

英語初級学習者のパラグラフ・ライティング 評価基準の確立を目指して

Toward Establishing Criteria for Evaluating Paragraph Writing of
Pre-intermediate Learners of English

MITA, Kaoru

三 田 薫

英語コミュニケーション学科教授

SHIMODA, Atsuko

霜 田 敦 子

英語コミュニケーション学科非常勤講師

抄録：

日本人学生の英作文力不足は長年問題となっているが、それは文法知識や語彙の不足に留まらず、学生がエッセイを英語で書く機会が圧倒的に少ないことが要因となっている。学生が自分の考えを説得力を持って書く機会が少ない理由の1つは、日本人初級学習者向けの英文エッセイの評価基準が確立しておらず、教員の採点負担が大きいことにある。本研究では日本人の英語初級学習者の英文エッセイを測定する評価基準を確立する。エッセイ1本についてレベル1から4の基準を含む「分析評価」と、良し悪しの印象で1から4まで判断するという、印象による「全体評価」を複数の評価者が行う。両評価の相関分析により、「分析評価」の信頼性、妥当性の確立を目指す。また印象による「全体評価」の評価者4名がそれぞれ最高評価のエッセイに付けた評価の理由説明をテキストマイニングで分析し、英文の質の高さを決定する要素を明らかにする。

Abstract：

Japanese students' lack of English writing skills has been a long-standing problem. It not only originates from students' lack of grammatical knowledge and vocabulary, but also from the availability of fewer opportunities to write essays in English. Further, the current standards for evaluating the essays written by Japanese beginner-level learners in English are insufficient, and the burden of correcting and grading essays is too heavy for their teachers. This study establishes an evaluation standard to assess the English essays written by Japanese beginner-level English learners. For each essay, multiple evaluators conduct an "analytic scoring," which includes the criteria from levels 1 to 4 and a "holistic scoring," which

is a judgment from 1 to 4, based on the impression of goodness or badness. The reliability and validity of the "analytic scoring" are established through the correlation analysis of both evaluations. Text mining is used to justify the ratings provided by the impression raters to the highest-rated essays and identify the factors determining the quality of English writing.

キーワード：第2言語ライティング，ルーブリック，パラグラフ・ライティング，トピックの展開，分析評価，全体評価，印象評価，相関分析，内容分析，テキストマイニング

Keywords : Second language writing, rubrics, paragraph writing, topic development, analytic scoring, holistic scoring, general impression, correlation analysis, content analysis, text mining

1. はじめに

大学生の英語力はCEFR（ヨーロッパ言語共通参照枠¹⁾）でA2レベル（基礎段階の言語使用者）が大多数であり（GTEC Academic 実施大学の結果²⁾），さらに深刻なのは英作文力の低さである。文部科学省が高校3年生を対象に行った「平成29年度英語力調査」では、「与えられたテーマに対して，限られた時間の中で自分の意見や考えを説得力を持って書いて表現する力を測定する」意見展開問題を含むテストで，15.1%の生徒が無解答または0点のため採点対象から除外されており，まとまった英文を書くこと自体を放棄する学生が多い実態が明らかになった。

学生の英作文力不足は，単に文法知識や語彙の不足に起因するものではない。高校卒業まで学生がまとまった内容を英語で書く機会が圧倒的に少ないことが要因となっている。その理由の1つが，日本人英語初級学習者向けの英文エッセイ評価基準が未だに確立しておらず，採点が手探りとなる等，教員の負担が大きいことにある（江口，2012；大井，2015；平林，2016³⁾）。

一方，高等学校の英語教育では，2022年度高等学校学習指導要領の下で，「論理・表現」という科目が新たに加わることとなった。これは文部科学省が高校英語教育で，これまで以上に文章構造や発信技術に重きを置いたパラグラフ・ライティングを奨励していることを示している（山本，2020）。本稿では1つのエッセイについてWiseman（2012）のTopic Developmentという評価項目を応用して4段階それぞれの評価基準を明記した「分析評価」と，単にエッセイの良し悪しの印象で1から4の評価をする「（印象による）全体評価」（以下，「印象評価」と呼ぶ）を複数の評価者が行い，両評価の相関分析から「分析評価」ルーブリック⁴⁾（評価基準）の信頼性，妥当性を確立することを目指す。また外部評価者4名による印象評価で最高評価となったエッセイについて，外部評価者それぞれがコメントした理由をテキストマイニングで分析し，英文の質の高さを決定する要素を明らかにする。

第2節ではL2（第二言語）英文ライティングのルーブリックに関する先行研究を紹介し，第3節でリサーチクエスション，第4節で調査方法，第5節で調査結果を述べ，第6節で考察を行い，第7節でまとめる。

2. 先行研究

三田&霜田(2021)で使用した英文ライティング評価ルーブリックでは、「内容」の評価基準として Topic Development という概念を用いた。今回の調査では、この Topic Development の設定に信頼性と妥当性があるかについて検証する。

英文ライティングの評価法は全体評価 (holistic scoring) と分析評価 (analytic scoring) の2つに大別され、その比較研究が行われてきた。Carr (2000) は、これら2つのライティング評価法を分けることの意義を明確に主張し、自身の実証研究で分析評価の尺度は学習者に有益な情報をより多く提供できる可能性があるとして述べている。

Wiseman (2012) は L2 ライティングを評価する方法として全体評価ルーブリック (holistic scoring rubric) と分析評価ルーブリック (analytic scoring rubric) を用いて ESL (English as a second language) 学習者のライティングを評価し、それぞれの特徴と有効性を分析した。その結果、どちらのルーブリックも学習者の能力を明示的に区分しているが、分析評価ルーブリックを使用した方が、能力のレベル別の差がより明確に示され、全体評価ルーブリックを使用した場合より機能上優れていると結論づけている。そして、全体評価ルーブリックは短期的にみると時間と費用の節約になるというメリットがあるが、分析評価ルーブリックは長期的には学習者のライティング力の診断とクラス分けの目的のためにより好ましい採点法であると主張している。本稿の英文ライティングにおける「内容」の評価基準 Topic Development は、Wiseman の分析評価ルーブリックを応用している (pp.91-92)。

