第32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

自動英文解析ツールを用いた英作文採点の妥当性検証:Coh-Metrix と Text Inspector の指標に基づいて

研究者:茨城県／筑波大学大学院　在籍　佐々木 大和

《研究助言者:村木 英治》

概要

本研究では，英文解析ツールであるCoh-MetrixとText Inspectorの自動英作文評価への利用可能性に関して2つの調査を行った。まず，調査1では，Coh-MetrixとText Inspectorを用いて，英検が公表している1級から3級の英作文の模範解答例のテキスト分析を行い，テキスト特性を検証した。結果として，受験級が高くなるほど，トピックが難しくなり，幅広い単語や時制・相を用いた難易度が高い英作文を書くように求められていることがわかった。次に，調査2では，Coh-MetrixとText Inspectorを用いて，実際の日本人英語学習者の英作文のテキスト分析を行い，算出された指標の傾向と英語力との関係，英作文得点との関係を調査した。結果として，英作文のテキスト特性に関しては，受験級が上がるにつれて，英作文の馴染みやすさや時制や相の利用，難易度，語の多様性が増加する傾向にあることがわかった。一方，求められる語数が多くなるほど，簡単な馴染みのある文や同じ動詞を繰り返し使うことがわかった。算出された指標と学習者の英語力の関係に関しては，求められるトピックや語数が異なることで，英語力を予測する指標が変わる可能性が示唆された。算出された指標と英作文評価との関係性については，受験級ごとに英作文評価と関係のある指標が異なっていたが，全ての受験級を通して，語の長さや頻度，親密度，多様性のような指標が予測変数に含まれていたことから，語に関する指標が自動英作文評価に影響を与えている可能性があることが示唆された。

1 はじめに

近年，大規模英語テストを入試に利用する学校が増えている。リーディングやリスニングテストに関しては，多肢選択式問題が取り入れられ，採点の妥当性が確保されている。一方，ライティングにおいては，採点前トレーニングを行ったり，熟練した採点者を確保したりしても，多肢選択式問題に比べ，主観性が介在してしまうため，採点者間でブレが生じる可能性がある。また，普段の英語の授業においても，学習者の英語ライティング能力を測定するために，教師が英作文を課す場合が多い。しかし，教師にとって，この英作文を採点・評価する作業というのは，かなりの労力と時間を要する。そこで，近年，英作文の自動評価が注目を浴びてきている。一方，英作文自動評価ツールには有料のものも多く，手を出しにくい場合も多い。

以上の背景を踏まえ，本研究では，近年，読解研究において，ウェブベースの無料の自動英文解析ツールとして活用されているCoh-Metrix（Graesser, McNamara, Louwerse, & Cai, 2004）とText Inspector（Bax, 2012）の英作文自動評価への応用可能性を検証する。このツールは自然言語処理とデータマイニングを利用し，さまざまな指標を用いて，英文の言語的特徴を客観的に評価することが可能である。

2 先行研究

2.1 ライティング

ライティングは，書き手が持つ文章を書くための技能や知識を利用する認知プロセスでもあり，ある特定の目的や読み手を伴う場面において行われる状況内の活動（situated activity）である（Polio & Friedman, 2017）。ライティングは文字を使って，相手に情報などを伝える活動であり，インターネット等で世界中の人々と繋がる可能性が高まる現代社会において，英語で情報を伝えるニーズは高まっている（佐野，2013）。しかし，平成30年に告示された高等学校学習指導要領（文部科学省，2018）において，「書くこと」の言語活動が適切に行われていないことが指摘され，その問題を踏まえ，話すことや書くことによる発信能力の育成を目的とした「論理・表現」という外国語科目が新設された。

「1. はじめに」で述べたように，英語のテストにおいて，英語ライティングの能力を測定するために，生徒に英作文を書かせる場合が多い。また，授業においても，英語ライティング向上のために，英作文課題を課す教員が多い。しかし，その英作文の採点や評価には，かなりの労力と時間がかかり，教員の負担となっている。また，英作文の評価には，評価者の主観が介在する可能性が高く，評価者トレーニングを行うなど妥当性の確保が難しい。そこで，ツールを使った英作文の自動評価を利用する教員やテスト会社が増えてきている。

2.2 英作文の評価法と自動評価

英作文の評価法としては，総合的評価（holistic scoring）と分析的評価（analytic scoring）が挙げられる。前者は，評価者が英作文全体を読んで，点数を1つ与える評価法であり，後者は，初めにいくつか観点を定め，その観点ごとに点数を与える評価法である。英作文の自動評価には，総合的評価が用いられることが多い（小林・金丸，2012; Shermis & Burstein, 2003）。

英作文の自動採点（automated essay evaluation）とは，コンピュータ・プログラムを通して，作文を評価したり，採点したりすることである。近年，Educating Testing Serviceの e-rater Scoring Engineや日本英語検定協会の Writing Tutorなど，様々な機関が英作文の自動採点にコンピュータ・プログラムを利用している。しかし，これらの自動評価は教員や学習者がいつでも使えるものではなく，有料の場合も多い。そこで，本研究では，無料のウェブベースの英文解析ツールであるCoh-MetrixとText Inspectorの英作文自動評価への利用可能性を検証する。

2.3 英文解析ツール：Coh-MetrixとText Inspector

Coh-MetrixやText Inspectorといった英文解析ツールは，さまざまな指標を用いて，英文の言語的特徴を客観的に評価することができるものである。Coh-Metrixは英文の結束性（文章中の要素の結びつき）を評価できる点が特徴的であり，複数の関連する指標を用いて，Referential Cohesion, Verb Cohesion, Deep Cohesionといった英文の結束性の指標を算出することが可能である。特に読解研究において，幅広く利用されてきている（McNamara, Graesser, McCarthy, & Cai, 2014）。

