英語初級学習者のパラグラフ・ライティング自動採点システム開発の試み Part2

…英文の内容の質を測るシステムの更新と英文の質向上の分析

Automated Scoring System for Paragraph Writing of Pre-Intermediate Learners of English Part 2

… Upgrades to the Assessment System for the Quality of Content of English Writing and Analysis of English Writing Improvement

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：

短期大学１年生向け英語必修科目で実施しているライティングテストのデータを用いた自動採点システムを開発した．人工知能の機械学習に，過去の同一テーマのライティングテストの英文データを，教師評価と共に入力し，クラウドコンピューティングで一般公開されている人工知能の機械学習のサービスを用いて「教師あり学習」を行った．自動採点システムは英文の「内容の質」のみを採点対象とし，Level 1 から Level 4 の 4 段階の採点を行う．このシステムを用いて2023 年度の英語必修科目のライティングテストの英文 106 件を採点したところ，自動採点システム（Model C）と教師評価の一致率は 49.1%，相関係数は 0.600** であった．同システムを授業内で学生に使用させた上でアンケート調査を行い，それをテキストマイニングで分析した．

Abstract：

We developed an automated scoring system using data from past writing tests on one theme in a required English course for first-year junior college students and teacher evaluations. These data were input into a cloud-based artificial intelligence “supervised” machine-learning system that scored the texts’ “quality of content” at four levels (1–4).

Subsequently, 106 writing tests for the required English course in Academic Year 2023 were scored; the agreement rate between the automated scoring system and the teacher's evaluation was 49.1% (R = 0.600**). The students used the system in class and completed a questionnaire, which was analyzed using text mining.

キーワード：自動採点システム，第 2 言語ライティング，パラグラフ・ライティング，テキストマイニング，人工知能，機械学習，教師あり学習，一致率，内容の質，典型事例

Keywords：Automated Scoring System, Second Language Writing, Paragraph Writing, Text Mining, Artificial Intelligence, Machine Learning, Supervised Learning, Agreement Rate, Quality of Content, Exemplars

1. はじめに

著者らは学生の英文ライティングの「内容の質」を自動評価するシステムの開発を試みてきた（三田・霜田, 2023a）．本稿は，その後に得られた学生英文のデータを加えてシステムを更新した後の精度と，自動採点システムを使用した学生の英文の変化を検証することを目的としている．

この調査継続中の 2022 年 11 月に生成 AI の大規模言語モデル ChatGPT が登場し，それがさまざまな分野で話題となっている．教育の分野での利用に関する議論も始まり，英語ライティングの評価，添削，フィードバックに大きな変革をもたらす可能性を示す研究がすでに行われている（Mizumoto. A, Eguchi. M, 2023: 竹ノ内, 2023）．また柳瀬（2023）は，非英語話者が英語による学術論文を書く際の機械翻訳と生成 AI の活用例を示し，人間と AI が相互補完的に作業を進める英文ライティングを推奨している．一方で，柳瀬（2022）は，機械翻訳によって知的生産が英語的な語り一辺倒になってしまうことの危うさについて「バベルの塔」になぞらえて警鐘を鳴らしている．また川西（2023）は，「一教師としては，十分な準備期間がなく，また，教育や発達に関するエビデンスがないままにこうしたテクノロジーが広まることに頭を悩ませる部分はおおいにある」（p. 107）と述べている．本研究で開発中の英文ライティング自動採点システムについても，こうした議論を踏まえつつ，如何に改善していくかが今後の課題となるだろう．学生の英文ライティング力の向上をどの立ち位置で目指すのか，一人一人の教師自身が問われる時代となった．

筆者らは短期大学で 1 年次英語必修科目（Integrated English）[1)] において，過去数年間にわたり特定のテーマで年 3 回ライティングテストを行い，そのデータを分析してきた（三田・霜田, 2020, 2021a, 2021b, 2022a, 2022b）．2021 年度には，調査データの一部について人工知能の機械学習を用いた自動採点システムを開発するという試みを開始した．このシステム開発は，授業担当者の英文ライティングの採点に関わる作業負荷の削減と学習者の自律的学習の促進を目的としている．短期大学の 1 年次英語必修科目という限定的な学習環境における自動採点システムの実装を目指し，自動採点システムの開発に関する専門知識を持たない筆者らが，公開クラウドベース

の機械学習を利用して教育的コンテキストでのモデル構築を試みた．今回は自動採点システム Model C 開発の経緯と授業で学生に使用させた結果について報告する．

第 2 節では英語教育の現場における自動採点システムについての先行研究を紹介し，第 3 節でリサーチクエスチョン，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

自動採点システムでは初期モデル（Model A）より，出力される評価得点と同時に 4.4 節の表 3 のように短いフィードバックも自動的に提示されるようになっている（例：評価 1：1 点… Detail 文がありません．理由の詳細や具体例をそれぞれの理由に付け加えてみましょう）．第 2 期モデル（Model B）開発の際，自動採点システムで出力される評価得点と同時に短いフィードバックの表示だけではなく，過年度の学生が同じトピックで作成した英文をモデル英文（典型事例）として提示する機能を追加し，Model A, B どちらの出力画面にも付加した（三田・霜田, 2023b）．これにより自動採点システム Model A, B は，教師の自動採点ツールとしてだけではなく，学生の形成的評価につながるツールとしても機能する可能性を持つことになった．

自動採点システムのフィードバックに関する研究から，フィードバックには様々な効果があることが明らかになっている（小林・石井, 2019: 石井・近藤, 2020a,b: Van Beuningen, C. G., De Jong, N. H., & Kuiken, F., 2012）．Model A, B に搭載したこの自動フィードバックでは，学生は自分の書いた英文の評価得点（レベル 1～ 4 ）と同時に，評価得点の一つ上のレベルのモデル英文を見ることができ，自らの英文を向上させるためのヒントを学ぶことができる．本研究で使用した「内容の質」を測る自動採点システムでフィードバックとして用いたモデル英文は「パフォーマンス評価研究」における「典型事例（exemplars)」の概念と一致する．

典型事例（exemplars）とは，質または能力のレベルを定めた水準の典型的なものとなるように選ばれた重要な例と定義され（Sadler，1987），パフォーマンス評価において各学習者が目指す見本として自己評価力を高める機能があるとされている．本研究でフィードバックとして採用した優れたモデル英文は Sadler の「典型事例」にあたるもの，すなわち学生が目標とのギャップを自己評価する機会を提供するものであると考える．岩田（2020，2022）は，「自己評価力」について「評価基準を理解し，それを自分のパフォーマンスに適用することで，目標とのギャップを適切に把握する力」と定義している．丹原他（2020）は，教員の用意した典型的なパフォーマンスの事例を指導に用いることで，学生が「評価する」だけでなく「分析する」などの活動も期待できると指摘している．

平林（2016）の開発した日本人初級学生用ルーブリックは，「総語数」「語彙や文体の難易度」「エラーが少なく熟達した英文」の 3 点でライティングを評価するもので GTEC の全体評価と 90％以上の一致度があることが検証されている．この研究で興味深いのは，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が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．現在開発中の自動評価システムは，これまで難しかったライティングの「内容の質」の評価を補完する可能性を期待することができるだろう．

3. リサーチクエスチョン

(1)　学生のライティング英文の教師による採点と自動採点システム（Model C）による採点の一致率はどの程度であるか

(2)　自動採点システムを授業で導入する前と導入した後の学生の英文にはどのような変化が見られるか

(3)　自動動採点システム（Model C）を利用した学生のアンケートから示唆されるものは何か

4. 調査方法

4. 1. 自動採点システムの概要

筆者らは，Amazon 社がクラウドコンピューティングを介して提供する Amazon Machine Learning の機械学習機能を活用し，自動採点システムの開発に関する実験的研究を実施した．自動採点システムの分野の先行研究において導入されている「特徴量」は，本研究では導入していない．また 1 回のテストで大量のデータが収集できる検定試験や入学試験に比べて，入力できるデータ件数も圧倒的に不足している．データ数の不足を克服するため，1）ライティングのテーマを 1 つに限定する，2）毎回同じテーマでテストを実施することにより，データの積み上げを可能にする（学生には同じテーマであっても毎回異なる対象を選択することを義務付けている），3）テストで使用する表現も可能な限り限定する，4）エッセイの構成や，使うべきディスコースマーカーを指定する，5）文法やスペリングのエラーは採点の対象とせず，「内容の質」に限定して 4 段階の評価を行うといった方策を導入している．このアプローチを用いて，2020 年度，2021 年度，2022 年度の授業内でライティングテストを年 3 回実施し，さらに 2023 年度 1 回目のデータを加え，それらのデータに教師の採点結果を「正解」として加えて機械学習の「教師あり学習」を実施した．

ライティングテストの主題を一貫して同一にする背景には，データ量の増加という実用的な動機のみならず，学生が特定のテーマに反復的に取り組むことで，パラグラフ・ライティングの構造や内容の深化に関する学びを定着させるためという意図がある．そのため，このライティングテストは総括的評価（学習の最後に，生徒の学力の達成度を確認するために行う評価）ではなく，形成的評価（学生がライティングを改善できるようにサポートするための評価）に重点を置いている．

(1) 本稿で用いる自動採点システム（Model C）の概要

① AI の機械学習の「教師あり学習」を用いる.