Ghalib & Al-Hattami (2015) では、EFL (English as a foreign language) 学習者のライティングを、評価基準を示さない全体的印象 (general impression)⁵⁾ による評価で調査したところ、評価者間で評価にばらつきがあり、学習者間のライティング能力の妥当な識別ができないという結果となった (p.226)。この結果を受けて、彼らは評価者による評価の揺れが最小限に抑えられるような、信頼性と妥当性の確保された評価基準の研究を開始した。すなわち全体評価ルーブリックと分析評価ルーブリックを、計量心理学的統計手法を使って比較分析し、その結果、2つの評価には強い相関性があるということを示した。その上で、分析評価ルーブリックはライティング力を測る基準をより明確に学習者に示すことができるという点で、全体評価ルーブリックよりも有効であると指摘している。また彼らは、評価方法の正確さと一貫性の向上を目指す研究として Wiseman (2012) の研究を紹介している。

国内では中西&赤堀(2004)が日本人英語教師とネイティブ英語教師で英文エッセイの評価の観点が異なることを紹介している。すなわち日本人は「内容、単語の選択、文法の正確さ」に焦点を当てる傾向があるが、アメリカ人は「パラグラフや構成、文と文のつながり」に注目していることを示し、「書き手の意図が読み手に十分に伝わるかどうか」に着目するネイティブ評価の重要性を指摘している。

江口(2012)は、担当教員の経験に依存した評価に偏りがちなライティング教育の現状を打破する方法として、ルーブリックを活用した評価法を提唱している。ルーブリックが教育現場において一般的な評価法にならない理由として、ルーブリックの作成および評価に時間と手間がかか

る点、評価者間で評価に揺れが生じるためその信頼性が疑問視されている点を指摘する。それらを解決するため、江口は「内容分析」の技法を応用した具体的なライティング・ルーブリックを考案し、その有用性を考察した。その中で、1) 短時間で客観的に学習成果の公平な評価ができるか、2) 効率的に採点できるか、3) 学生の学習に役に立つかの3点が検証された。その結果、江口の考案したライティング・ルーブリックは、効率的な評価、および学生の内容理解と学習意欲の向上をもたらすことが確かめられた。ただし評価者が著者一人であったため、信頼性の確保には複数の評価者による評価が必要であること、またより詳細で具体的な評価基準からなるルーブリックの開発と活用が今後必要になることが指摘されている。

平林 (2016) は、「自由英作文」⁶⁾ のルーブリック開発を目的とした研究を行った。日本では社会の国際化により発信型の英語教育活動が求められる一方、自分の考えや意見を英語にして記述する自由英作文の学習が必ずしも普及していない。その原因の一つが自由英作文の評価基準となるルーブリックが十分に整備されていないこと、それにより「学習者にとっては自由英作文に対する学びに支障をきたしがちになり、指導者も指導方法に戸惑うことになる。(p.23)」ことを平林は指摘している。

平林のルーブリックの対象レベルは「初級英語学習者」と「上級英語学習者」である。その理由は、CEFRの6つのレベル(下からA1, A2, B1, B2, C1, C2)では日本人の初級英語学習者のライティング力を正確に測ることが困難であり、また上級英語学習者の評価記述が抽象的すぎるため必ずしも有用とは言えないためであるとしている。そこで、初級学習者向けには日本人学習者の8割が属すると言われているCEFR-J A2レベル⁷⁾、上級学習者向けにはCEFR B2およびC1レベルを評価するルーブリックをデザインし、その妥当性を検証している。

平林の開発したルーブリックは分析評価ルーブリックであるが、統計的に検証した「特定要因」を評価の観点とする独自の特徴を持つ。また、CEFR-J A2レベルとCEFR B2およびC1レベルそれぞれの英作文の全体評価とそれらの分析評価ルーブリックをもとに算出した得点との相関関係を調査し、ルーブリックを検証している。その結果、それぞれのレベルの全体評価と分析評価の間には正の強い相関関係があることが明らかになり、開発されたルーブリックは全体評価に代用しうるものであることが検証されている。

興味深いのは、外れ値となったいくつかの自由英作文を平林が読み直したところ、外れ値となった要因が内容面での優劣に関係していると観察されていることである。たとえば、ルーブリックの得点が低いと全体評価が高い例では、「内容自体の面白さ」「興味深さの切り口が優れている」「内容面の洞察が優れている」といった点が全体評価を高くしていることを平林は観察している。平林はこれらの外れ値となった英作文については、開発されたルーブリックの評価観点だけでは、内容面での評価が十分に機能しない可能性を指摘している。

平林のデザインした自由英作文のルーブリックは、日本における初級学習者と上級学習者の評価の峻別を可能にする。これによりルーブリックが自由英作文の学習や指導の有効なツールとなり、同時に指導者は評価の効率性を高められる可能性がある。平林の研究は、社会的に重要性が認識されているものの普及が進まない自由英作文教育の現状打破につながる研究と言えるだろう。

3. リサーチクエスション

- (1) 英文ライティングの分析評価の評価項目「内容の質」を測るために Topic Development を用いた基準には信頼性、妥当性があるか。
- (2) 英文ライティングにおいて、内容の質の高さを決定する要素は何か。

