

AIを活用した自学自習支援システム

—条件英作文の効果的なフィードバック—

小久保 知咲

1. はじめに

英語のライティング能力は、現代のグローバル社会で重要なスキルとされる。特に、条件英作文は高校入試や英語検定で頻出し、文法・語彙の知識だけでなく、条件の理解や論理的な文章構築が求められる[1]。そのため、適切なフィードバックを通じた継続的な学習が不可欠である。

従来の学習では、教員が添削し、誤りの指摘や改善案を提示してきた[2]。しかし、この方法は時間や労力の制約があり、大人数への十分なフィードバックが難しい。特に、条件英作文では文法や語彙に加え、条件適合性の評価も必要であり、単純な自動添削ツールでは対応が難しい。

近年、オンライン学習の普及とともに、AIを活用した自学自習システムが注目されている[3]。AIにより即時かつ個別最適化されたフィードバックが可能となり、従来の指導では難しかった学習支援が期待される。しかし、市場の多くのシステム[4][5]は自由英作文向けであり、条件英作文のフィードバック機能が不足している。

本研究では、条件英作文に特化した自学自習システムを開発し、学習者が自律的にスキルを向上できる仕組みを提案する。生成AIを活用し、文法や語彙の正確さに加え、条件適合性や論理構成を評価し、具体的な改善案を提供することを目的とする。

2. 条件英作文の自学自習システムの現状

2.1 条件英作文の自学自習の概要

条件英作文は、多くの高校入試や英語検定で頻出する問題形式の一つである。条件を理解し、適切な英文で表現する力が必要となるため、条件英作文は学習者にとって高度な能力を求められている。そのため、単なる正誤判定ではなく、適切なフィードバックが重要となる。フィードバックが学習者の気づきを促し、改善に向けた具体的なアドバイスとなるためである。

横野[2]によれば、「英文の添削は、現実には多大な時間と労力を要する作業である」と指摘されている。学習者の人数が多い場合、教員がすべての解答を丁寧に添削することは時間的な制約から困難であり、結果として多くの学習者が十分なフィードバックを得る機会を失うことになる。このため、条件英作文の学習を個別に効率よく支援する仕組みが必要であるといえる。

2.2 自学自習システムの必要性

近年、オンライン学習環境の普及は[3]、学習者にとって新たな利便性を提供する一方で、従来の教員による添削指導に依存する方法では対応しきれない側面もある。AIを活用した自学自習システムは、このような学習ニーズに応

えるための有力な手段として注目されている。AI技術を用いることで、学習者の解答に対して即座にその誤りを指摘したり、具体例の提示や、より高度な解を示したりすることで、学習者は効率的かつ効果的に学習を進めることができる。加えて、教員の負担も軽減される。

2.3 条件英作文に特化した自学自習システム

現在サービス提供されているAIツール[4][5]などでは、学習者が英作文を作成し、それに対してAIがフィードバックを提供する機能が備わっている。しかし、これらのシステムは自由英作文を対象としており、条件付きの問題に対応していない。そのため、条件を満たしているかの評価や、具体的な改善案の提示が不十分である。また、文法や語彙だけでなく、条件に基づいた論理構成の評価も重要だが、現行のシステムでは対応が難しい。このような課題を解決するために、条件英作文に特化した自学自習システムを開発する必要がある。本研究では、生成AI¹を活用して条件英作文の学習を支援し、学習者に対して即時的かつ適応的なフィードバックを提供するシステムを提案する。

3. 自学自習システムの構成

3.1 使用するツール

本研究のシステムはPythonを用いて開発し、Web上で動作するアプリとして構築した。これは、学習者がインターネット接続さえあれば、時間や場所に制約されることなく条件英作文の学習を進めることを可能とするためである。開発にはPythonのWebアプリフレームワークであるDjango²を採用し、問題の出題やフィードバックの提供を含む動的なWeb機能を効率的に実現した。また、条件英作文の解答に対する適切なフィードバックを提供するために、OpenAI社のChatGPT API³(GPT-4-miniモデル⁴)を活用した。

3.2 GPTを用いた自学自習システム

本研究では、条件英作文の学習支援において、GPT4-miniのAPIを用いた自学自習システムの開発をする。ChatGPTは大規模な言語モデルをベースとしており、ChatGPT APIを使うことで、ユーザの入力に対して、開発したシステム上でリアルタイムに文章を作成することができる[6]。これを利用することで、学習者が提出した条件英作文の解答に対して迅速かつ的確なフィードバックを提供することが可能となる。

本システムでは以下の2つの特徴を重視して設計する。

1. 即時性: 学習者が入力した解答に対して即時フィードバックを行うことで、効率的な学習を実現可能とする。
2. レベルに応じた適応性: 初級・中級・上級の3つレベルのフィードバックを実装し、幅広い学習者に対して適応

¹ 話、ストーリー、画像、動画、音楽などの新しいコンテンツやアイデアを生み出すことのできる AI の一種

² <https://www.djangoproject.com/>

³ ソフトウェアやアプリケーション同士をつなぐ仕組み

⁴ <https://openai.com/ja-JP/index/gpt-4o-mini-advancing-cost-efficient-intelligence/>

的なフィードバックを提供する。

このようにGPTを活用することで、従来の添削指導に匹敵する質のフィードバックを自動で提供できるため、学習者が一人で効率的かつ効果的に条件英作文を学ぶことが可能となる。