② 各英文についての教師の評価を「教師あり学習」の「正解」として与える.

③ 「内容の質」について Level 1 から Level 4 を予測する「分類」を行う.

(2) 英文採点インターフェイスのシステム構築の前提条件

① API（Application Programming Interface）を使用して英文の採点結果を取得，表示する.

② 複数の評価モデルを切り替えて使用することを可能とする.

(3) 外部インターフェイス（WebAPI）

① Amazon Machine Learning エンドポイント

② テキストの送信と評価の取得に使用

(4) 機械学習に用いるデータ

2020 年度，2021 年度，2022 年度に 3 回ずつ実施したライティングテストの英文データ（計 9 回分）と 2023 年度 1 回分の英文データ，ならびに全 10 回分の教師評価の採点データ

第 1 期モデル（Model A）は 2022 年 9 月授業開始前に開発され，第 2 期モデル（Model B）は 2022 年 1 月最終授業前に開発されている．第 3 期モデル（Model C）は，2023 年後期授業開始前に開発された．Model A, B, C それぞれが使用した英文エッセイは以下の表 1 の通りである．前期始め，前期末，後期末の 3 回分のライティングテストの英文をデータベースとして使用している.

表 1　自動採点システムの各モデルで使用したデータベースの英文エッセイ

Model A	2021 年度 3 回分
Model B	2021 年度 3 回分，2020 年度 3 回分
Model C	2020 年度 3 回分，2021 年度 3 回分，2022 年度 3 回分，2023 年度 1 回分

Model C を開発するにあたって，機械学習に読み込ませるデータである 2020 年度から 2022 年度までの 3 年分の英文の教員評価について，執筆者 2 名で見直しを行った．この見直しにより，過去の Model 開発の際，判断があいまいになりがちだった部分が解消され，評価を確定することができた．見直しで実際に変更された教員評価はわずかであった．データ入力や外部インターフェイスのシステム構築は，前回同様専門の業者[4] に委託した.

各モデルの開発では，それ以前のモデルにデータを追加するのではなく，その時点までに揃っているデータをまとめて機械学習に読み込ませている．Model C 開発には，表 2 の件数の英文をデータベースとして使用した.

表 2　Model C 開発に用いたデータベースの英文エッセイ

英文回収時期	英文件数
2020 年度 5 月 7 月 1 月	484
2021 年度 4 月 7 月 1 月	509
2022 年度 4 月 7 月 1 月	394
2023 年度 4 月のみ	119
合計	1506

Model C 開発時には 2023 年 7 月の英文は回収されていたが，その英文は後期初めの授業で学生に Model C で自動採点させるために使用するという理由から，Model C 用データベースには加えていない．

4. 2. 機械学習入力用英文のトピック

機械学習の入力データとなる 2020，2021，2022 年度及び 2023 年度第 1 回のライティングテストは，すべて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．

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 20 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」（意見文）である．意見文（Opinion Essay）は英語検定試験において頻繁に出題されるジャンルであるため，実用面からもそうした試験に準拠したトピックとした．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．以下の表現で書いてください．
>
> The place I would like to visit most is（　　）. There are three reasons.

年 3 回のライティングテストとも同じテーマを用いている．ただし学生には必ず 3 回とも別な場所を選んで書くよう指示し，同じ場所が選ばれている英文には一切加点されないことを伝えている．

2023 年度 1 回目のライティングテストは，あらかじめ配布した用紙に 20 分間手書きさせた上で，試験終了後に，manaba の画面に手書きした英文を打ち込ませた．試験時間および試験方法が今年度変更になった理由は，タイピングに不慣れなことが原因で本来の実力が発揮できないこ

とを解消するためである．エッセイの指示文は 2022 年度までのライティングテストと同じである．

4. 3. 学生英文の評価基準

機械学習の「教師あり学習」で用いる教師評価は，「内容の質」に関する以下の 4 つの評価基準 Level 1 から Level 4 で評価した．ただし，未完のもの，理由が 3 つ無いもの，トピックが違うもの，単語羅列で意味が伝わらないものなどは，最低限の修正を施し Level 1 以上の英文にリライトした[5)]．「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．「内容の質」については，Wiseman（2012）のライティング・ルーブリックにおける Topic Development という評価項目を応用している．

Level 1：Detail 文が無いもの
Level 2：Detail 文が少しあるが内容が限定的なもの
Level 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
Level 4：Level 3 の英文の中で特に優れているもの

以下は各評価のサンプル英文である．「内容の質」の評価であるため，文法やスペリングのエラーが含まれていても，自動採点システムでは採点の対象としていない．以下は 1 から 4 の各レベルの英文サンプルである．

① Detail 文が無いもの「Level 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「Level 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「Level 3」例
The place I would like 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 Korean dramas. Second, I want to buy Korean cosmetics. They are so good for

my skin. So, I often use Korean cosmetics. Finally, 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. For these reasons, I would like to visit Korea.

④ 「Level 3」の中で特に優れているもの「Level 4」例

The place I would like to visit most is Australia. There are three reasons. First, I love animals. Australia is a great place to get close to rare animals, such as koalas and kangaroos. My parents showed me a picture of them holding a koala and a crocodile when they traveled to Australia a long time ago, so I want to hold a koala or a crocodile too. Second, I want to visit the many World Heritage sites, for example, Ayers Rock, the Great Barrier Reef, and the Opera House. In particular, I want to see an orchestra concert at the opera house because Australia is famous for classical music. Finally, it is summer in Australia when it is winter in Japan. I don't like the cold. Therefore, I would like to spend time in Australia during the winter season in Japan. Moreover, the time difference between Australia and Japan is about an hour or two. I've never traveled overseas before, so I'm glad that there isn't much of a time difference. For these reasons, I would like to visit Australia.

4. 4.　自動採点システムの表示画面

今回開発した自動採点システム（Model C）の表示画面は以下の図 1，図 2 の通りである．図 1 は英文入力前の画面，図 2 は英文入力・自動採点後の画面である．