4. 調査方法

4.1. 評価対象英文

本研究の調査対象者は、本短期大学で必修科目 Integrated English⁸⁾ を履修する1年生のうち、学期初め(5月)に実施した GTEC Academic を受検し、さらに事前(5月)と事後(翌年1月)の両方のライティングテストを受検した学生161名のライティングテスト結果を今回の調査対象とした。対象のライティングテストを受けた学生の学科内訳は、日本語コミュニケーション学科76名、英語コミュニケーション学科85名であった。学生の英語レベルは、GTEC Academic のスコアを CEFR のレベルに当てはめたところ、以下の4つのグループに分かれた。

- ① CEFR B2, B1 レベル：27名 (16.8%)
- ② CEFR A2 レベル上：43名 (26.7%)
- ③ CEFR A2 レベル下：62名 (38.5%)
- ④ CEFR A1 レベル：29名 (18.0%)

ライティングテストは、事前も事後もオンラインで実施した。具体的には学習管理システム(LMS)の manaba ver.2.95 (朝日ネット) を用いて受験させた。調査時にはコロナ禍で学生が登校できなかったため、学生は5月の事前テストも翌年1月の事後テストも自宅のパソコンを用いて manaba にアクセスし、受験することとなった。

ライティングテストの所要時間はブレインストーミングを含めて15分間であり、テスト用画面には、ライティングの入力用スペースだけではなく、受験者個人のブレインストーミング内容を記録するためのスペースも設けた。ライティングテストのトピックは「好きな場所」である。エッセイの指示文は以下の通りである。

ライティングテストトピック「好きな場所」

自分の行ってみたいところを決め、その場所と、行きたい理由を3つ書いて下さい。海外でも国内でも結構です。以下の表現で始めてください。

The place I want to visit most is (). There are three reasons.

4.2. 学生英文の評価基準

学生のライティングテスト結果は、以下の4つの評価項目のループリックで評価した。本研究は以下の4項目のうち、④「内容 (content)」に焦点を当てている。

ライティングテスト分析評価ルーブリックの評価項目

- ① 「流暢さ (fluency)」 語数
- ② 「エッセイの構成 (structure)」 (授業で設定したディスコースマーカーの有無)
- ③ 「文法の正確さ (accuracy)」 (授業で設定した指定文法エラーの有無)
- ④ 「内容 (content)」

リサーチクエスション (1) に含まれる Topic Development とは、元々 Wiseman (2012, p.91) の分析評価ルーブリック (analytic scoring rubric) の中の 1 項目である。Wiseman は「英文の質」を表すものとして、Topic Development (トピックの発展性) を設け、6 段階の評価基準で評価する分析評価ルーブリックを作成した。本調査では、この項目を応用し、以下のような 4 段階の評価基準を設けた。以下の 1) から 4) の各英文は、それぞれの評価を受けた学生の英文である。4 段階評価の中の「Detail 文」とは、詳しい説明や具体例を挙げて、主張に説得力を与える文のことである。

「Topic Development」の評価基準 (4 段階)

- 1) Detail 文が無いもの「1」(例)

The place I want to visit most is Okinawa. There are three reasons. First, I like hot place. Second, beach is beautiful and rich in nature. And finally, I've never eaten Okinawa food, and I want to try them.

- 2) Detail 文が少しあるが内容が限定的なもの「2」(例)

The place I want to visit most is Hawaii. There are three reasons. I want to eat delicious food in Hawaii. For example, I like garlic shrimp. I want to surf because my father is doing it. I want to go to the beach in the evening. Because the setting sun is beautiful.

- 3) Detail 文が複数ありトピックが発展し内容が深まったと考えられるもの「3」(例)

The place I want to visit most is Korea. there are three reasons. First, I want to go to different cities and go shopping. because I've been watching Korean dramas every day lately. therefore, the reason I want to go to Korea is probably influenced by the Korean drama. Second, I want to buy Korean cosmetics. they are so good for my skin. so, I often use Korean cosmetics. Third, there are many delicious foods in Korea. I especially like spicy food. so, I want to eat a lot of spicy food when I go to Korea.

- 4) 「3」の中で特に優れているもの「4」(例)

The place I want to visit most is Okinawa. There are three reasons. First, I have been to Okinawa only once. In addition, the weather was very bad when I went to Okinawa. So, I want to see the beautiful sea of Okinawa on a sunny day. Next, I want to eat delicious food in Okinawa. When I went to Okinawa, I ate delicious food such as Okinawa soba and Seta Andagi. And I like them very much. So, if I go to Okinawa again, I want to eat them. Finally, I like Okinawa time. Because it is very slowly, we can relax in Okinawa time. That's why I want to go to Okinawa.

4.3. 学生英文の評価者と評価内容

学生 161 名分の前期始め、後期終わりのライティングテスト英文計 322 本を以下の評価者 6 名が評価した。

評価者内訳（仮名）

- ① ATSUKO（日本人）
- ② EIKO（日本人）
- ③ BRIT（ネイティブ）
- ④ JANEY（ネイティブ）
- ⑤ 教員 2 名（執筆者日本人）

外部評価者①～④の 4 名は、英文添削サービスを行う会社 IDIY⁹⁾ を通して依頼した評価者である。外部評価者 4 名には、以下のレベル 1 から 4 の「印象評価」、およびレベル 4 (Excellent to very good) についてその評価を付けた理由をコメントすることを依頼した。彼らにはサンプルとして前述「Topic Development」の評価基準による学生英文を示した。

IDIY スコアリング講師への依頼内容

英語が専門でないジュニアカレッジの学生が 15 分で書いたエッセイです。スペルミス・文法のミスなどは気にせず、エッセイの印象を評価頂き、最もよかったエッセイについてのみ「なぜその評価が良かったのか」一言コメントをお知らせください。