それに対し，Text Inspectorは英文に対して，ヨーロッパ言語共通参照枠（Common European Framework of Reference for Languages; CEFR）に基づいたScorecardを算出することができる。このScorecardでは，Percentage（100%が英語母語話者レベルを表す），Number of Metrics Used（参考にした指標の数），CEFR Levelを確認することができる。また，このScorecardは，リーディング，ライティング，リスニングといったテキストの種類によって，参考にする指標を変えて算出されている。

以上のような英文解析ツールをライティングに応用した研究がなされている。Coh-Metrixを用いた研究に関しては，特に，Coh-Metrixで算出された指標と評価者による英作文評価の関係性を検証している研究が多い（e.g., Crossley & McNamara, 2010; 2012; McNamara, Crossley, & McCarthy, 2010; Guo, Crossley, &

McNamara, 2013; Zedulius, Millis, & Schooler, 2019)。母語話者の英作文を対象とした研究では，英作文の語彙の洗練さ（lexical sophistication）や統語的複雑さ（syntactic complexity）の指標が採点者による評価に影響を与え，Coh-Metrix の特徴である結束性に関わる指標はあまり影響を与えていない（McNamara et al., 2010）という結果や，評価者による結束性・一貫性に関わる評価にも，結束性に関わる指標は関連していない（Crossley & McNamara, 2010）という結果が得られている。また，英作文の創造性に着目すると，英作文の馴染みやすさ（narrativity）や語や内容の重複（referential cohesion），語の具象性（word concreteness）などの言語的特徴が採点者による評価に影響を与えていることが示唆されている（Zedulius et al., 2019）。一方，外国語としての英語を学ぶ学習者の英作文を対象とした研究では，英作文の語の多様性（lexical diversity），語の頻度（word frequency），語の意味内容（word meaningfulness），相の繰り返し（aspect reputation），語の親密度（word familiarity）が採点者による評価を予測するという結果（Crossley & McNamara, 2012）や，タスクの種類が異なっていても，英作文の長さなどの言語的特徴が採点者による評価を予測するという結果（Guo et al., 2013），語彙に関する指標が採点者による評価と関係するという結果（Aryadoust & Liu, 2015）が得られている。

Coh-Metrix で算出された指標と書き手の熟達度（i.e., 英作文の質）の関係に関しても数はあまり多くはないが，研究が行われている（小林・金丸，2012; Latifi & Gierl, 2020）。母語話者を対象とした研究では，複数のトピックの英作文間で比較しており，語や内容の重複（referential cohesion）と語の多様性（lexical diversity）に関する特性が英作文の質の評価に最も影響を与えているが，英作文のレベルやトピック，書き手の熟達度，利用したルーブリックなどにより影響を与える指標が異なる（Latifi & Gierl, 2020）という結果が得られている。また，外国語として英語を学ぶ学習者を対象とした研究に関して，日本人英語学習者を対象に行った小林・金丸（2012）では，e-rater によって評価された書き手の熟達度を推定するにあたって，異なり語数の値が最も熟達度の推定に寄与していることが示された。

Text Inspector を用いた研究については，数が限られている。Bax, Nakatsuhara, and Waller（2019）では，文章の繋がりを作る談話標識といった metadiscourse marker に着目し，学習者を対象に研究を行った。結果として，熟達度の高い書き手は熟達度の低い書き手よりも metadiscourse marker を使わないが，幅広く使用するということが示された。

2.4 本研究の概要と目的

以上のような先行研究を踏まえ，本研究では，外国語として英語を学ぶ日本人を対象にし，Coh-Metrix や Text Inspector といった自動英文解析ツールを用いた英作文自動評価の妥当性を検証する。調査1では，英検が公表している英作文問題の模範解答例のテキスト特性を Coh-Metrix と Text Inspector を用いて検証する。次に調査2では，日本人英語学習者の英作文を対象に，Coh-Metrix と Text Inspector で算出されたテキスト特性の傾向，英語力，英作文得点との関係を検証する。

3 調査1

3.1 目的

調査1では，英検が公表している英作文問題の模範解答例のテキスト特性を Coh-Metrix と Text Inspector で分析することで，各級において求められている英作文の特性を明らかにすることを目的とする。具体的には，1級から3級までの英作文問題の模範解答例を対象とし，異なる受験級において，どのようなテキスト特性が見られるかを明らかにする。検証課題（Research Question: RQ）は以下のとおりである：

RQ1-1 Coh-Metrix や Text Inspector により算出される指標について，英検1級から3級の英作文模範解答例の特性はどのようになっているか。

3.2 方法

3.2.1 マテリアルの収集

英作文問題の模範解答例のテキスト特性を明らかにするため，英検1級から3級までの模範解答例を収集した。具体的には，英検3級に英作文問題が導入された2017年度第1回から2019年度第3回までの英作文問題の模範解答例を各受験級（1級，準1級，2級，準2級，3級）からそれぞれ9テキスト，計45テキスト収集した。

3.2.2 分析指標

本調査で使用した指標はCoh-Metrixから9つ（Readability, Narrativity, Syntactic Simplicity, Word Concertedness, Referential Cohesion, Deep Cohesion, Verb Cohesion, Connectivity, Temporality），Text Inspectorから4つ（Percentage, Gunning Fog, Metadiscourse [Type], Lexical Diversity）である。