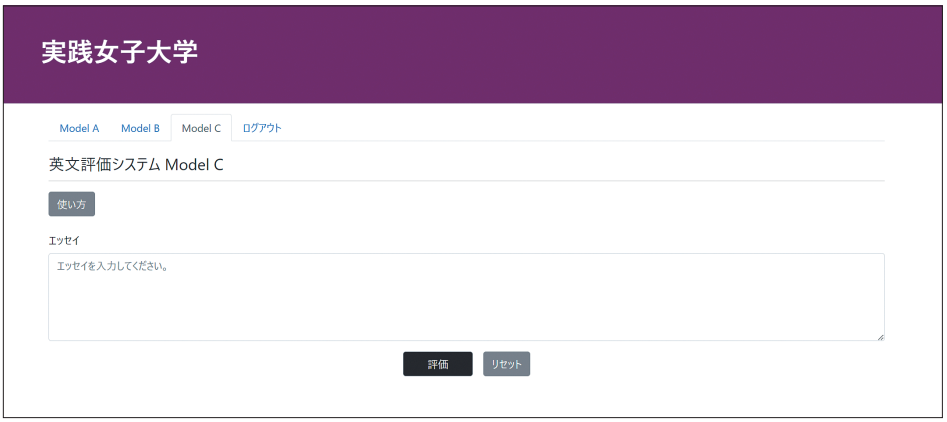

図 1　英文評価システムの英文入力前画面

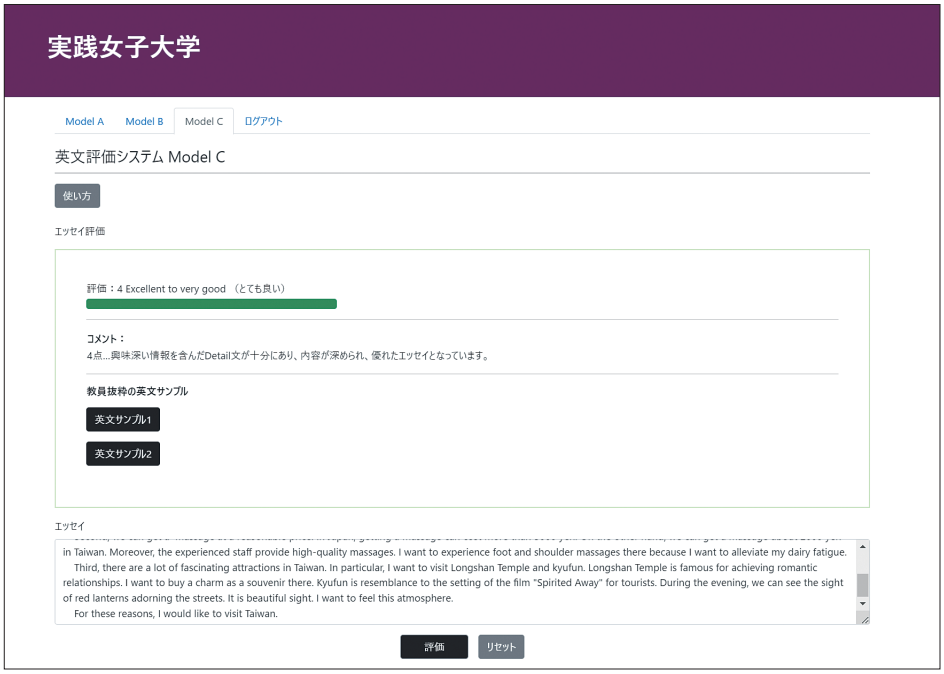

図２　英文評価システムの英文入力・自動採点後画面

表３は，英文の採点と同時に表示されるレベル別のフィードバック表現である．

表３　自動採点システムに表示されるフィードバック表現

Level	自動採点システムのフィードバック表現
1	評価：1 Very poor（良くない） コメント：1 点…Detail 文がありません．理由の詳細や具体例をそれぞれの理由に付けてみましょう．
2	評価：2 Fair to poor（あまり良くない） コメント：2 点…Detail 文が平凡です．読んだ人に印象が残るような情報（例えば自分の経験や皆が知らないような情報など）を入れてみましょう．
3	評価：3 Good to Average（良い） コメント：3 点…興味深い情報を含んだ Detail 文が複数あり，内容が深められています．
4	評価：4 Excellent to very good（とても良い） コメント：4 点…興味深い情報を含んだ Detail 文が十分にあり，内容が深められ，優れたエッセイとなっています．

第２期モデル（Model B）開発の際，自動採点システムで出力される評価得点と同時に表示される自動フィードバックとして，過年度の学生が同じトピックで作成した英文を，英文サンプル（典型事例）として提示する機能を追加した（図３）．Model A, B どちらの出力画面にも付加し（三田・霜田, 2023b），Model C にも継承された．これにより自動採点システム Model A, B, C

は，教師の自動採点ツールとしてだけではなく，学生の形成的評価につながるツールとしても機能する可能性を持つことになった.

図 3　英文評価システムのモデル英文画面

4. 5. 分析方法

2020 年度 1 回目から 2023 年度 1 回目までのライティングテストのデータを基に開発された自動採点システム Model C を，2023 年 9 月の授業で学生に使用させた．その上で以下の 3 つの分析を行った.

(1)　今回開発した Model C の自動採点と教師採点との一致率
(2)　自動採点と教師採点で差が大きかったものの理由の分析
(3)　自動採点システムを導入する前と導入した後の学生の英文の変化の分析
(4)　学生アンケートの自由記述のテキストマイニング分析

5. 調査結果

5. 1. 自動採点と教師採点の一致率と相関係数

2023 年 9 月の後期授業開始時に学生に Model C を使用させた．学生の英語レベルは CEFR A2 中心である（注 6)[6]．学生には 7 月に実施したライティングテストで自ら作成した英文エッセイを自動採点システムで採点させた．表 4 は，学生英文 106 件を Model A, B, C で自動採点した結果と教師採点の一致率である．

表 4　2023 年 7 月の学生英文についての自動採点と教師採点の一致率

	一致英文件数	不一致英文数	英文総件数	一致率（%）
Model A	50	56	106	47.2%
Model B	60	46	106	56.6%
Model C	52	54	106	49.1%

表 5 は，Level 1 から Level 4 の 4 段階評価の自動採点と教師採点の平均と標準偏差である．

表 5　自動採点と教師採点の平均と標準偏差（N=106）

	M	*SD*
Model A	3.104	1.0413
Model B	2.849	.9339
Model C	3.160	.8634
教員採点	2.811	.7574

表 6 は，4 段階評価の自動採点と教師採点の相関係数である．相関係数は 1% 水準で有意である．4 段階評価の Model A, B, C の自動採点と教師採点の相関係数は，いずれも「比較的強い相関がある」（0.4〜0.7）ことを示している．

表 6　自動採点と教師採点の相関係数（N=106）

	Model A	Model B	Model C	教師採点
Model A	1	.672**	.649**	.605**
Model B	.672**	1	.644**	.592**
Model C	.649**	.644**	1	.600**
教師採点	.605**	.592**	.600**	1

**$p < .01$

表 7 は，106 件の英文の Level 1 から Level 4 の自動採点と教師採点の件数である．

表 7　2023 年度 7 月英文件数

	Level 1	Level 2	Level 3	Level 4	計
教師採点	0	42	42	22	106
Model A	8	28	15	55	106
Model B	7	34	33	32	106
Model C	7	11	46	42	106

表 7 の教員評価で Level 1 の英文が 0 件である理由としては，英文の中にわずかでも Detail 文に相当する内容が含まれている場合には，Level 2 と採点するという教員評価の決め事があることが影響している．例えば次のような語数がわずかな英文では，自動採点システムは Level 1 と採点しているが，教員採点では以下の下線部のように Detail 文が含まれているため，Level 2 となる．

自動採点が Level 1，教師採点が Level 2 の英文

I'd like to go to Korea. There are three reasons. First, I like K-POP. If I can go to Korea, I want to go to the place rerated K-POP. Second, I want to eat authentic delicious Korean food. Third, Korea is easy to go from Japan even if it's your first abroad. This is the reason the place I'd like to go to the most.

表 8 は自動採点と教師採点の差によるエッセイの数である．Model C は 106 件のうち，52 件で自動採点と教師採点が一致しており，また差が 1 の採点が 49 件，合わせて 101 件（95.3%）が差 1 以内となる（表 9）．

表 8　自動採点と教師採点の差による英文エッセイの件数

教師採点との差	Model A	Model B	Model C
0	50	60	52
1	47	40	49
2	9	6	5
3	0	0	0
合計	106	106	106

表 9　教師採点との差が 1 以下のエッセイ件数の割合（%）

Model A	Model B	Model C
91.5	94.3	95.3

5. 2. 自動採点と教師採点の差が大きかったもの

Model C の自動採点と教師採点で評価 Level に 2 以上の差のあった英文は 5 件であった（表 10）．

表 10　自動採点と教師採点の差ごとのエッセイ数その差

自動採点－教師採点	Model A	Model B	Model C
0	50	60	52
1	32	18	38
2	8	5	5
3	0	0	0
-1	15	22	11
-2	1	1	0
-3	0	0	0
合計	106	106	106

表 11 は差が 2 以上だった 5 件の英文の Model C の自動採点と教師採点である．いずれも自動採点が 4，教師採点が 2 であった．

表 11　評価 Level の差が 2 以上だった英文の Model C の自動採点と教師採点

No.	自動採点	教師採点	差
1	4	2	2
2	4	2	2
3	4	2	2
4	4	2	2
5	4	2	2

この結果から自動採点が教師採点より高い点数を出す傾向が読み取れる．以下の 5 つの英文は自動採点が教師採点より 2 点高かった（自動採点 4，教師採点 2）英文の例である．

以下のエッセイ 1 は，3 つ目の理由に同じ内容の記述が繰り返されていたため教師採点を 2 点としたケースである．自動採点では内容の重複が採点に反映されない可能性がある．

【エッセイ 1】

The place I would like to visit most is Korea. There are three reasons. The first is that Korean food is very spicy and very much to my liking. Second, I want to go to Starbucks in korea because they have a lot of limited edition food and drinks and limited edition goods. Third, I want to buy a lot of Korean cosmetics. The third reason is I want to buy a lot of Korean cosmetics because Korean cosmetics are very cheap and there are a lot of cute ones. For these reasons, I would like to visit Korea."