レベル 1. Very poor

レベル 2. Fair to poor

レベル 3. Good to average

レベル 4. Excellent to very good

評価者⑤である教員 2 名（執筆者日本人）は、それぞれが 322 本の英文に「Topic Development」の分析評価基準に基づく 1~4 の「分析評価」を行った上で、互いに異なってい

る評価については話し合って調整した。

4.4. 分析方法

リサーチクエスチョン (1) については、外部評価者4名一人一人の「印象評価」と教員2名の調整後の「分析評価」について、相関分析を行った。

リサーチクエスチョン (2) については、外部評価者4名それぞれが、「印象評価」でレベル4の評価に付けた理由コメントについて、評価者ごとにテキストマイニング¹⁰⁾を行った。テキストマイニングには、KH Coder 3を使用した。

5. 調査結果

5.1. 分析評価と印象評価の相関分析

学生の英文エッセイ322本について、外部評価者4名が行った「印象評価」と、教員2名が行った「分析評価」を比較した。表1は各評価者によるレベル別のエッセイ数である。

表1 各評価者によるレベル別の英文エッセイ数

	レベル1	レベル2	レベル3	レベル4	合計
ATSUKO	105	91	92	34	322
EIKO	94	143	61	24	322
BRIT	56	101	123	42	322
JANEY	5	30	156	131	322
教師	103	101	86	32	322

表2は評価者4名それぞれの印象評価と教員の分析評価について、ピアソンの積率相関係数、平均、標準偏差を表している。

表2 評価者4名の印象評価と教員の分析評価の相関係数と平均、標準偏差 (N=322)

	ATSUKO	EIKO	BRIT	JANEY	教員	M	SD
ATSUKO	—					2.171	1.0041
EIKO	.782**	—				2.047	.8830
BRIT	.820**	.772**	—			2.469	.9276
JANEY	.738**	.673**	.731**	—		3.283	.6955
教員	.830**	.793**	.810**	.678**	—	2.146	.9829

** $p < .01$

相関係数はいずれも1%水準で有意である。EIKOとJANEY、教員とJANEYの相関係数は「比較的強い相関がある」(0.7未満)ことを示している。それ以外の相関係数はすべて「強い相関がある」(0.7以上)ことを示している。

5.2. 印象評価者4名のコメントのテキストマイニング

続いて4名の外部評価者が評価したレベル4エッセイについて、各評価者がコメントした理由説明を、テキストマイニングで分析した。

5.2.1. ATSUKO

ATSUKOの理由コメントの総抽出語数は1,438語(72文)であった。抽出語の頻出語上位5件は「良い」(31回)、「思う」(26回)、「とても」(22回)、「する」(20回)、「まとめる」(16回)であった(表3)。

表3 ATSUKOのコメントの頻出語数上位30語

No	抽出語	品詞	出現回数	No	抽出語	品詞	出現回数
1	良い	形容詞	31	16	伝える	動詞	7
2	思う	動詞	26	17	魅力	名詞	7
3	とても	副詞B	22	18	完成	サ変名詞	6
4	する	動詞B	20	19	高い	形容詞	6
5	まとめる	動詞B	16	20	思い	名詞	6
6	英文	名詞	14	21	表現	サ変名詞	6
7	具体	名詞	14	22	訪れる	動詞	6
8	理由	名詞	13	23	韓国	地名	5
9	文章	名詞	12	24	経験	サ変名詞	5
10	伝わる	動詞	11	25	自然	形容動詞	5
11	きれい	形容動詞	8	26	自分	名詞	5
12	行く	動詞	8	27	大変	形容動詞	5
13	特に	副詞	8	28	ある	動詞B	4
14	上手	形容動詞	7	29	オーストラリア	地名	4
15	短時間	副詞可能	7	30	沖縄	地名	4

共起ネットワークを作図したところ、「良い」「思う」「英文」「理由」など一般的な語句が出現し特徴が読み取りにくかった(図1)。

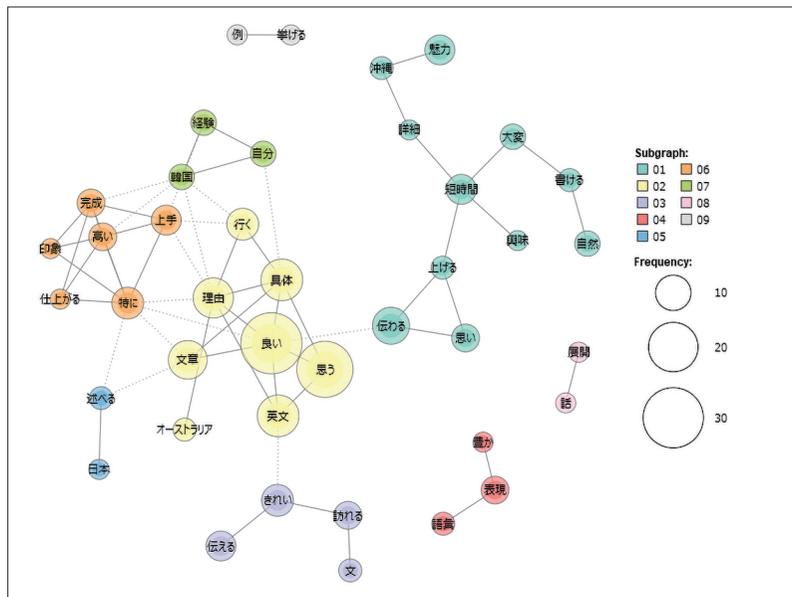


図1 ATSUKO 1

そのため、KH Coderの「品詞による語の取捨選択」で「良い」「思う」「英文」「理由」「文章」を「使用しない語」に指定し、オプションで「地名」を外したところ、特徴的な語句の結びつきがより鮮明な共起ネットワークが出現した(図2)。

