析で得られた結果の一般化につなげていく必要がある。第1に、本研究では2つのトピックをまとめて分析を行ったため、トピックの要因を統制することができなかった。実際に2つのトピックの平均点にも

多少の開きがあることが確認されているため(表3参照), 今後はサンプルサイズを増やしてトピックごとに調査する必要がある。これにより, 決定木分析の予測精度の向上も期待できる。

第2に, 本研究では重回帰分析にステップワイズ法を用いたことにより, 結果の解釈が難しい結束性指標がスコアを予測する変数として抽出された場合があった。例えば, テキスト一貫性の判断を予測する指標を検証するRQ1では, 形容詞の重複がスコアの予測に寄与することが示された。しかし, 形容詞レマの隣接文間での繰り返し使用がプラスの影響を及ぼすことを報告する先行研究は, 私の知る限りではない。結束性指標がライティング評価に与える影響は十分に明らかになっていないものの(e.g., Kojima & Kaneta, 2022), 先行研究で示された結果の傾向や研究目的に基づいて, スコアに寄与する結束性指標をある程度整理した上で検証する方法も実施する必要があるだろう。

第3に, 本研究ではライティングスコアの分散の説明率が20%以下と低い結果が報告された。先行研究においても同様の説明率であるが(e.g., Tywoniw & Crossley, 2019), これは結束性マーカーの誤用を考慮していないことが要因の1つとして考えられる。TAACOではエッセイに出現している結束性マーカーに基づいて指標を数値化しているため, 不適切な使用であったとしても指標値に反映されてしまう。そのため, 結束性マーカーの誤用に焦点を当てて再度検証することも求められる。また, 結束性マーカーが使用されるべきテキスト内の適切な位置や距離の観点からも検証することで(鈴木, 2023; 保田, 2024), ライティングスコアを予測する結束性指標の特定につながるだろう。

その他の要因として, 結束性指標とルーブリックの評価方法の特性の違いも挙げられる。結束性指標などの比率尺度は, ルーブリックによる評価尺度よりも狭い構成概念の測定に焦点を当てる(Koizumi et al., 2022)。つまり, 個々のつながりを測定する結束性指標では評価者によるテキスト一貫性の判断の20%の予測に止まったという結果は, 一貫性の判断には結束性以外の要素も大きく関わることを示唆するとも考えられる。Organizationの観点では一貫性以外の要素の影響も考えられるものの(Ishikawa, 2018a), 高いスコアが付与されるエッセイでは, 洗練性の高い語彙や統語的に複雑な文を使用しながらまとまりのあるテキストを作成することが想定される(Crossley, 2020)。そのため, 一貫したテキスト形成には結束性以外も関わることになる。したがって, ライティング熟達度の主要な予測因子である語彙の洗練性や複雑性などの指標も含めた検証も今後行う必要性がある。

謝辞

本研究を実施する機会をくださった公益財団法人 日本英語検定協会の皆様や選考委員の先生方に厚く御礼申し上げます。中でも, 研究助言者である寺内一先生には非常に有益なご助言を賜りましたこと, 深く感謝申し上げます。また, ご指導いただきました筑波大学大学院の平井明代先生と小泉利恵先生に心より感謝申し上げます。なお, 本報告書に関して, 開示すべき利益相反関連事項はありません。

引用文献

- Bae, J., Bentler, P. M., & Lee, Y.-S. (2016). On the role of content in writing assessment. *Language Assessment Quarterly*, 13(4), 302-328. <https://doi.org/10.1080/15434303.2016.1246552>
- Cameron, C. A., Lee, K., Webster, S., Munro, K., Hunt, A. K., & Linton, M. J. (1995). Text cohesion in children's narrative writing. *Applied Psycholinguistics*, 16(3), 257-269. <https://doi.org/10.1017/s0142716400007293>
- Connor, U. (1990). Linguistic/rhetorical measures for international student persuasive writing. *Research in the Teaching of English*, 24(1), 67-87. <http://www.jstor.org/stable/40171446>
- Crossley, S. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research*, 11(3), 415-443. <https://doi.org/10.17239/jowr-2020.11.03.01>
- Crossley, S. A., Kyle, K., & McNamara, D. S. (2016a). The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behavior Research Methods*, 48(4), 1227-1237. <https://doi.org/10.3758/s13428-015-0651-7>
- Crossley, S. A., Kyle, K., & McNamara, D. S. (2016b). The development and use of cohesive devices in L2 writing and their relations to judgments of essay quality. *Journal of Second Language Writing*, 32, 1-16. <http://dx.doi.org/10.1016/j.jslw.2016.01.003>