以下のエッセイ 2 は，3 つの理由それぞれに Detail 文があり，理由の具体例や詳細が記述され

ており語数的にも 3 点に相当するものであるが，文法エラーが多く，結果として読みにくい英文となっている．「内容の質」の評価においてスペルミスや初歩的な文法エラー（local error）は減点の対象とはしていない．しかし，読み手がエラーを補って憶測して内容を理解しなければならないような global error が多いものは，読み手に心理的負担を与えるため「内容の質」が低下する．そのためエッセイ 2 を教師採点で 2 点とした．エッセイ 2 は，自動採点ではスペリングや文法の誤りに基づく読みにくさの判断ができない例と考えられる．

【エッセイ 2】

I would most like to go to place is Hokkaido. I have three reasons. First, that traveling to easy for me. I don't need a passport and don’t need exchange money. Second, I want to to eat delicious food, for example, miso ramen, seafood and jingiskan. I especially eat seafood. Third, I want to refresh and relax. I always surroundid by buildings in the city. I want to see beautiful scenary. For these reasons, I want to go to Hokkaido.

以下のエッセイ 3 は，3 つの理由それぞれにかろうじて Detail 文があるため教師採点を 2 点とした．しかし「3 つの理由」とテスト出題で指定されているにもかかわらず冒頭が I have two reasons で始まっており，またいくつかのスペルミス，文法エラー，内容の重複があったため，3 点以上の評価とはならなかった．一方自動採点では 4 点である．これもエッセイ 2 のケースと同様，スペリングや文法の誤りを判断できない例と考えられる．

【エッセイ 3】

The plane would most of the Hawaii. I have two reasons. First, I want to swim in the sea because it is beautiful and I want to sea a lot of fishs and corals. Second, I want to eat a lot of Hawaiian foods for example, pancakes and loco moco, hamburgers etc. Finally, I want to get on a sea turtle plane because I like planes especialiy ANA. So, I want to get on a sea turtie plane. For these reasons, The plane would most of the Hawaii.

以下のエッセイ 4 は，特にスペースエラーが多く，2 つの単語がつながっているなど，大変読みにくい．ただし，想像力を駆使して理解すると具体的で詳細な Detail 文が含まれているため教師採点を 2 点とした．しかし内容は評価できるものの，視覚的な読みにくさから 3 点以上を与えることができない．自動採点では，スペルミスと同様に機械的エラー（mechanics）も採点に反映されない例と考えられる．

【エッセイ 4】

why i whot to go to Australia. Becausekoalas are cute and the sweetsare delicious. Above all, it is rich in hature. It also has a good ehviromeht, is hygiehic, and has the world`s largest

coral reet, the Great Barrier Reef. There are mahy world Heritage sites, such as the Great Barrier Reef,the world`s largest coral reef, and the gigahtic Ay you oan see, Australia has aiot to offer I would like to visit Australia.

エッセイ 5 は，3 つの理由に続くそれぞれの Detail 文で具体例と理由が述べられているため教師採点が 2 点以上となる．ただし，文法エラー，特に because の誤用が多発しているため 3 点以上を与えることができない．仮に because エラーを local error と捉え無視するとすれば評価が 3 点となる可能性もある．自動採点が 4 点と採点した理由は不明である．

【エッセイ 5】

The place I would most like to visit Fukuoka. There are three reasons. First, because I want to eat Hakata Ramen. Actually, I like ramen. I want to try ramen in Fukuoka. Second, because I want to visit Itoshima. Especially, the sea is beautiful in Itoshima. Finally, because I want to see a game of the Fukuoka Softbank Howks. Because, there are many famous players in Fukuoka Softbank Howks. For these reasons, I would like to visit Fukuoka.

5. 3. 自動採点システムを導入する前と導入した後の学生の英文の変化

自動採点システムの授業内での使用を開始したのは，Model A が開発された後の 2022 年度後期授業初めである（表 12）．また自動採点システムに英文サンプルの表示機能が搭載されたのは，2022 年度後期末のライティングテスト前からである[7)]．

表 12　自動採点システムの授業内使用開始時期と英文サンプル搭載時期

	授業内使用時期	英文サンプル搭載
Model A	2022 年度後期開始時	開発時は非搭載
	2022 年度後期末	2022 年度後期末より搭載
Model B	使用せず	開発時に搭載
Model C	2023 年度後期開始時	開発時に搭載

以下，表 13, 14 は，Topic Development の 5 段階教員評価（Level 0 から Level 4）の 2021 年度，2022 年度のレベル別エッセイ数，表 15, 16 は両年度のレベル別のエッセイ数の割合を表している．2021 年度は自動採点システムと英文サンプルが授業で導入される前，2022 年度，2023 年度は導入された後の結果である．

表 13　2021 年度のレベル別エッセイ数

2021 年度	Level 0	Level 1	Level 2	Level 3	Level 4	総数
4 月	43	52	53	18	2	168
1 月	21	6	35	66	40	168

表 14　2022 年度のレベル別エッセイ数

2022 年度	Level 0	Level 1	Level 2	Level 3	Level 4	総数
4 月	22	34	67	4	0	127
1 月	4	3	18	40	62	127

表 15　2021 年度のレベル別エッセイ数の割合

2021 年度	Level 0	Level 1	Level 2	Level 3	Level 4	合計
4 月	26%	31%	32%	11%	1%	100%
1 月	13%	4%	21%	39%	24%	100%

表 16　2022 年度のレベル別エッセイ数の割合

2022 年度	Level 0	Level 1	Level 2	Level 3	Level 4	合計
4 月	17%	27%	53%	3%	0%	100%
1 月	3%	2%	14%	31%	49%	100%

表 15，表 16 のように，事前ライティングテストでは Level 4 のエッセイ数の割合が 2021 年度に 1 %，2022 年度に 0 % であった．事後ライティングテストでは，Level 4 のエッセイ数の割合が 2021 年度 24% であるのに対して，自動採点システムと英文サンプルが授業で導入された 2022 年度では 49% と大幅に増えている．

5. 4.　自動採点システムを利用した学生のアンケート分析

自動採点システム（Model C）を 2023 年度学生に使用させ，アンケート調査を行った．学生には 2023 年 7 月の前期最終授業のライティングテストで作成した英文を，9 月の後期初回授業中，自動採点システムで採点させた．自動採点システム使用後に行った記述式アンケート結果についてテキストマイニング[8] を行い，自由回答の頻出語として抽出された単語同士の関係性を可視化するために共起ネットワーク分析を行った．テキストマイニングには，KH Coder 3 を使用した[9]．アンケートの質問は以下の 4 問である．

問 1. 今回の自動採点システムについてよいと思う点を 2 つ書いてください.
問 2. 今回の自動採点システムについて悪いと思う点を 2 つ書いてください.
問 3. 自動採点と共に表示される 2 つの英文サンプルの感想を書いてください. 英文サンプルは, 過年度の先輩のエッセイです. 参考になったと思う人は, 具体的に何が参考になったかを書いてください.
問 4. 今回の自動採点システムについての感想をお聞かせください.