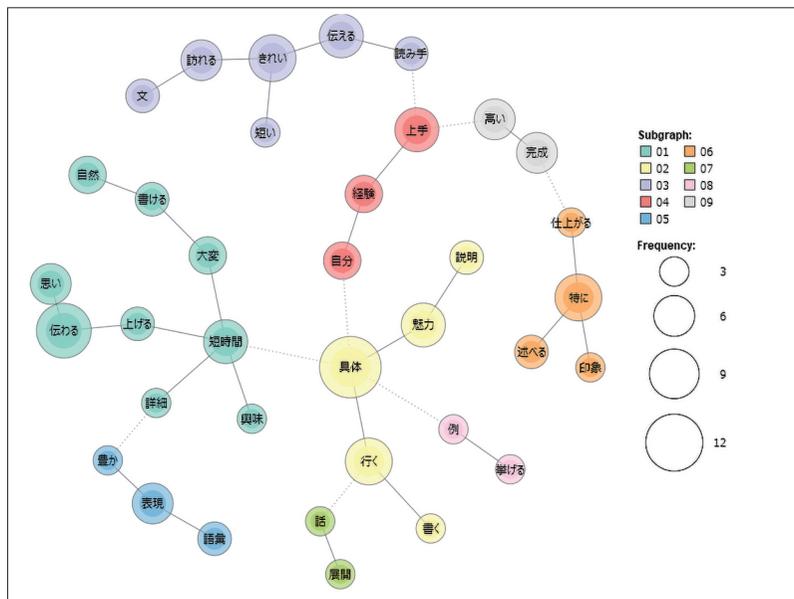


図2 ATSUKO 2

図2で多く現れた単語は「短時間」「具体」「きれい」「上手」「表現」である。「短時間」という単語は、以下の単語グループ「興味」「詳細」「上げる」「伝わる」「思い」「大変」と強く共起している。「詳細」は「表現」「豊か」「語彙」とも共起関係にある。また、「具体」は単語グループ「行く」「魅力」「説明」と強い共起関係にあり、「例」「挙げる」や「自分」「経験」「上手」とも共起関係にある。このことから、ATSUKOが優れたエッセイの評価の基準を、制限時間で書く能力、語彙の豊かさ、自分の体験といった具体例を挙げて説明していることに置いていることがわかる。以下はコメントの抜粋である。

「短時間で大変よくまとまった文章を仕上げたと思います。訪問したことの無いフランスについて学んだこと、興味を持った点を具体的にまとめていた点が良かったです。英文もとてもきれいです。」

「ディズニーワールドへの強い思いが文章全体から伝わってきました。自分の過去の経験、そしてこれからやりたいことを具体的に生き生きとした文章にまとめている点が良いです。」

5.2.2. EIKO

EIKO コメントの総抽出語数は245語(24文)であった。抽出語の頻出語上位5件は「充実」(18回)、「理由」(18回)、「具体」(10回)、「構成」(7回)、「内容」(7回)であった(表4)。

表4 EIKOのコメントの頻出語数上位9語

No	抽出語	品詞	出現回数	No	抽出語	品詞	出現回数
1	充実	サ変名詞	18	6	工夫	サ変名詞	4
2	理由	名詞	18	7	文章	名詞	3
3	具体	名詞	10	8	詳細	形容動詞	2
4	構成	サ変名詞	7	9	美しい	形容詞	2
5	内容	名詞	7				

語の取捨選択を行わずに共起ネットワークを作図し、さらに共起ネットワークの調整で「最小スパニング・ツリーだけを描画」を選択したところ、図3が出現した。

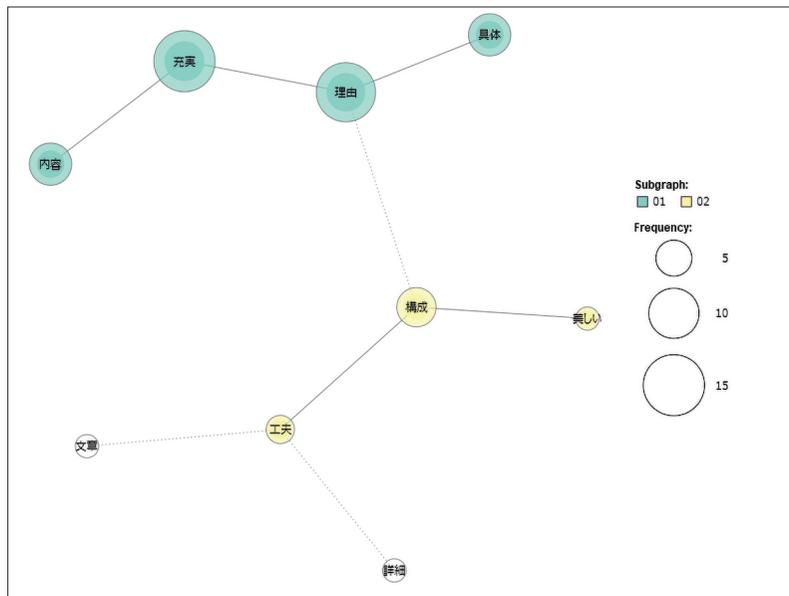


図 3 EIKO

図 3 で多く現れた語は、「理由」と「構成」である。「理由」は「具体」「充実」「内容」と強く共起している。また、「理由」は「構成」とも共起関係にあり、「構成」は「美しい」「工夫」と強い共起関係にある。ERIKO は、優れたエッセイの評価の基準を、理由の具体性や充実度、構成の工夫や美しさに置いていることがわかる。以下はコメントの抜粋である。