Coh-Metrixにおける指標に詳しい説明は以下のとおりである：

Readability：
テキスト中の内容語の重複や統語的類似性，語の頻度を考慮した読みやすさの指標。
Narrativity：
テキストがどれだけ日常的であるか，つまり馴染み深いかを示す指標。
Syntactic Simplicity：
テキスト中の文がどれだけ少ない語で，かつ簡単な馴染みのある統語構造が使用されているかを示す指標。
Word Concreteness：
テキスト中の内容語に具体的な語がどの程度含まれているかを示す指標。
Referential Cohesion：
文間やテキスト全体で重複する語や考えがどの程度含まれているかを示す指標。
Deep Cohesion：
テキスト中に因果的もしくは論理的な接続語がどの程度含まれているかを示す指標。
Verb Cohesion：
テキスト中に動詞の重複がどの程度含まれているかを示す指標。
Connectivity：
テキスト中に逆説・付加・比較の接続語がどの程度含まれているかを示す指標。
Temporality：
テキスト中に時間的な手がかりがどの程度含まれているか，時制や相がどの程度一致しているかを示す指標。

また，Text Inspectorにおける指標の説明は以下のとおりである：

Percentage：
100%が，母語話者が書く高いレベルのアカデミックな英作文を示す。
Gunning Fog：
1文あたりの平均単語数と長い綴りの単語を考慮した英作文の読みやすさの指標。
Metadiscourse（type）：
接続詞といった談話標識がどの程度使用されているかを示す指標。
Lexical Diversity：
語の多様性，つまりどの程度幅広い語が使用されているかを示す指標。

3.3 手順

テキスト分析ツールのCoh-Metrix（http://www.cohmetrix.com/）とText Inspector（http://www.textinspector.com）を用いて，英検が公表している英作文問題の模範解答例を分析した。

3.4 結果と考察

表1はCoh-Metrixで算出した指標を受験級ごとに集計した記述統計である。結果より，級が上がるにつれて，Narrativityの値が減少傾向にあることがわかる。級が上がるにつれて，テキストの馴染みやすさが減少することから，受験者にとって難しいトピックが扱われていることがわかり，特に準1級と2級間の差は大きい。また，Temporalityにおいても，同様の減少傾向が見られる。級が上がるにつれて，時間を示すような語句を使わなくなったり，英作文に含まれる時制（現在形や過去形）や相（完了形や進行形）が多様化したりする傾向にあることがわかる。さらに，

Readabilityも準2級から1級にかけて減少傾向が見られ，難易度の高い英文を書くように求められていることがわかる。3級の値が準2級とあまり変わらないことに関しては，語数が少ないため，内容語の重複が少ないことに起因すると考えられる。

一方，統語的な簡単さを示すSyntactic Simplicityや語の具体性を示すWord Concretenessは，級が上がるにつれて減少すると予測していたが，一貫した結果は得られなかった。Syntactic Simplicityに関しては，3級が最も値が低いが，これは語数が少ないために値が低く出てしまった可能性がある。準1級と2級は類似した値が算出され，1級で値が減少しているため，文中の語数が増え，また難しい統語構造を使うことを求められていることがわかる。Word Concretenessにおいては，3級が最も高く，準2級，2級，準1級の値は似ており，1級で減少している。1級では，難しいトピックが扱われる傾向にあることから，英作文で使用する語の抽象度が上がると考えられる。

Coh-Metrixの特徴である結束性・一貫性の値に関しては，級ごとで一貫した結果は得られていないが，どの指標においても1級の値が最も小さい。求められる語数が上がるにつれて，相対的に結束性や一貫性を示す接続語が少なくなり，英作文の結束性・一貫性が低くなる可能性がある。

■表1: Coh-Metrixによる英作文模範解答の級ごとの特徴

	1級 (n=9)		準1級 (n=9)		2級 (n=9)		準2級 (n=9)		3級 (n=9)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Readability	11.37	2.33	16.60	3.05	26.55	5.42	34.85	3.74	32.82	8.40
Narrativity	14.49	5.98	28.90	15.94	68.46	15.11	86.13	5.98	88.86	22.04
Syntactic Simplicity	57.42	18.33	65.12	18.86	61.43	15.99	88.87	9.89	53.50	28.72
Word Concreteness	33.45	19.28	50.79	20.11	52.99	27.60	50.26	29.25	82.47	17.64
Referential Cohesion	19.82	18.93	30.59	23.94	70.89	26.03	66.59	30.27	83.78	25.03
Deep Cohesion	73.57	16.32	92.61	8.53	81.79	20.77	98.15	2.91	80.33	28.81
Verb Cohesion	20.45	14.80	36.24	23.65	83.41	13.33	68.94	33.51	49.35	28.59
Connectivity	11.06	8.37	1.61	1.73	36.98	33.52	9.27	13.05	28.23	23.95
Temporality	17.23	18.80	22.12	26.03	40.01	33.96	44.38	26.69	64.33	37.22

次に，表2にText Inspectorで算出した指標を受験級ごとに集計した記述統計を示す。表より，級が上がるにつれて，Gunning Fog，Lexical Diversityの値が上昇していることがわかる。Gunning Fogに関しては，英作文を読む際に，読み手がどの程度の能力が必要かを示す指標である。Coh-Metrixの結果と同様，級が上がるにつれて，難しい英作文を書くように求められていることがわかる。また，Lexical Diversityに関しても，級が上がるにつれて，様々な種類の語を使うように求められていることがわかり，2級と準1級の間の値の差が大きい。