5. 4. 1. 自動採点システムの良いと思う点

質問 1「今回の自動採点システムについてよいと思う点を 2 つ書いてください」に対する学生の自由記述回答の総抽出語数は 2,281 語（181 文）であった. 抽出語の頻出語上位 5 件は「英文」(60 回),「サンプル」(52 回),「自分」(47 回),「評価」(30 回),「参考」(19 回）であった.

語の取捨選択を行わずに共起ネットワークを作成し, さらに共起性の強い線だけの描画に絞る「最小スパニングツリーだけを描画」を選択したところ, 自動採点システムの良い点に関して 5 つのサブグループが形成された（図 4).

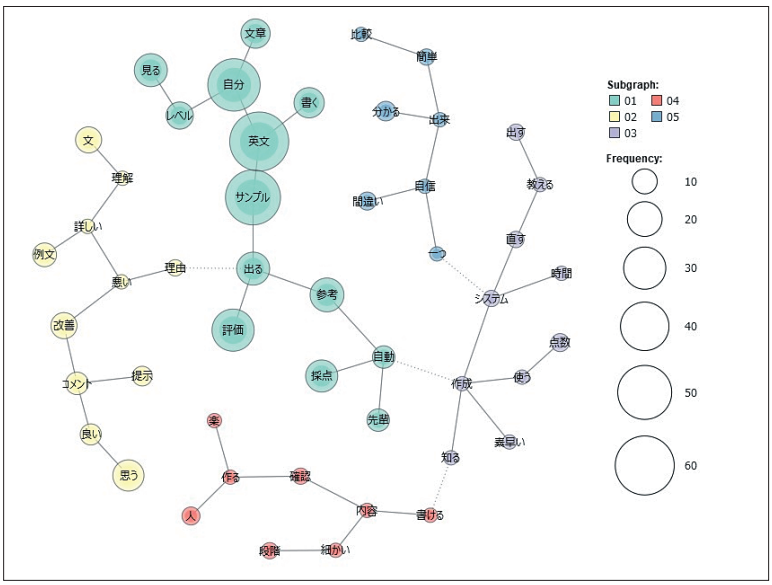

図 4　自動採点システムについて良いと思う点

1 つ目のサブグループでは,「英文」,「自分」,「文章」,「レベル」,「見る」,「サンプル」,「出る」,「評価」,「参考」が示されていることから, 学生は自分の英文のレベルを見ることができ, またサンプルが出ることで評価の参考になることを良い点としていることがわかる. 2 つ目のサブグループには「改善」,「悪い」,「理由」,「詳しい」,「例文」,「理解」,「コメント」,「提示」,「良い」が示されており, 改善のための悪かった理由や詳しい例文, またはコメントが提示されること点を良い点としている. 以下はコメントの抜粋である.

「英文に対してのコメントが簡潔に示されており，ランダムで英文サンプルを閲覧できるので，良いと思った.」

「教員抜粋の英文サンプルの英語サンプルがあり，改善方法がわかりやすい.」

3つ目のサブグループでは「システム」,「直す」,「時間」,「作成」,「素早い」,「点数」,「知る」が示され，学生は自動採点システムで素早い時間で点数が作成されることを良い点としていることがわかる．4つ目のサブグループには「内容」,「細かい」,「段階」,「確認」,「作る」,「楽」が示され，英文内容の細かい段階を確認することができ作文が楽になることを良い点としている．5つ目のサブグループには「比較」,「簡単」,「出来」,「分かる」,「自信」,「間違い」,「一つ」が示されていることから，自分の出来が簡単に比較でき間違いに気付けることを良い点としていることが推測される．以下は上コメントの抜粋である．

「自分の書いた英文を時間をかけることなく，すぐに評価をしてもらい，結果を見ることができるのはとても良いと思いました.」

「評価の段階が細かく分かれている．英文に対する不安を軽くしてくれる．英文を作ることが楽になる.」

「自分の英文に自信を持てる．間違いがないか不安な時に役に立つと思う.」

5. 4. 2.　自動採点システムの悪いと思う点

質問 2「今回の自動採点システムについて悪いと思う点を 2 つ書いてください」に対する学生の自由記述回答の総抽出語数は 960 語（129 文）であった．抽出語の頻出語上位 5 件（同率 3 位 3 件を含む）は「評価」(14 回),「英文」(11 回),「もう少し」(7 回),「採点」(7 回),「自分」(7 回）であった.

語の取捨選択を行わずに共起ネットワークを作成し，さらに共起性の強い線だけの描画に絞る「最小スパニングツリーだけを描画」を選択したところ，自動採点システムの悪い点に関して 10 のサブグループが形成された（図 5）.

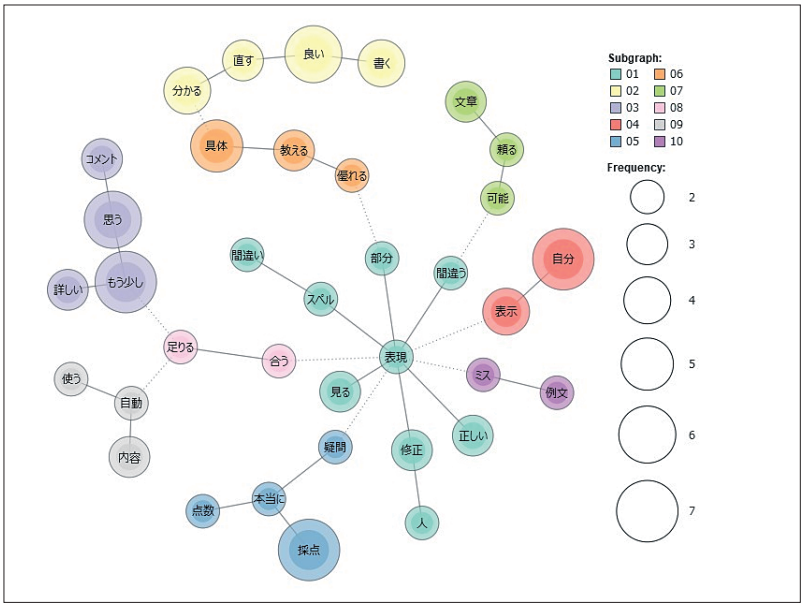

図５　自動採点システムについて悪いと思う点

１つ目のサブグループでは，「表現」，「修正」，「人」，「正しい」，「間違う」，「部分」，「スペル」，「間違い」，「見る」が示されていることから，悪い点として正しい表現は様々あるので人による修正の方が正しいのではないか，またスペルの間違いが指摘されない点を疑問に思っていることが推察される．２つ目のサブグループには「良い」，「書く」，「直す」，「分かる」が示されており，もっと良くなる方法や直し方が書かれていない点を指摘している．３つ目のサブグループでは「もう少し」，「詳しい」，「思う」，「コメント」が示されており，もう少し詳しいコメントが欲しいことを望んでいることが推察される．

「人がやったほうが具体的に英文の修正箇所の指摘ができるのではないかという点がある．
　本当にその採点が正しいのかが分からない．」
「どう直せばもっと良くなるか書いてない．」
「もう少し詳しくコメントや評価をしてほしいと思った．」

４つ目のサブグループには「自分」，「表示」が示され，自分の文章に足りない点が表示されていないことを不満に思っていることが分かる．５つ目のサブグループには「採点」，「本当に」，「点数」，「疑問」が示され，採点点数が本当に正しいのか疑問を持っていることがわかる．６つ目のサブグループには「具体」，「教える」，「優れる」が示され，具体的に何が優れているのか教えてほしいと考えていることが推察される．

「自分の文章に対する，直すべきところの指摘やアドバイスの表示がない.」
「4 月の時点のものも採点にかけたところどちらも評価は 3 でした．本当に厳格な評価がされているのか疑問に思いました」
「どこの部分が特に優れているのか，改善したほうが良いのかを具体的に教えてほしい」

7 つ目のサブグループには「文章」,「頼る」,「可能」が示され，頼りすぎると文章力が落ちる可能性を悪い点と考えていることが分かる．8 つ目のサブグループには「足りる」,「合う」が示され，自分の英文が条件に足りていないのに高いレベルが表示され，評価が合っていないのではないかという疑問を持っていることがわかる．9 つ目のサブグループには「内容」,「自動」,「使う」が示され，自動採点システムを使うことで内容が似てしまう可能性を指摘している．10 つ目のサブグループには「例文」,「ミス」が示され，サンプル例文にミスがあることを指摘している．以下はコメントの抜粋である．

「頼りすぎると逆に文章力が落ちる可能性がある.」
「語数が全く足りていないのにもかかわらず「内容が深められ」という評価は合っていないのではないか.」
「AI なので，みんなが自動採点を使うことで内容が似てきたり，典型的な英文になり一緒になってしまう.」
「例文に翻訳ミスがある .」

5. 4. 3. 英文サンプルの感想

質問 3「自動採点と共に表示される 2 つの英文サンプルの感想を書いてください．英文サンプルは，過年度の先輩のエッセイです．参考になったと思う人は，具体的に何が参考になったかを書いてください」に対する学生の自由記述回答の総抽出語数は 2,677 語（139 文）であった．特徴を読み取りやすくするために「思う」という一般的な語を「語の取捨選択」で「使用しない語」に指定したところ，抽出語の頻出語上位 5 件は「書く」(51 回),「理由」(34 回),「自分」32 回),「参考」29 回),「具体」21 回）であった.