「理由が具体的で充実している。」

「構成が工夫されてあって理由が詳細である。」

5.2.3. BRIT

BRIT コメントの総抽出語数は 568 語 (71 文) であった。抽出語の頻出語上位 6 件は *be* (17 回), *very* (12 回), *vocabulary* (12 回), *really* (10 回), *minute* (9 回), *well* (9 回) であった (表 5)。

表 5 BRIT のコメントの頻出語数上位 30 語

No	抽出語	品詞	出現回数	No	抽出語	品詞	出現回数
1	be	Verb	17	16	great	Adj	5
2	very	Adv	12	17	content	Noun	5
3	vocabulary	Noun	12	18	wide	Adj	5
4	really	Adv	10	19	writing	Noun	5

表6 JANEY のコメントの頻出語数上位 30 語

No	抽出語	品詞	出現回数	No	抽出語	品詞	出現回数
1	you	PRP	216	16	make	Verb	43
2	well	Adv	107	17	work	Noun	43
3	do	Verb	85	18	essay	Noun	37
4	be	Verb	84	19	assignment	Noun	36
5	have	Verb	77	20	it	PRP	36
6	reason	Noun	73	21	great	Adj	29
7	interesting	Adj	70	22	good	Adj	29
8	detail	Noun	61	23	clearly	Adv	27
9	personal	Adj	61	24	job	Noun	26
10	visit	Verb	61	25	structure	Noun	26
11	give	Verb	47	26	very	Adv	21
12	read	Verb	44	27	structure	Verb	20
13	want	Verb	44	28	extra	Adj	18
14	write	Verb	44	29	which	W	17
15	clear	Adj	43	30	like	Verb	16

共起ネットワークを作図したところ、*you, do, have* など一般的な語句が出現し特徴が読み取りにくかった (図6)。

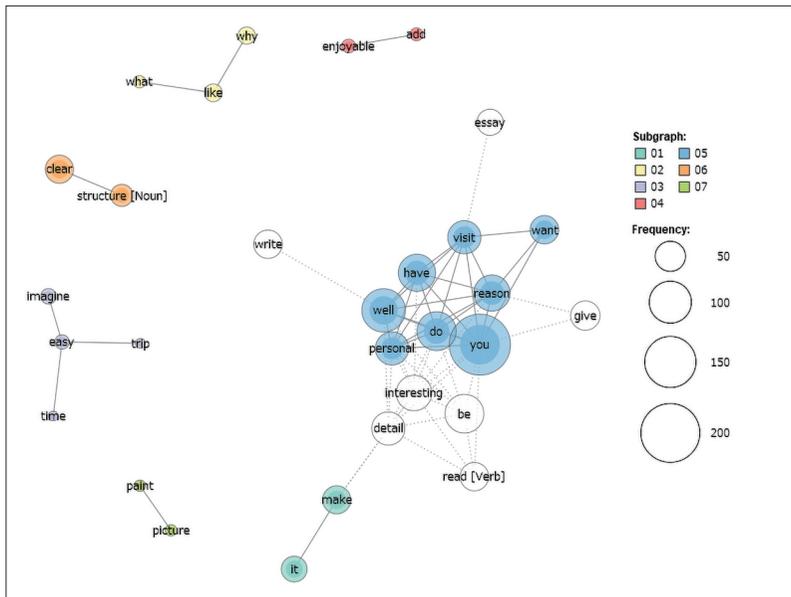


図6 JANEY 1

そのため、共起ネットワークのオプションで「最小／最大出現数による語の取捨選択」の「最大出現数」を70とし *you, well, do, be, have, reason* を外した。また *want, assignment, visit* といった一般的な語を「語の取捨選択」で外した。さらに共起ネットワークの調整で「最小スパンニング・ツリーだけを描画」を選択した。すると結びつきがより鮮明な共起ネットワークが出現した(図7)。

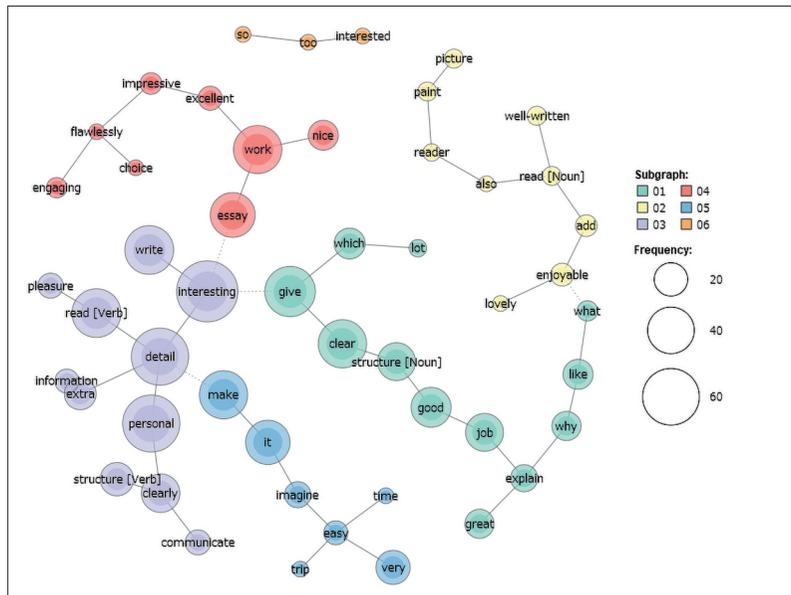


図7 JANEY 2

図7で多く現れた語は、*structure, detail, work, read*である。*structure*は*clear, good*と強く共起している。また、*detail*は*personal, information, interesting*と、*work*は*excellent, impressive, flawlessly, engaging*と、*read*は*well-written, enjoyable, lovely, paint, picture*と強い共起関係にある。JANEYは、優れたエッセイの評価の基準を、構成の良さ、個人的で詳しく興味深い内容、人を惹きつける印象の良いエッセイ、読み手に情景が思い浮かぶような楽しい読み物であることに置いていることがわかる。以下はコメントの抜粋である。

“This is an impressive essay about why you would like to visit France. It is flawlessly written, and you have clearly structured your work. The interesting details with personal reasons to back up your choice make it engaging. Excellent work!”