Percentageは母語話者らしいアカデミックな英作文が書けているかを表す指標であり，Text Inspectorにおいて，英作文の得点を表す。準2級から1級にかけて，得点が上昇傾向にあり，母語話者らしい英作文を書くように求められていることがわかる。また，他の指標と同様，準1級と2級の間に大きな値の差がある。

一方，Metadiscourseに関しては，減少傾向にある。これは様々な談話標識の数を表すものであり，接続詞など結束性・一貫性に関わる指標である。Coh-Metrixでの指標において，結束性・一貫性に関しては一貫した結果は得られず，1級が最も値が低くなるという結果であったが，Text Inspectorでは，一貫して減少していることがわかった。

■表2: Text Inspectorによる英作文模範解答の級ごとの特徴

	1級 (n=9)		準1級 (n=9)		2級 (n=9)		準2級 (n=9)		3級 (n=9)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Percentage	77.69	4.58	70.69	3.35	48.20	6.60	37.09	7.85	46.72	9.42
Gunning Fog	16.76	1.10	15.06	1.54	9.45	0.92	7.60	1.74	7.43	2.59
Metadiscourse	9.03	2.08	10.71	2.78	14.65	3.03	17.05	4.52	20.79	3.93
Lexical Diversity	111.83	23.55	102.28	20.44	68.78	15.96	56.17	7.92	n/a	n/a

3.5 調査1のまとめ

調査1では，英作文問題の模範解答例のテキスト特性を明らかにするために，1級から3級までの英作文模範解答例をCoh-MetrixとText Inspectorを用いて分析を行った。結果より，受験級が高くなるほど，トピックが難しくなり，幅広い単語や時制・相を用いたテキスト難易度が高い英作文を書くように求められていることが示唆された。

調査2では，実際に日本人英語学習者が書いた英作文を対象にし，Coh-MetrixやText Inspectorを用いてテキスト分析を行い，学習者の英語力とどのような関係があるのかを調査した。

4 調査2

4.1 目的

調査2では，日本人英語学習者の英作文をCoh-MetrixとText Inspectorを用いて分析し，算出された指標の特徴やその指標と学習者の英語力にどのような関係があるのかを調査する。また，Text Inspectorで算出される英作文の得点（Percentage）が他のCoh-MetrixやText Inspectorの指標とどのような関係があるのかを検証する。検証課題（RQs）は以下のとおりである：

RQ2-1 学習者の英作文における，Coh-MetrixとText Inspectorで算出された指標は級ごとにどのような傾向があるのか。

RQ2-2 学習者の英作文におけるCoh-MetrixとText Inspectorで算出された指標は学習者の英語力とどのような関係があるのか。また，級ごとで異なるのか。

RQ2-3 Text Inspectorで算出された英作文の得点（Percentage）は他のCoh-MetrixやText Inspectorの指標とどのような関係があるのか。また級ごとで異なるのか。

4.2 方法

4.2.1 協力者

私立大学に通う日本人英語学習者30名が調査に参加した。協力者は全員大学1年生であり，専攻は人文系であった。協力者は全員日本語母語話者で，少なくとも6年以上日本の教育機関で英語を学んでいた。自己申告の英語資格のアンケートによると，協力者は英検3級から2級程度であり，CEFRのA1からB1レベルであると推察される。実験は3日間に分けて行われたため，全ての実験を完遂した22名を分析対象とした。

4.2.2 マテリアル

(1) 英作文課題

英検の過去問題より，3級，準2級，2級から1つずつトピックを選定した。選定したトピックは次の表3のとおりである。受験級に関しては，大学1年生を高校卒業程度とし，2級までとした。指定語数はそれぞれの級にならい，3級のトピックは25語から35語，準2級は50語から60語，2級は80語から100語とした。指示は実際の英検の出題

形式に則り,「以下の質問について, あなたの考えとその理由を2つ英文で書きなさい」とし, 2級のトピックではポイントを3つ(Convenience, Cost, The environment)与えた。解答は紙ベースで行われた。協力者の英作文の語数の平均は, 3級では35.45語(*SD* = 10.13), 準2級では39.73語(*SD* = 17.38), 2級では61.32語(*SD* = 26.05)であった。

■表3: 英作文の受験級とトピック

受験級	英作文のトピック
3級	What day of the week do you like the best?
準2級	Do you think it is important for children to play sports?
2級	It is often said that people today use too much electricity. Do you agree with this opinion?

(2) 英語力測定課題

協力者の英語力を測定するために, 英検IBAを用いた。リーディングとリスニングの2技能を測定するものを使用し, 2級から3級を測定することが可能なテストBを使用した。形式は英検と同様のもので, すべて多肢選択式問題であった。リーディング問題は語彙・文法問題20問, 会話文問題5問, 長文問題10問で構成されており, リスニング問題は会話文問題15問, 英文問題15問で構成されている。

4.2.3 手順

調査は3日間に分けて一斉に実施された。全体の所要時間は95分程度であった。1日目と2日目に協力者は英作文課題を行った。1日目に英検3級と準2級のトピック, 2日目に英検2級のトピックに関する英作文を書き, それぞれ25分間で行った。この英作文課題の際, 協力者は実際の試験を想定し, 辞書等で単語を調べることはできなかった。

3日目に英検IBAテストが実施された。リーディングが25分, リスニングが20分の計45分で行われた。この際, 英作文課題と同様, 辞書等の使用は禁止されていた。