「思う」を外した共起ネットワークを作成し，さらに「最小スパニングツリーだけを描画」を選択したところ英文サンプルに関して 9 つのサブグループが形成された（図 6).

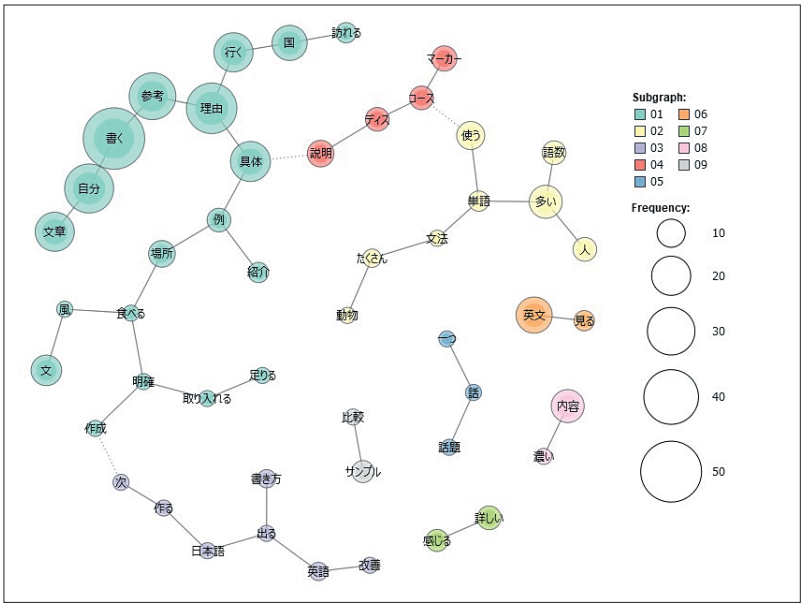

図６　英文サンプルの感想

１つ目のサブグループでは，「書く」，「自分」，「文章」，「参考」，「理由」，「具体」，「例」，「場所」，「紹介」が示されていることから，先輩の文章が自分の書くものの参考になったこと，さらに理由と具体例が紹介されていることでサンプルが役に立っていることが推察される．２つ目のサブグループには「多い」，「語数」，「人」，「単語」，「使う」，「文法」，「たくさん」が示されており，先輩の英文の語数の多さに驚きサンプルに使用されている単語や文法の多様性に気がついたという感想も多く示されている．３つ目のサブグループでは「改善」，「英語」，「出る」，「書き方」，「日本語」，「作る」，「次」が示されており，サンプル英文を読むことで改善が期待され，日本語訳も出るため次にどのように英文を作るかわかると感じたことが推察される．以下はコメントの抜粋である．

「行きたい国の理由をこまかく書いてあったり，実際に体験したことも書いてあって参考になった.」

「語数が多く具体的な例が出せている．行きたい場所や食べたいものについて軽く紹介ができている．ほかの人も行きたくなるような文章が書けている.」

「日本語の文章を読むだけでも，次のレベルになるためにはどのように文章を作ればよいのかがわかるところがよいと思った.」

４つ目のサブグループには「ディスコースマーカー」，「説明」が示され，ディスコースマーカーの使い方や説明の仕方が参考になったことが分かる．５つ目のサブグループには「話題」，「話」，「一つ」が示され，一つの話をどのように展開するかについてヒントを得ることができた

と指摘している．6 つ目のサブグループには「英文」,「見る」が示され，先輩の英文を見ることでどのように書けばいいのか分かったことが推察される．以下はコメントの抜粋である．

「訪れたい国を先に示し，なぜそうなのか具体例を挙げながら説明していたので分かりやすかった．ディスコースマーカーや，字下げを積極的に使っていた．」
「一つの話題や理由についてとても深く話を掘り下げられている．」
「先輩の英文を見ることで自分に足りないところが知ることができて，足りないところを参考にして取り入れていこうと思いました．」

7 つ目のサブグループには「感じる」,「詳しい」が示され，内容が詳しく書かれていることをサンプルから感じ取ったことがわかる．8 つ目のサブグループには「内容」,「濃い」が示され，内容の濃さが参考になったことが推察される．9 つ目のサブグループには「サンプル」,「比較」が示され，自分の文とサンプルを比較することで自分に足りないものに気が付くきっかけとなったことがわかる．以下はコメントの抜粋である．

「理由の内容が詳しく分かりやすく書いてあって読みやすいし，話題の広げ方もとても自然で違和感がないなと感じた．」
「内容が濃く，読んでいて飽きない文になっている．」
「自分の文章とサンプルを比較して，自分の文章には理由にオリジナリティがないということが分かった．」

5. 4. 4.　自動採点システムについての感想

質問 4「今回の自動採点システムについての感想をお聞かせください」に対する学生の自由記述回答の総抽出語数は 2,553 語（131 文）であった．特徴を読み取りやすくするために「思う」という一般的な語を「語の取捨選択」で「使用しない語」に指定し，「自動採点システム」を複合語として強制抽出したところ，抽出語の頻出語上位 7 件（同率 5 位 3 件を含む）は「自分」(41 回),「英文」(39 回),「良い」(23 回),「採点」(21 回),「システム」(19 回),「書く」(19 回),「評価」(19 回）であった．

さらに「最小スパニングツリーだけを描画」を選択したところ自動採点システムを使った感想に関して 8 つのサブグループが形成された（図 7）．

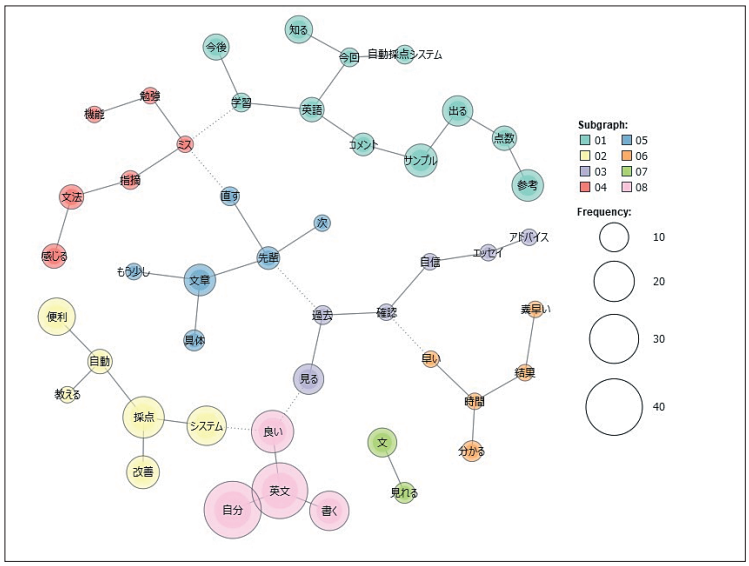

図７　自動採点システムの感想

１つ目のサブグループでは，「参考」，「点数」，「出る」，「サンプル」，「コメント」，「英語」，「学習」，「今後」，「今回」，「自動採点システム」，「知る」が示されていることから，点数とサンプルやコメントの出る自動採点システムが今後の学習に役立つと感じた学生が多かったことがわかる．２つ目のサブグループには「採点」，「システム」，「改善」，「自動」，「便利」，「教える」が示されており，この採点システムが自動で評価を教えてくれることが便利で，また自分の英文作成の改善につながると感じていることが推察される．３つ目のサブグループでは「見る」，「過去」，「確認」，「自信」，「エッセイ」，「アドバイス」が示され，過去のサンプルエッセイとアドバイスを見ることで自分の英文を確認でき，評価してもらうことで自信につながったことがわかる．以下はコメントの抜粋である．