“Your essay was interesting to read because of the clear picture you have painted of what you would like to do in England. Great work!”

“It is very easy to imagine your trip to the US after reading your well-structured and detailed reasons for wanting to visit. Well done!”

6. 考察

リサーチクエスト (1) 「英文ライティングにおいて、Topic Development を判断基準とした内容を測る分析評価は信頼性、妥当性があるか」については、外部評価者4名が英文エッセイの良し悪しで判断する「印象評価」と、教員2名による「分析評価」の相関分析を行い、いずれの評価者の組み合わせも1%水準で有意に（比較的）強い相関関係が示された。そのため、「英文の質 (contents)」に関して Wiseman (2012) の Topic Development を応用した分析評価の4段階の評価基準について、信頼性と妥当性が確認されたと考える。

リサーチクエスト (2) 「英文ライティングにおいて、内容の質の高さを決定する要素は何か」については、4名の外部評価者（日本人2名、ネイティブ2名）のコメントをテキストマイニングにより内容分析した。その結果、共通した概念がいくつか抽出された。

注目すべきは、内容の展開に関連する単語が4名の外部評価者全員のコメントに現れていた点である（日本人2名では「具体」「詳細」、ネイティブ2名では「develop」「detail」）。これにより、「具体」に関連する概念が、エッセイの内容が優れていると判断されるための主要な要因であるという結論が得られた。

次に、「構成」「structure」という単語が、外部評価者4名のうち3名のコメントに多く現れていることが観察された。また「構成」という単語が現れなかった1名の評価者のコメントにも「大変よくまとまった文章」という記述があり、これは「構成」を連想させるものと推測される。よって、リサーチクエスト (2) の「内容の質の高さを決定する要素」とは、説明が「具体的に記述されていることと、整った「構成」で書かれていることの2点であると判断することができる。

さらに、テキストマイニングの結果から、英文の優れた内容の特徴づける概念を捉えることができた。「語彙」「vocabulary」に共起した「豊かさ」「rich」「descriptive」「varied」といった記述から、良いエッセイにおける「語彙」の重要性を読み取ることができる。また、「information」と「pack」が共起し、「content」と「lot」が共起していることから、「情報量の多さ」が高く評価されていることがわかる。

ネイティブ評価者のコメントでは、「picture」に共起する語として「paint」「place」が現れている。情景をありありとイメージできるような描写を高評価の理由とするコメントが、ネイティブ評価者2名に共通していることは興味深い。これらの共起語は、優れたエッセイの主要要素の1つである「具体」性の下位概念と読み取ることができる。

最後に、外部評価者4名のうち2名が「短時間」「just 15 minutes」など、15分間という限られた時間で書き上げたことを高く評価していることは興味深い。内容の良し悪しに直接かかわる要素ではないが、制限時間で好印象のエッセイを書きあげる技術もライティングスキルの評価基準の1つとなり得ることを示している。

英文エッセイの「全体評価」は、評価者によって評価が変わるという先行研究がある。今回の調査でも、外部評価者4名のコメントから、それぞれ独自の観点で評価していることが読み取れる。例えば JANEY は、学生に直接呼びかけることを意識して書いていることが、主語を *you*

としていることや激励の表現からわかる (例: *Your essay was interesting to read because of the clear picture you have painted of what you would like to do in England. Great work!*). 4名の中でレベル4の最高評価を最も多くつけたのも JANEY である. 一方, 最低限の語数で辛口の評価をしているのが EIKO である (例: 理由が具体的で充実している). 評価平均も EIKO は最低, JANEY は最高であるが, 両者の評価の相関分析の結果, 比較的強い相関 (.673**) を示していた. 先行研究では評価者により評価にばらつきがあることが指摘されているが, 今回の調査結果からは, 評価のスタンスが異なる評価者同士でも, レベル1から4の「印象評価」においては相関がみられることが確認できた.

中西&赤堀 (2004) にあるように, 本調査でも日本人評価者とネイティブ評価者で評価の視点が異なることが明らかになった. 日本人評価者には内容の具体性とエッセイとしての構成を評価する傾向がある一方, ネイティブ評価者には語彙の豊かさや読み手が情景を思い描きながら読めるような内容を評価する傾向がみられた.

今回テキストマイニングにより抽出された概念は, 教員が学生の英文を評価するためのみならず, 学生の英文ライティング指導にも役立てることができる. 学生が英文の自己評価やピア評価をする際, 今回明らかになった概念を1つずつ項目にしたチェックリスト (例: 「序論・本論・結論が書かれていますか (構成)」「自分の経験が入っていますか (具体)」) を用いて確認させるなどの指導により, 従来曖昧になりがちだった項目である「内容の質」に対する理解, 良い評価を得るための条件を, 学生自身に意識させるなどの工夫が考えられる.

7. 終わりに

学生の英文エッセイ 322 本について Wiseman (2012) の Topic Development を応用したレベル1~4の「分析評価」と, 単にエッセイの良し悪しの印象によるレベル1~4の「印象評価」を複数の評価者が行い, 両評価の相関分析から「分析評価」ルーブリックの信頼性, 妥当性を確認した. また外部評価者4名それぞれが最高評価を付けた英文エッセイの理由説明をテキストマイニングで分析し, 「内容の質」の高さを決定する要素を抽出した. 今回の結果を学生の英文ライティング指導に活用すると同時に, 英文ライティングの「内容の質」の評価基準についての調査研究を続けていきたい.