4.2.4 採点・分析

英作文課題に関しては, 調査1と同様, ウェブベースのテキスト分析ツールであるCoh-MetrixとText Inspectorを用いて, 学習者が取り組んだ各級の英作文を分析した。Coh-Metrixでは, 綴りの誤りを分析することができないので, 綴りの誤りに関しては, 調査者が書き起こす際に修正した。また, 英作文問題の模範解答例では, 3級, 準2級, 2級と1段落で構成されているため, 協力者の英作文も1段落に統制を行った。

英語力測定課題に関しては, 日本英語検定協会で採点がされ, 得点は成績表のCSEスコアを参考にした。満点スコアは1300点である。

RQ2-1に関して, 研究1と同様の分析指標を用いて, 記述統計を算出し, 受験級ごとの傾向を確認した。

RQ2-2に関して, Coh-MetrixやText Inspectorで算出される指標が膨大であるため, Crossley and McNamara(2012)を参考に, 英語力と算出された指標で各級ごとの英作文で相関分析を行い, 相関係数が大きい指標を選定し, 変数の絞り込みを行った。絞り込まれた変数を独立変数, 英検IBAテストで測定された英語力を従属変数に重回帰分析を行った。

RQ2-3に関しても, RQ2-1と同様の手順で変数の絞り込みを行った。絞り込まれた変数を独立変数, Text Inspectorで算出された英作文の得点(Percentage)の値を従属変数に重回帰分析を行った。

4.3 結果と考察

(1) 英語力測定課題

表4は英検IBAテストの記述統計である。

■表4: 英語力測定課題の記述統計

	M	95%CI	*SD*	*Min*	*Max*
リーディング	422.81	[410.02, 435.61]	32.35	338	497
リスニング	383.15	[368.78, 397.52]	36.33	316	484
総合点	805.96	[783.24, 828.69]	57.44	668	938

(2) 英作文課題

英作文課題において，Coh-Metrixで算出した指標を受験級ごとに集計した記述統計を表5に示す。調査1の結果を基に，考察をしていく。調査1の結果と同様，Narrativityは級が上がるにつれて，値が減少していく傾向が見られた。この結果により，級が上がるにつれて，質問の内容が難しくなり，馴染みがあまりないトピックになることが示唆された。また，Temporalityに関しても同様の減少傾向が見られたが，模範解答の3級の値とほぼ同じであり，同じような時制や相を使って，表現していることが示された。

一方，調査1では見られなかった値の上昇傾向がSyntactic SimplicityとVerb Cohesionに見られた。級が上がるにつれて，求められる語数が多くなり，複雑な文構造を使用することができない学習者が似たような馴染みのある文構造を用いて英文を書いている可能性がある。また，同様の傾向がVerb Cohesionにも見られ，求められる語数が多くなるにつれて，同じような動詞を繰り返し使用するようになるためであると考えられる。

■表5: Coh-Metrixによる協力者の英作文の級ごとの特徴

	2級		準2級		3級	
	M	*SD*	*M*	*SD*	*M*	*SD*
Readability	32.50	8.22	38.50	10.99	35.95	12.04
Narrativity	52.70	22.00	79.51	19.43	81.46	18.70
Syntactic Simplicity	93.98	8.11	89.41	14.82	88.90	25.67
Word Concreteness	30.04	27.44	67.19	31.39	43.72	31.95
Referential Cohesion	50.84	33.12	82.23	28.63	72.58	33.37
Deep Cohesion	79.32	31.06	95.22	20.77	82.90	22.69
Verb Cohesion	77.22	23.09	72.16	30.31	49.49	34.04
Connectivity	5.81	20.14	42.47	35.11	21.29	26.76
Temporality	62.94	31.22	65.10	33.13	68.39	29.76

次に，英作文課題において，Text Inspectorで算出した指標を受験級ごとに集計した記述統計を表6に示す。Percentageにおいて，調査1の模範解答例では，3級が2級と同程度であり，準2級から級が上がるにつれて値が大きくなるという傾向が見られていたが，本研究の結果では，級が上がるにつれて，値が徐々に上がっているという傾向が見られた。Gunning Fogに関しても，上昇傾向が見られており，級が上がるにつれて，文が長くなったり，2音節以上の単語を使うようになったりしていることが示唆される。Lexical Diversityに関しては，模範解答例で示されたほど数値は高くないが，こちらも増加傾向にあることがわかる。

一方，Metadiscourseに関しては，調査1で見られたような減少傾向は見られず，一貫した結果は得られなかった。

■表6: Text Inspectorによる協力者の英作文の級ごとの特徴

	2級		準2級		3級	
	M	*SD*	*M*	*SD*	*M*	*SD*
Percentage	45.10	4.38	33.21	8.03	28.85	9.40
Gunning Fog	8.96	1.45	6.12	1.71	5.10	1.60
Metadiscourse	18.67	5.09	23.51	12.11	21.32	3.60
Lexical Diversity	37.03	28.80	12.23	17.22	4.23	10.93

(3) Coh Metrix と Text Inspector で算出された指標と学習者の英語力の関係

分析にあたって, Coh-Metrix と Text Inspector で算出される指標が膨大であるため, 算出された受験級ごとの英作文の指標と協力者の英語力でスピアマンの相関分析を行った。協力者の英語力と有意な相関関係が見られたのは, 3級の英作文において, Coh-Metrix では, 三人称単数代名詞の使用(rs = .51, p = .02), Text Inspector では, 異なり語100語ごとの語の頻度(rs = .47, p = 0.3), 準2級の英作文において, Coh-Metrix では, 動詞の重複(p = -.45, p = .04)と 一人称複数代名詞の使用(rs = -.424, p = .05), Text Inspector では, なし, 2級の英作文において, Coh-Metrix では, 意図を表す語の使用(rs = .50, p = .02) と動名詞の利用(rs = .46, p = .03), Text Inspector では, なしであった。