引用文献

- Crossley, S. A., Kyle, K., & Dascalu, M. (2019). The Tool for the Automatic Analysis of Cohesion 2.0: Integrating semantic similarity and text overlap. *Behavioral Research Methods*, 51(1), 14–27. <https://doi.org/10.3758/s13428-018-1142-4>
- Crossley, S. A. & McNamara, D. S. (2010). Cohesion, coherence, and expert evaluations of writing proficiency. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 32(32).
- Crossley, S. A., & McNamara, D. S. (2012). Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading*, 35(2), 115–135. <https://doi.org/10.1111/j.1467-9817.2010.01449.x>
- Duggleby, S. J., Tang, W., & Kuo-Newhouse, A. (2016). Does the use of connective words in written assessments predict high school students' reading and writing achievement? *Reading Psychology*, 37(4), 511–532. <https://doi.org/10.1080/02702711.2015.1066910>
- Eguchi, M., & Kyle, K. (2020). Continuing to explore the multidimensional nature of lexical sophistication: The case of oral proficiency interviews. *The Modern Language Journal*, 104(2), 381–400. <https://doi.org/10.1111/mljl.12637>
- 福田陽子 (2020). 「英語学習者のライティングにおける一貫性・結束性の量的分析から質的分析へ—「読み手からみた不自然さ」を左右する要因とは—」『EIKEN BULLETIN』, 32号, 44–67. https://www.eiken.or.jp/center_for_research/pdf/bulletin/vol32/vol_32_p44-p67.pdf
- Grant, L., & Ginther, A. (2000). Using computer-tagged linguistic features to describe L2 writing differences. *Journal of Second Language Writing*, 9, 123–145. [https://doi.org/10.1016/s1060-3743\(00\)00019-9](https://doi.org/10.1016/s1060-3743(00)00019-9)
- Graesser, A.C., McNamara, D.S., Louwerse, M.M. & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, and Computers*, 36, 193–202. <https://doi.org/10.3758/BF03195564>
- Guo, L., Crossley, S. A., & McNamara, D. S. (2013). Predicting human judgments of essay quality in both integrated and independent second language writing samples: A comparison study. *Writing Assessment*, 18, 218–238. <https://doi.org/10.1016/j.awas.2013.05.002>
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Longman.
- 平井明代・岡秀亮・草薙邦広 (2022).『教育・心理系研究のためのRによるデータ分析—論文作成への理論と実践集』東京書籍。
- Ishikawa, S. (2013). The ICNALE and sophisticated contrastive interlanguage analysis of Asian learners of English. *Learner Corpus Studies in Asia and The World*, 1, 91–118. <https://doi.org/10.24546/81006678>
- Ishikawa, S. (2018a). The ICNALE edited essays: A dataset for analysis of L2 English learner essays based on a new integrative viewpoint. *English Corpus Studies*, 25, 117–130. https://jaecs.com/jnl/ECS25/ECS25_117-130.pdf
- Ishikawa, S. (2018b). Comparison of three kinds of alternative essay-rating methods to the ESL composition profile. *International Journal of Computer-Assisted Language Learning and Teaching*, 8(4), 32–44. <https://doi.org/10.4018/IJCALLT.2018100103>
- Jacobs, H. L., Zinkgraf, S. A., Wormuth, D. R., Hartfiel, V. F., & Hughey, J. B. (1981). *Testing ESL composition: A practical approach*. Newbury House.
- Kim, J. (2022). The use of cohesive devices in Korean EFL writing across different proficiency levels. *Korean Journal of English Language and Linguistics*, 22, 1078–1100. <https://doi.org/10.15738/kjell.22.202210.1078>
- 小林雄一郎・濱田彰・水本篤 (2020).『Rによる教育データ分析入門』オーム社。
- Koizumi, R., In'ami, Y., & Jeon, E. H. (2022). L2 speaking and its internal correlates: A meta-analysis. In E. H. Jeon & Y. In'ami (Eds.), *Understanding L2 proficiency: Theoretical and meta-analytic investigations* (pp. 307–158). John Benjamins. <http://doi.org/10.1075/bpa.13.10koi>
- Kojima M., Kaneta T. (2022). L2 writing and its internal correlates: A meta-analysis. In Jeon E. H., In'ami Y. (Eds.), *Understanding L2 proficiency: Theoretical and meta-analytic investigations* (pp. 109–158). John Benjamins. <https://doi.org/10.1075/bpa.13.05koj>
- 倉橋佑輔 (2021).「英語ライティングにおける結束性」『神戸大学国際コミュニケーションセンター論集』, 17巻, 85–110. <https://doi.org/10.24546/81012587>
- Maie, R., Eguchi, M., & Uchihara, T. (2023). Arbitrary choices, arbitrary results: Three cases of multiverse analysis in L2 research. *Research Methods in Applied Linguistics*, 3(2), 100124. <https://doi.org/10.1016/j.rmal.2024.100124>
- MacArthur, C. A., Jennings, A., & Philippakos, Z. A. (2019). Which linguistic features predict quality of argumentative writing for college basic writers, and how do those features change with instruction? *Reading and Writing*, 32(6), 1553–1574. <https://doi.org/10.1007/s11145-018-9853-6>
- McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). The linguistic features of writing quality. *Written Communication*, 27(1), 57–86. <https://doi.org/10.1177/0741088309351547>
- McNamara, D. S., Crossley, S. A., & Roscoe, R. (2013). Natural language processing in an intelligent writing strategy tutoring system. *Behavior Research Methods*, 45(2), 499–515. <https://doi.org/10.3758/s13428-012-0258-1>
- McNamara, D. S., & Kintsch, W. (1996). Learning from text: Effects of prior knowledge and text coherence. *Discourse Processes*, 22(3), 247–288. <https://doi.org/10.1080/01638539609544975>
- McNamara, D. S., Crossley, S. A., & Roscoe, R. (2013). Natural language processing in an intelligent writing strategy tutoring system. *Behavior Research Methods*, 45(2), 499–515. <https://doi.org/10.3758/s13428-012-0258-1>
- Myhill, D.A. (2008). Towards a linguistic model of sentence development in writing. *Language and Education*, 22(5), 271–288.