[注]

1. CEFR とは, Common European Framework of Reference for Languages の略称である.
2. GTEC[®] (ジテック/Global Test of English Communication) とは, 株式会社ベネッセコーポレーションが実施している英語力を測定するためのスコア型英語4技能検定である.
3. 日本人学習者の英文ライティングについて, 多くの研究者が, 英文エッセイ評価基準が確立しておらず, 教員の負担が大きいことを指摘している. 例えば江口 (2012) は「ルーブリックが一般的な評価方法となっているとは言いがたい. その理由の一つに, ルーブリックを作成することやルーブリックを使って評価することに時間と手間がかかるということがある。」(pp. 73-74), 「ライティングの評価は手間がかかり, 評価者内でも評価に揺れが生じるのが通常である」(p. 82) と指摘している.
4. スティーブンス & レビ (2014) は, ルーブリックを「ある課題をいくつかの構成要素に分け, その要素ご

- とに評価基準を満たすレベルについて詳細に説明したもので、様々な課題の評価に使うことができる」(p. 2)と説明している。
5. さらに Ghalib & Al-Hattami (2015) は、次の Wiseman の説明を引用している。"Holistic scoring is "a global approach" to scoring that is underscored by the idea that "writing is a single entity which is best captured by a single scale that integrates the inherent qualities of the writing" (Wiseman, 2012, p. 59)。
 6. 平林は「自由英作文」という言葉について Richard et al. (1992:147) を引用し、「(自由英作文とは) 学生が与えられた時間内 (例えば 3 分) で、あるトピック (テーマ) に対してできるだけ多くの英文を記述するものである。その目的はライティングにおける流暢さ (この場合、情報用の豊かさ、すなわち、書く量を表す) の向上やライティングという課題に伴うアイデアの産出をはかるために、英文法や (語彙など) の正確さを気にすることなく、できるだけ多くの英文を記述させることである」と定義している。
 7. CEFR-J とは、CEFR を日本人英語学習者のためにカスタマイズしたものである。
 8. Integrated English は、短期大学の全学必修英語科目で Integrated English a (1 年前期)、Integrated English b (1 年後期) 両方を開講している。前後期とも週 2 コマの授業で、そのうち 1 コマは日本人教員、もう 1 コマはネイティブ教員が担当する。調査年度はコロナ感染拡大の影響で、オンライン授業開始時期が遅くなったことにより、1 回目調査が 5 月となった。
 9. この会社は世界中に英文添削講師を抱え、11 年の実績がある。
 10. テキストマイニングは、文章データを単語ごとに切り取り、量的な方法で分析し、その結果を視覚化する内容分析の手法である。

[参考文献]

- Carr, Nathan. T. (2000). A Comparison of the Effects of Analytic and Holistic Rating Scale Types in the Contest of Composition Tests, *UCLA Issues in Applied Linguistics*, 11 (2)
- Ghalib, T. K., & Al-Hattami, A. A. (2015). Holistic versus Analytic Evaluation of EFL Writing: A Case Study. *English Language Teaching*, 8 (7), 225-236.
- Wiseman, C. S. (2012). A comparison of the performance of analytic vs. holistic scoring rubrics to assess L2 writing. *International Journal of Language Testing*, 2 (1), 59-92.
- ダネル・スティーブンス, アントニア・レビ (佐藤浩章監訳, 井上敏憲, 俣野秀典訳) (2014). 『大学教員のためのルーブリック評価入門』, 玉川大学出版部.
- 江口真理子. (2012). 内容分析の手法を用いたライティング・ルーブリック具体化の試み. *鳥根県立大学総合政策学会総合政策論叢*, 24, 73-84.
- 平林健治. (2016). 重回帰分析により抽出した評価の観点に基づく自由英作文のルーブリックのデザイン (Doctoral dissertation, 慶應義塾大学).
- 石岡恒憲. (2008). 小論文およびエッセイの自動評価採点における研究動向 (<特集> テキストの自動評価). *人工知能*, 23 (1), 17-24.
- 清野光雄, & 竹内孔一. (2019, August). ニューラルネットワークを利用した日本語小論文の自動採点の検討. In *IEICE Conferences Archives*. The Institute of Electronics, Information and Communication Engineers.
- 工藤洋路, & 根岸雅史. (2002). 自由作文の採点方法による採点者間信頼性について. *全国英語教育学会紀要*, 13, 91-100.
- 三田薫 & 霜田教子. (2020). 学生の英文ライティング力向上の分析 ——Fluency が伸びた学生の日本語の干渉によるエラーと表現力の変化—. *Jissen English Communication*, 50, 6-33.
- 三田薫, & 霜田教子. (2021a). 学生の英文ライティング力向上の分析 その 2 : 文法・構造・論理の重点的指導によるライティングの習熟度別変化. *Jissen English communication*, 51, 14-46.
- 三田薫, & 霜田教子. (2021b). 学生の習熟度別 英文ライティング力向上の分析一弱点克服の重点的指導によるライティングの変化一. 実践女子大学短期大学部紀要 = *Jissen Women's Junior College Review*, 42, 63-83.
- 中西千春, & 赤堀侃司. (2005). 日本人英語教師とネイティブ英語教師のライティング評価の相違. *日本教育工学会論文誌*, 28, 229-232.
- 大井恭子. (2015). 4 技能試験時代のライティングに必要な指導と評価とは. *英語教育*, 12, 10-12.
- 大澤真也. (2006). 実践と研究をつなぐライティング評価法. 広島修大論集. 人文編, 47(1), 221-236.
- 山本昭夫. (2020). 「論理・表現」で新しいことと継承されること. *チャートネットワーク*, 93, 18-20. <https://www.chart.co.jp/subject/eigo/cnw/93/93-5.pdf> (2021 年 10 月 4 日閲覧)