この結果を基に, それぞれの級において, 協力者の英語力(英検IBAの総合点)を従属変数に, 有意な相関関係が見られた変数を独立変数に強制投入法による重回帰分析を行った。分析を行う前に, 多重共線性の確認を行ったところ, 相関係数が.80を上回る変数はなかった。重回帰分析の結果, それぞれの級に関して, 有意な回帰モデルが得られた, $F(1, 20) = 7.05$, p = 0.2; $F(2, 19) = 4.76$, p = 0.2; $F(2, 19) = 4.33$, p = 0.3; $F(1, 20) = 4.41$, p = .05。それぞれの級の結果が表7から表10に示されている。

結果より, 3級においては, Coh-Metrixにおける三人称単数代名詞の使用, 英作文で使用される語の頻度が学習者の英語力をそれぞれ26%, 18%予測することができ, 準2級においては, 動詞の重複と一人称単数代名詞の使用で学習者の英語力を33%, 2級においては, 意図を表す語と動名詞の使用で学習者の英語力を31%予測することができる。ちなみに準2級の英作文においては, βの値が負であるため, 動詞の重複や一人称単数代名詞が少ない協力者ほど, 英語力が高いことがわかる。

まとめると, それぞれの受験級において, 固有の予測変数は存在するが, 共通した予測変数は見つからなかった。したがって, 英作文のトピック, もしくは求められる語数によって, 英語力を予測する指標は異なる可能性が示唆された。

(4) Coh-Metrix で算出された指標と Text Inspector で算出された英作文得点の関係

(3) と同様, 重回帰分析で使用する変数を絞るため, それぞれの受験級ごとの英作文において, Text Inspector で算出された英作文得点 (Percentage) と Coh-Metrix で算出された指標でスピアマンの相関分析を行った。その後, 変数間の相関関係を確認し, .80を超える相関係数の変数は, 英作文得点との相関係数が大きい変数を採用した。その結果, 重回帰分析の独立変数に選ばれたのは, 3級の英作文においては, 語(文字数)の長さ(rs = .47, p = .03), Connectivity(rs = .50, p = .02), 文間の名詞の重複(rs = -.43, p = .04), 意図を表す動詞の利用(rs = .-.46, p = .03), 内容語の語の頻度(rs= -.46, p = .03), 準2級の英作文においては, 語(文字数)の長さ(rs = .51, p = .02), 内容語の親密度(rs = -.62., p = .002), 2級の英作文において, 語数(rs = -.70, p < .001), 語の多様性(rs = .53, p = .01)だった。

この結果を基に, それぞれの級において, 協力者の英作文得点(Percentage)を従属変数に, 採用された変数を独立変数に強制投入法による重

■表7: 3級英作文(Coh-Metrix)における強制投入法による重回帰分析の結果

	英語力					
	B	*SE B*	95%CI	β	*t*	*p*
三人称単数代名詞の使用	8.52	3.21	[1.83, 15.21]	0.51	2.65	0.02

CI = confidence interval, $R = .51$, $R^2 = .26$ ($p = .02$)

■表8: 準2級英作文 (Coh-Metrix) における強制投入法による重回帰分析の結果

	英語力					
	B	*SE B*	95%CI	β	*t*	*p*
動詞の重複	-73.38	38.92	[-154.85, 8.09]	-0.35	-1.89	0.08
一人称単数代名詞の使用	-1.91	0.80	[-3.57. -0.25]	-0.45	-2.40	0.03

CI = confidence interval, $R = .58$, $R^2 = .33$ ($p = .02$)

■表9: 2級英作文(Coh-Metrix)における強制投入法による重回帰分析の結果

	英語力					
	B	*SE B*	95%CI	β	*t*	*p*
意図を表す語の使用	0.82	0.43	[-0.09, 1.72]	0.37	1.89	0.07
動名詞の使用	1.43	0.80	[-0.25, 3.12]	0.35	1.78	0.09

CI = confidence interval, $R = .59$, $R^2 = .31$ ($p = .03$)

■表10: 3級英作文(Text Inspector)における強制投入法による重回帰分析の結果

	英語力					
	B	*SE B*	95%CI	β	*t*	*p*
語の頻度	0.00	0.00	[0.00, 0.01]	0.43	2.10	0.05

CI = confidence interval, $R = .43$, $R^2 = .18$ ($p = .03$)

回帰分析を行った。その結果, 3級, 準2級, 2級の英作文に関して, 有意な回帰モデルが得られた, $F(5, 16) = 5.87$, $p = 0.03$; $F(2, 19) = 12.94$, $p < 0.01$; $F(2, 19) = 3.93$, $p = 0.4$。それぞれの級の結果は表11から表13に示されている。

表より, 有意な予測3級英作文においては, 語(文字数)の長さが長く, 逆説・付加・比較の接続語, 同じ名詞の繰り返し, 意図を表す動詞が少なく, 内容語の語の頻度が低ければ, 英作文の得点が高くなり, 準2級英作文においては, 語(文字数)の長さが長く, 内容語の親密度が低ければ, 英作文の得点が高くなり, 2級の英作文においては, 語数が少なく, 語の多様性が低ければ, 英作文の得点が高くなることが示された。

それぞれの級に共通する全く同じ予測変数は見つからなかった。これは英作文のレベルやトピック, 書き手の熟達度, 利用したルーブリックなどにより英作文の得点に影響を与える指標が異なる(Latifi & Gierl, 2020)という先行研究の結果と一致するものである。一方, 英検3級における語の頻度, 準2級における語の親密度, 2級における語の多様性のような語に関するCoh-Metrixの指標がText Inspectorで算出される英作文の得点に影響を与えていることが示唆された。