引用文献

- https://doi.org/10.1080/09500780802152655
文部科学省 (2018).『高等学校学習指導要領(平成30年告示)解説【外国語編 英語編】』開隆堂.
Nation, P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31(7), 9-13.
Natsukari, S. (2012). Use of I in essays by Japanese EFL learners. *JALT journal*, 34(1), 61-78. https://doi.org/10.37546/JALTJJ35.1-4
O'Reilly, T., & McNamara, D. S. (2007). Reversing the reverse cohesion effect: Good texts can be better for strategic, high-knowledge readers. *Discourse Processes*, 43(2), 121-152. https://doi.org/10.1080/01638530709336895
Plakans, L. (2008). Comparing composing processes in writing-only and reading-to-write test tasks. *Assessing Writing*, 13(2), 111-129. https://doi.org/10.1016/j.asw.2008.07.001
Plakans, L., & Gebril, A. (2013). Using multiple texts in an integrated writing assessment: Source text use as a predictor of score. *Journal of Second Language Writing*, 22(3), 217-230. https://doi.org/10.1016/j.jslw.2013.02.003
Plakans, L., & Gebril, A. (2017). Exploring the relationship of organization and connection with scores in integrated writing assessment. *Assessing Writing*, 31, 98-112. https://doi.org/10.1016/j.asw.2016.08.005
R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/
沢谷佑輔・鈴木智己 (2016).「英語ライティングにおける結束性と評価の関係性—まとまりのあるライティングを目指した実践的研究—」『北海道英語教育学会紀要』, 15巻, 35-54. https://doi.org/10.24675/helesje.15.0_35
鈴木駿吾 (2023).「第二言語スピーキング能力の機能的達成度は自動判定できるか?—学際的な研究課題の紹介—」『日本音響学会誌』79巻, 177-183. https://doi.org/10.20697/jasj.79.3_177
Therneau, T., & Atkinson, B. (2022). rpart: Recursive partitioning and regression trees. https://CRAN.R-project.org/package=rpart
Tywoniw, R., & Crossley, S. (2019). The effect of cohesive features in integrated and independent L2 writing quality and text classification. *Language Education and Assessment*, 2(3), 110-134. https://doi.org/10.29140/lea.v2n3.151
Tsunemoto, A., & Trofimovich, P. (2024). Coherence and comprehensibility in second language speakers' academic speaking performance. *Studies in Second Language Acquisition*, 1-23. https://doi.org/10.1017/S0272263124000305
保田幸子 (2024).『「書く力」の発達—第二言語習得論と第二言語ライティング論の融合に向けて—』くろしお出版.

資料1：補足データ

本研究で用いたデータ及び分析コードは以下のリンクから入手可能です: <https://osf.io/h5vpr/>

資料2：テキスト一貫性を予測する決定木分析で抽出された結束性指標の記述統計と相関行列

結束性指標	M	SD	1	2	3	4	5	6
1. adjacent sentence overlap content lemmas	.172	.069	—					
2. adjacent sentence overlap function lemmas	.252	.076	.415	—				
3. adjacent two-sentence overlap adjective lemmas (sentence normed)	.195	.200	.254	.371	—			
4. word2vec similarity (two adjacent sentences)	.887	.029	.248	.412	.290	—		
5. sentence linking	.034	.013	-.030	.092	.147	.040	—	
6. adjacent sentence overlap noun and pronoun lemmas	.253	.077	.776	.535	.089	.353	-.089	—

資料3：エッセイの質を予測する決定木分析で抽出された結束性指標の記述統計と相関行列

結束性指標	M	SD	1	2	3	4	5	6
1. binary adjacent two-sentence overlap content lemmas	.799	.147	—					
2. binary adjacent sentence overlap pronoun lemmas	.383	.215	-.010	—				
3. demonstratives	.022	.013	.159	.174	—			
4. positive intentional connectives	.009	.008	.189	.125	-.064	—		
5. temporal connectives	.011	.007	-.099	.255	-.086	.145	—	
6. order	.008	.007	-.204	.029	-.037	-.121	.534	—