「採点システムを使ってみて自分が書いた文章がどのくらいの点数なのかを知ることができました．また，過年度の先輩のエッセイがあるおかげでどのように書いたらいい文章になるかを知ることができたのでとても参考になりました．」

「自分で文を評価するのは，なかなか難しいし間違いに気づけないけど自動採点してもらえて，まだ自分の文は改善できるのだなと気づけました．」

「自分が作った英文を一度確認し，評価してもらうことで英文に自信がついた．過去の学生が書いたエッセイを見られ，とても参考になった．」

４つ目のサブグループには「機能」，「勉強」，「ミス」，「指摘」，「文法」，「感じる」が示され，自動採点システムの機能により文法ミスも指摘されれば良いと感じたことが推察される．５つ目のサブグループには「具体」，「もう少し」，「文章」，「先輩」，「次」，「直す」が示され，もう少し

具体的に文章を書いた方が良いことを先輩のサンプルから学び，次に直すことを意識できたことがわかる．6 つ目のサブグループには「素早い」,「結果」,「時間」,「早い」,「分かる」が示され，素早く採点結果が出ることと改善点がわかりやすいことが良いと評価していることが推察される．以下はコメントの抜粋である．

「英文についての改善点を指摘してもらえるようになると，より良いシステムになるのではないかと感じました.」
「今回の自動採点システムによって，もう少し文章を具体的に書いたほうが良いということを気づくことができた.
「自分の書いた英文を時間をかけることなく，すぐに評価をしてもらい結果を見ることができるのはとても良いと思いました.『英文を書いてみたけど正解が分からない …』という不安がなくなる.」

7 つ目のサブグループには「文」,「見れる」が示され，先輩のサンプル文が見られることが役に立ったことがわかる．8 つ目のサブグループには「自分」,「英文」,「書く」,「良い」が示され，自分の英文を書く上で良いシステムだと思ったことが多数見られた．以下は各サブグループのコメントの抜粋である．以下はコメントの抜粋である．

「自分の作った文が採点されてコメントと，英文サンプルが見れるのは英語力向上につながりそうだと思いました.」
「正しい文章を書きたいと思ったときやテストの前などに，家でも簡単に英文を書く練習ができそうでいいシステムだと思いました.」

6. 考察

本研究では 2020 年度 1 回目から 2023 年度 1 回目までの計 10 回分のライティングテストの英文 1506 件を機械学習の「教師あり学習」の入力データとして用いて自動採点システム Model C を開発した．またそれを 2023 年度英語必修科目の授業内で学生に使用させた．この自動採点システムは，特定の教育現場で用いるための自動採点システムである．Amazon のクラウド上のサービスを利用し，技術者の協力により開発されたスモールスタートの AI モデルである．

自動採点システムで採点する項目は英文の「内容の質」に限定している．すなわち主張を補足し説明する Detail 文の充実度によって Level 1 から Level 4 までの採点を行うシステムである．ライティング評価では欠かせない文法エラーやスペリングエラーは，この自動採点システムでは取り上げていない．

学生たちには 2023 年度前期最終授業で実施したライティングテストの自分の英文を Model C に入力させ、採点結果を確認させた．リサーチクエスチョン（1）「学生のライティング英文の教師による採点と自動採点システム（Model C）による採点の一致率はどの程度であるか」につい

ては，2023 年度の英文の計 106 件の自動採点が教師採点と一致した割合は 49.1% であった．Model C の自動採点と教師採点の相関係数は，.600（1%水準で有意）で，「比較的強い相関がある」（0.4〜0.7）ことを示している．

Model C は 106 件のうち，52 件で採点結果が教師採点と一致しており，また差が 1 の採点が 49 件，合わせて 101 件（95.3%）が教師採点との差 1 以内に収まった．Model C の自動採点と教師採点で差が 2 以上の英文は 5 件のみであった．5 件いずれも Model C の採点が 4，教員採点が 2 であった．5 件で採点が異なる理由は，自動採点の評価に影響を及ぼさない「内容の重複」「global error による読みにくさ」「定型表現からの逸脱」「機械的エラー（mechanics）による視覚的読みにくさ」「同じ文法エラーの多発」に起因するものと思われる．

リサーチクエスチョン（2）「自動採点システムを授業で導入する前と導入した後の学生の英文にはどのような変化が見られるか」について，自動採点システムが導入される前と後で「内容の質」の高評価エッセイ数にエッセイ数に大きな違いが見られた．表 15 と表 16 は自動採点システムと英文サンプルが授業で提供される前の 2021 年度と、授業で提供されるようになった 2022 年度の各レベルの事前・事後のエッセイ数の割合を示している。1 月の Level 4 のエッセイの割合が、2021 年度は 24%であるのに対し、2022 年度は 49%と、倍以上に増加していた。このような「内容の質」向上の要因としては，2022 年度から導入された以下の 3 つの授業内容が影響していることが考えられる．

1）自動採点システムで Level 1 から Level 4 の採点結果が瞬時に表示されるため，学生自身が目指すべきレベルが明確になる
2）採点結果と同時に表示される 1 つ上のレベルの英文サンプルが英文入力ごとに 2 つずつ表示されるため，内容の質向上のヒントが得られる
3）ライティングテストで用いる定型表現を指定にすることにより，パラグラフ・ライティングで英文が書きやすくなる

上記 3）の定型表現の指定については，2022 年度授業で定型文として “The place I would like to visit most is（　　）. There are three reasons. First, Second, Finally, … For these reasons, I would like to visit（　）.” を覚えてライティングテストに臨むように強調した．またその定型表現を入れたライティングテストルーブリックを前期初めの授業で配布している．藤田（2023）は，高校 2 年生に対する実践例として，文章やパラグラフの最初の部分や特定の表現を与えることが，英文を書くことがそれほど難しいことではないと生徒に思わせるための足場掛け（scaffolding）となることを紹介している．定型表現の提示が，学生が英文の「内容」に思考を集中するための一助となっている可能性がある．

今回開発した自動採点システム Model C は，2020 年度から 2023 年度 1 回目までのライティングテストの英文データを用いて開発した．今後さらに 2023 年度 2 回目と 3 回目のデータを加えた Model D を開発し，一致率の変化を見ていきたい．

リサーチクエスチョン（3）「自動採点システムを利用した学生のアンケートから示唆されるものは何か」については，4 項目の記述アンケートから得られた学生のコメントについてテキストマイニングにより内容分析を行った．まず，自動採点システム（Model C）について良いと思う点として，多くの学生が自分の英文の改善に役立つとして得点とともに示されるコメントと先輩のサンプルを挙げている．特に，自分の英文内容の改善を先輩のモデル英文から学んだというアンケートのコメントが圧倒的に多いことから，フィードバックとして搭載した先輩のサンプルが自動採点システムの形成的評価ツールとして大いに機能していることが確認できた．また，評価点数を瞬時に知ることができることも良い点として多く指摘されている．教師だけでなく学生にとっても素早い評価結果は学習促進の点で有効であることがわかった．

一方，悪いと思う点としては，スペリングや文法の間違いが指摘されないことが挙げられた．この点に関しては，この自動評価システムでの採点は英文の「内容の質」に限定しており，文法エラーやスペリングエラーは，この自動採点システムでは評価しないという点を学生が理解していないことによるものであるため，使用前に「内容の質」に限定した評価点であることを周知しておく必要がある．また，自分の英文がもっと良くなる方法や直し方をもっと詳しく具体的に知りたいという意見も見られ，フィードバックとして搭載した先輩のサンプル英文から自分で改善点を見つけられない学生に対してのほかの方法の可能性を今後の課題としたい．また，評価点数が本当に正しいのか疑問に思うという意見もあり，この点に関しては，自動採点システムの精度を上げることを目指したモデルの開発を続けていかなければならない．