■表11: 3級英作文における強制投入法による重回帰分析の結果

	英作文得点					
	B	*SE B*	95%CI	β	*t*	*p*
語（文字数）の長さ	19.36	7.78	[2.82, 35.89]	0.40	2.48	0.03
Connectivity	-0.03	0.09	[-0.15, 0.09]	-0.09	-0.52	0.61
文間の名詞の重複	-10.29	6.44	[-23.94, 3.36]	-0.25	-1.60	0.13
意図を表す動詞の利用	-0.13	0.06	[-0.25, -0.01]	-0.37	-2.30	0.04
内容語の語の頻度	-20.159	7.26	[-35.54, -4.77]	-0.44	-2.78	0.01

CI = confidence interval, $R = .80$, $R^2 = .46$ ($p = .003$)

■表12: 準2級英作文における強制投入法による重回帰分析の結果

	英作文得点					
	B	*SE B*	95%CI	β	*t*	*p*
語（文字数）の長さ	7.14	2.35	[2.22, 12,07]	0.45	3.03	0.01
内容語の親密度	-0.92	0.22	[-1.39, -0.46]	-0.63	-4.19	< 0.01

CI = confidence interval, $R = .76$, $R^2 = .58$ ($p = .001$)

■表13: 2級英作文における強制投入法による重回帰分析の結果

	英作文得点					
	B	*SE B*	95%CI	β	*t*	*p*
語数	-0.10	0.06	[-0.22, 0.03]	-0.58	-1.64	0.12
語の多様性	-1.36	10.97	[-24.31, 21.60]	-0.04	-0.90	0.90

CI = confidence interval, $R = .54$, $R^2 = .29$ ($p = .04$)

4.4 調査2のまとめ

調査2では，Coh-Metrix や Text Inspector で算出された指標の英作文評価への利用可能性を検証するために，日本人英語学習者の英作文に対する Coh-Metrix と Text Inspector の指標を算出し，その指標の英語力の予測率，また，Text Inspector で算出された英作文得点への Coh-Metrix の指標の予測率を調査した。結果として，(1) 日本人英語学習者の英作文を英文解析ツールで，分析したところ，受験級が上がるにつれて，英作文の馴染みやすさや時制や相の利用，難易度，語の多様性が上昇傾向にあり，反対に，統語の簡単さや動詞の一貫性には減少傾向が見られた。一方，(2) 英作文の受験級（求められるトピックや語数）によって日本人英語学習者の英語力を予測する指標が異なる可能性，また，(3) 英文解析ツールによって算出された英作文得点は，受験級ごと予測変数は異なっていたが，特に，語に関する指標（e.g.，長さ，頻度，親密度，多様性）の影響が大きい可能性が示唆された。

5 結論と今後の課題

本研究では，無料で利用可能なウェブベースの英文解析ツールである Coh-Metrix と Text Inspector の自動英作文評価への利用可能性に関して2つ調査を行った。まず，調査1では，Coh-Metrix と Text Inspector を用いて，英検が公表している1級から3級の英作文の模範解答例のテキスト分析を行い，テキスト特性を検証した。結果として，受験級が高くなるほど，トピックが

難しくなり，幅広い単語や時制・相を用いた難易度が高い英作文を書くように求められていることがわかった。次に，調査2では，Coh-MetrixとText Inspectorを用いて，実際の日本人英語学習者の英作文のテキスト分析を行い，算出された指標の傾向と英語力との関係，英作文得点との関係を調査した。結果として，英作文のテキスト特性に関しては，受験級が上がるにつれて，英作文の馴染みやすさや時制や相の利用，難易度，語の多様性が増加する傾向にあることがわかり，日本人英語学習者は，級が上がるにつれて内容的にも形式的にも難しい英作文を書くことが示唆された。一方，統語の簡単さや動詞の一貫性は減少傾向にあり，求められる語数が多くなるほど，簡単な馴染みのある文や同じ動詞を繰り返し使うことがわかった。算出された指標と学習者の英語力の関係に関しては，英作文の受験級によって，英語力と関係のある指標が異なることがわかった。求められるトピックや語数が異なることで，英語力を予測する指標が変わる可能性が示唆された。算出された指標と英作文評価との関係性については，英語力との関係性と同様，受験級ごとに英作文評価と関係のある指標が異なっていた。しかし，全ての受験級を通して，長さや頻度，親密度，多様性のような語に関する指標が予測変数に含まれていたことから，語に関する指標が自動英作文評価に影響を与えている可能性があることが示唆された。

以上の本研究の結果より得られた教育的示唆としては，以下の点が挙げられる。まず，調査1の結果より，受験級が上がるにつれて，馴染みやすさ，時制・相の利用，語の多様性に一貫した傾向が見られることがわかった。これより，教師が生徒の英作文を評価する際に，語の多様性や時制・相の利用に関して，客観的な指標を用いて，フィードバックすることが可能である。それぞれの受験級で求められている客観的な数値がわかれば，その指標に基づいて，具体的なフィードバックをすることが可能である。例えば，語の多様性に関しては，生徒の英作文に関して，算出された数値が受験級で求められている数値より低ければ，同じ語を使わずに，パラフレーズして書くように指導することが可能であり，リライトした際も，前回の英作文と数値を比較して，どの程度増加したのか，何が足りないのかを具体的にアドバイスすることが可能である。

また，調査2の結果より，英作文で求められるトピックや語数により，英語力を予測する指標が異なることが示唆された。これより，英文解析ツールを用いて，英作文を評価し，英語力を推定する際は，語数やトピックごとにまず傾向を掴むことが必要である。学習者はトピックや質問に含まれる語や文型を繰り返し使用する可能性がある。したがって，生徒の英語力を推定する際には，トピックや質問に合わせた指標の利用が必要である。さらに，調査2の結果から，英文解析ツールを用いた英作文の自動評価には，語に関する指標が有効である可能性が示唆された。Coh-MetrixやText Inspectorは大規模なコーパスを用いて，語に関する指標を算出しており，信頼性も高い。したがって，生徒の英作文を評価する際には，頻度や多様性といった語に関する指標を評価やフィードバックに有効利用できると考えられる。

本研究は，ウェブベースで利用可能で，読解研究に利用されているCoh-MetrixとText Inspectorを用いた英作文自動評価の妥当性を検証することを目的に行われた。妥当性を十分に検証することはできなかったが，Coh-MetrixやText Inspectorを用いた英作文自動評価の可能性を中心に議論を進めた。一方，本研究には以下のような限界点がある。

まず，調査1では，マテリアルとして英検3級に英作文問題が導入された2017年度から2019年度の英作文解答例を使用した。英作文の解答例は各回に1つしか公表されないため，長文読解テキスト等に比べ，分析対象とした英作文の数が少なくなってしまった。さらに，調査2においても，調査を3日間に分けて行ったため，協力者に比べ，分析対象とした英作文の数が少なくなってしまった。したがって，今後の研究では，より幅広いマテリアルを分析対象とし，また，協力者の数を増やしたり，協力者の熟達度の幅を広げたりする必要がある。

さらに，各級ごとのトピックが英作文に影響した可能性がある。先述したように，書き手はトピックや質問に含まれている単語や文法を使用する傾向が高い。したがって，今後の研究では，同じトピックで語数が異なる英作文課題や幅広

いトピックで同じ語数の英作文課題を使用したり，結果を一般化する手立てを考えたりする必要がある。

以上のような限界点を考慮し，研究を進めていくことで，英作文自動評価の妥当性についてより説得力のある示唆を出すことができると考えられる。

謝辞

本研究の実施，発表にあたりまして，公益財団法人日本英語検定協会と関係者の皆様，ならびに選考委員の先生方からの支援をいただき，心より御礼申し上げます。特に，助言者である村木英治先生には，研究の実施および報告書の執筆にあたりご指導いただきましたこと，深く感謝申し上げます。また，筑波大学の卯城祐司先生をはじめ，研究室の先輩・同輩・後輩の皆さまには，本研究の立案から実施，報告書の執筆にあたりまして，ご助言いただきました。心より感謝申し上げます。最後に，本調査に協力してくださった協力者の皆さまに深く御礼申し上げます。

参考文献（＊は引用文献）

＊ Aryadoust, V. & Liu, S. (2015). Predicting EFL writing ability from levels of mental representation measured by Coh-Metrix: A structural equation modeling study. *Assessing Writing, 24,* 35-58.
＊ Bax, S. (2012). Text Inspector, online text analysis tool.
＊ Bax, S., Nakatsuhara, F., & Waller, D. (2019). Researching L2 writers' use of metadiscourse markers at intermediate and advanced levels. *System, 83,* 79-95.
＊ Crossley, S. A., & McNamara, D. S. (2010). Cohesion, coherence, and expert evaluations of writing proficiency. In Catrambone, R. and Ohlsson, S. (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp.984-989). Austin, TX.
＊ Crossley, S. A., & McNamara, D. S. (2011). Understanding expert ratings of essay quality: Coh-Metrix analyses of first and second language writing. *International Journal of Continuing Engineering Education and Life Long Learning, 21,* 170-191.
＊ 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,* 115-135.
＊ Graesser, A. C., McNamara, D.S., Louwerse, M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers, 36,* 193-202.
＊ Guo, L., Crossley, S. A., & McNamara, D. S. (2013). Predicting human judgements of essay quality in both integrated and independent second language writing samples: A comparison study. *Assessing Writing, 18,* 218-238.
＊ 小林雄一郎・金丸敏幸．(2012).「Coh-Metrix とパターン認識を用いた課題英作文の自動評価」『じんもんこん2012論文集』259-266.
＊ Latifi S. & Gierl, M. (2020). Automated scoring of junior and senior high essays using Coh-Metrix features: Implications for large-scale language testing. *Language Testing.*
＊ McNamara, D.S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication, 27,* 57-86.
＊ McNamara D. S., Graesser, A. C., McCarthy P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix.* Cambridge University Press.
＊ 文部科学省 (2018).『学習指導要領（平成30年告示）解説 外国語編 英語編』東京：開隆堂出版．
＊ Polio, C. & Friedman, D. A. (2017). *Understanding, evaluating, and conducting second language writing research.* New York: Routledge.
＊ 佐野富士子 (2013).「ライティング」In JACET（大学英語教育学会）SLA 研究会（編著),『第二言語習得と英語科教育法』(pp. 248-261). 東京：開拓社．
＊ Shermis, M. D., & Burstein, J. (2003). Introduction. In M. D. Shermis & J. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. xiii-xvi). Mahwah, NJ: Lawrence Eribaum Associates.
＊ Zedelius, C. M., Mills, C., & Schooler, J. W. (2019). Beyond subjective judgements: Predicting evaluations of creative writing from computational linguistic features. *Behavior Research Methods, 51,* 879-894.