自動採点と共に表示される英文サンプルの感想としては，アンケート項目の 1 つ目「良い点」でも多く指摘があった通り，先輩のサンプルを読むことで様々な点を参考にしていることがわかった．アンケートの中の「具体的に何が参考になったか」の問いに対して，「理由や具体例の書き方」「語数」「単語や文法の多様性」「ディスコースマーカーの使い方」「一つの話題の展開方法」「詳しく濃い内容の書き方」「比較することで自分の英文に足りないもの」など，学生は多くの気づきを先輩のサンプルから得ていることが確認できた．今回の調査で，フィードバックとしての良い見本（典型例文）の効果を確認することができた．今後さらに有効なフィードバックを探求していきたい．

最後に，今回使用した自動採点システムの感想については，この自動採点システムが今後の英語学習に役立つ，また英文作成の改善につながる，今後も活用したいとする意見が多数見られた．一致率が不十分ながら，おおむね学生に好意的に受け入れられたようである．

今回試みた自動採点システムに関する学生のアンケート結果から多くの示唆を得ることができた．これまでライティング研究において「内容の質」を正確に評価することの難しさが指摘されてきた．本研究でもそれを乗り越えることを目的として，人工知能の機械学習による自動評価システムの開発と実践を試みてきたが，その評価の正確さという点では不十分である．しかしその副産物として搭載したフィードバックが，学生のパラグラフ・ライティングを大いに改善させるという成果を得ることができた．「最新技術の進歩を目の当たりすることができて感動した．これからもっと進歩すると思うと楽しみだ．」といった学生の感想を受けとめ，今後も引き続き自

動採点システムの開発と改善に取り組んでいきたい.

7. 終わりに

短期大学英語必修科目のライティングテストの英文データを用いて人工知能の機械学習を行い，特定の授業での限定使用の自動採点システムを開発した．これまで英文ライティング指導でネックとなっていた教師の採点負担を軽減し，また学生が自ら英文を修正する動機付けとすることを目的として，2020 年度より収集した学生の英文エッセイ 1506 件をデータとして用いて Model C を開発した．自動採点システムの採点結果と同時に表示される英文サンプルが，学生の英文ライティング力向上に活かされていることが示唆された．今後も自動採点システムが学生にとってより使いやすく有効なものとなるよう検討を続けていきたい.

謝辞

本研究は科学研究助成基金基盤研究 ©（課題番号 18K00814）の助成を受けたものである．本研究を進めるにあたり，株式会社ルーティングシステムズの大庭裕司氏，成田康孝氏から数多くの技術的サポートやアドバイスを受けた．ここに感謝の意を表する.

〔注〕

1. Integrated English は，短期大学の全学必修英語科目で 1 年前期に Integrated English a（1 年前期），Integrated English b（1 年後期）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 コマは外国人教員が担当する.
2. CLI（Coleman-Liau Index）は，リーダビリティー（語彙の難易，文長等の文体による読みやすさ）を表す指標である.
3. EFT（Error-free T-unit）平均語数は，グローバルエラーと T-unit から算出した数値で，その数値が高いほど書き手は統語的に熟達した作文を書く力を有する.
4. この業者はオンラインで英文添削サービスを提供している．世界中に英文添削講師を抱え，11 年の実績がある.
5. 以下はリライトされたサンプルである.
   (2021 年 7 月)：
   リライト前（文の途中で終わっているので 0 評価，75 語，Level 0）
   The place I want to visit most is Disneyland. There are three reasons.
   First, I have only been Disneyland 3 times in my childhood. So I want to go there to see the difference from when I was a child.
   Second, I saw Disney movie lately and I though If I go to Disneyland, I can meet to the Disney character. I heard that there is a "Greeting", so I definitely want to go.
   Finaliy,
   リライト後（3 つ目の理由を追加，86 語，Level 3）
   The place I want to visit most is Disneyland. There are three reasons.
   First, I have only been to Disneyland 3 times in my childhood. So I want to go there to see the difference from when I was a child.
   Second, I saw a Disney movie lately and I thought if I go to Disneyland, I can meet the Disney character. I heard that there is a "Greeting", so I definitely want to go.
   Finally, I want to ride the attractions. They must be fun.
6. CEFR とは，Common European Framework of Reference for Languages の略称である．CEFR A2 は，学生が入学時に受検する GTEC Academic の結果に基づいた英語レベルである．GTEC®（ジーテック / 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.
7. Model B が使用されなかったのは，2022 年度 7 月の学生英文を入力した結果，一致率が Model A の方が僅かに高かったためである（Model A は 60.5%，Model B は 60.1%）.

8. 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．
9. KH Coder とは，計量テキスト分析またはテキストマイニングのためのフリーソフトウェアである．

〔参考文献〕

石井雄隆・近藤悠介．(2020a).「自動採点研究とは？」石井雄隆・藤悠介（編）『英語教育における自動採点 --- 現状と課題』1–15. ひつじ書房.

石井雄隆・近藤悠介．(2020b).「教室における指導と自動採点」石井雄隆＆近藤悠介（編）『英語教育における自動採点 --- 現状と課題』117–130. ひつじ書房.

岩田貴帆．(2020)．協議ワークを取り入れたピアレビューによる学生の自己評価力向上の効果検証．*大学教育学会誌*，*42*(1)，115–124.

岩田貴帆．(2022, 6 月)．学生の自己評価力を向上させうる教授法の特徴と課題の検討．*大学教育学会口頭発表*，岡山理科大学.

川西慧．(2023)．AI Chatbot と外国語ライティング―ChatGPT は効率の他に何をもたらすか．*JACET 関西支部ライティング指導研究会紀要, 15*, 99–108

小林雄一郎・石井雄隆．(2019)．英語ライティング指導のための自動フィードバックシステムの開発に向けて．*日本大学生産工学部研究報告．B，文系／研究報告専門委員会 編*，*52*, 7–15.

竹ノ内朋子．(2023)．ChatGPT を利用した英検 2 級のライティング分析．*外国語教育メディア学会（LET）関西支部メソドロジー研究部会報告論集, 14*, 63–69.

丹原惇・斎藤有吾・松下佳代・小野和宏・秋葉陽介・西山秀昌．(2020)．論証モデルを用いたアカデミック・ライティングの授業デザインの有効性．*大学教育学会誌 = Journal of Japan Association for College and University Education, 42*(1), 125–134.

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）.

藤田眞理子．(2023)．高校 2 年生の英語表現 II におけるライティングの授業．*JACET 関西支部ライティング指導研究会紀要, 15*, 17–29.

三田薫・霜田敦子．(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・霜田敦子．(2021a)．学生の習熟度別英文ライティング力向上の分析―弱点克服の重点的指導によるライティングの変化―．*実践女子大学短期大学部紀要 = Jissen Women's Junior College Review, 42*, 63–83.

三田薫・霜田敦子．(2021b)．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・霜田敦子．(2022a)．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．*実践女子大学短期大学部紀要 = Jissen Women's Junior College Review, 43*, 65–83.

三田薫・霜田敦子．(2022b)．学生の英文ライティング力向上の分析 その 3：文法・構造・論理・内容の質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, 52*, 13–48.

三田薫・霜田敦子．(2023a)．英語初級学習者のパラグラフ・ライティングのための自動採点システム開発の試み．*実践女子大学短期大学部紀要 = Jissen Women's Junior College Review, 44*, 39–67.

三田薫・霜田敦子．(2023b)．学生の英文ライティング力向上の分析 その 4：ルーブリックを用いた指導と「内容の質」を測る自動採点システム導入によるライティングの変化 . *Jissen English Communication, 53*, 2–35.

柳瀬陽介．(2022).「機械翻訳はバベルの塔を築くのか―大学教養課程での英語ライティング授業からの考察―」『ことばと社会』24, 43–63.

柳瀬陽介．(2023). AI を活用して英語論文を作成する日本語話者にとっての課題とその対策．*情報の科学と技術*, *73*(6), 219–224.

Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics, 2*(2), 100050.

Sadler, D. R. (1987). Specifying and promulgating achievement standards. *Oxford review of education, 13*(2), 191–209.

Van Beuningen, C. G., De Jong, N. H., & Kuiken, F. (2012). Evidence on the effectiveness of comprehensive error correction in second language writing. *Language Learning, 62*(1), 1–41.

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.