3.3 システムの構成

本研究のシステム構成を図 1 システム構成に示す。まず、学習者は習熟レベルを選択する(A)。レベルの選択を行うことで、幅広い学習者に対してのフィードバックが期待され、充実した自学自習支援が可能となる。次に、問題文リストから無作為に出力された問題文を学習者に表示し回答を促す(B)。問題文リストとは、問題文が10問分格納されたファイルである。入力された学習者の解答は問題文と共にChatGPT APIに送られる(C)。入力された解答を解析し、学習者の習熟レベル別に、文法や語彙、条件を満たしているかどうかのフィードバックをGPTが行い、学習者に表示する(D)。その後、次の問題文を新たに出力して、学習者は次の問題へと遷移する(E)。

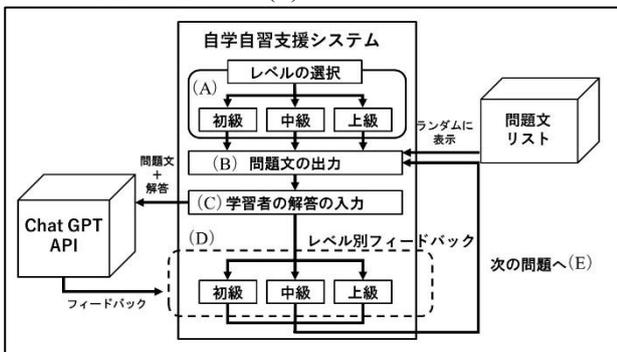


図 1 システム構成図

習熟レベル別のフィードバック項目は、表 1 のように定めた。表中の「○」は、該当する習熟レベルで提供されるフィードバック内容を示し、「-」はそのレベルでは提供されない項目を意味する。

表 1 フィードバック項目一覧

| | 初級 | 中級 | 上級 |
|------------|----|----|----|
| 文法 | ○ | ○ | ○ |
| 論理構成 | ○ | ○ | ○ |
| 例文の提示 | ○ | ○ | ○ |
| 条件適合性 | ○ | ○ | ○ |
| 高度な語彙の提案 | - | ○ | ○ |
| 自然な表現の例文 | - | ○ | ○ |
| 表現のニュアンス解説 | - | - | ○ |
| 応用問題の提供 | - | - | ○ |

初級、中級、上級、の各レベルでフィードバック項目を分けた理由は、学習者の習熟レベルに応じた適切なサポートを提供し、効率的な学習を促進するためである。特に、「文法」、「論理構成」、「条件適合性」、「高度な語彙の提案」については、これらが実際の英検採点基準として存在する[7]ことから本研究のフィードバック項目として採用した。また、「自然な表現の例文」や「表現のニュアンス解説」については、既存の英作文採点システムにおいて実装されているため[4][5]、本研究でも扱うこととした。

⁵ 評価値 0.5 を適用するのは「条件適合性」に限る。他の

学習者の習熟レベルに応じたフィードバックの差異を設けた例としては、中級および上級の学習者にのみ提示される項目「高度な語彙の提案」がある。初級者にとってこの項目は学習負荷が大きすぎると考えられるためである。一方で、中級の学習者には表現の幅を広げるために効果的な語彙を提案し、上級の学習者にはさらに洗練された表現を学ぶ機会を与える。これにより、学習者のレベルに応じた適切なサポートが可能となる。

このように、習熟度別のフィードバック項目を設定することで、既存の基準やシステムを参考にしつつ、学習者のレベルに応じた適切な指導を行い、効果的な学習をサポートできると考える。

4. 自学自習システムのフィードバックの評価

4.1 評価実験の手順

本研究における自学自習システムのフィードバックの精度を検証する。まず、高校入試の英語科目において条件英作文として出題された問題をインターネット上のサイトより引用し[8]、10個の問題文を用意する。問題文の条件数は最大2個である。これらに対して、フィードバック項目ごとにそれぞれ解答パターンを作成した。本研究では解答パターンを用いることにより、以下の2点においてフィードバックの精度を客観的かつ詳細に検証することを目的としている。

1. 正答を誤答とみなすことなく、正答であると判断できているかの確認
2. 誤答を正答とみなすことなく、誤りを適切に指摘できているかの確認

フィードバックの精度は表 2 で定義する評価値に基づき、複数の試行に対する平均評価値によって判断する。

表 2 評価値の定義について⁵

| 評価値 | 評価基準 |
|-----|--------------------------------------|
| 1 | フィードバックが完璧である場合 |
| 0.5 | 2 つある条件の適合性について、1 つのみ正しくフィードバックできた場合 |
| 0 | フィードバックがすべて間違った場合 |

本研究では、「文法」、「論理構成」、「条件適合性」の3項目について、正答と誤答に対するフィードバック精度を評価するために以下のような解答パターンを作成した。

表 3 解答パターン一覧

| | 文法 | 論理構成 | 条件適合性 |
|-----|-----------------|---------------------|-----------|
| 正答 | 正しい解答 | | |
| 誤答1 | 3単元"s"の誤り | 矛盾が生じている、文章のつながりが弱い | 条件1の誤り |
| 誤答2 | 不定詞と動名詞の使い分けの誤り | - | 条件2の誤り |
| 誤答3 | 前置詞の誤り | - | 条件1, 2の誤り |
| 誤答4 | 時制の誤り | - | - |

一方で、先に示した3項目以外のフィードバック項目については誤答パターンを作成しないこととする。いずれも誤りを指摘するというよりは、学習の幅を広げることを目的

フィードバック項目では、部分的な誤りが発生するケースが想定されないためである。

としているため誤答パターンを用意する必要はなく、提示された情報がレベルに対して適切か否かを判断すれば評価可能であるためである。

次に、適切なプロンプト⁶を検証するための予備実験を実施する。ここでは、用いる問題を1問、評価値を参照するフィードバック項目を条件適合性に限定し、プロンプトの内容が精度に及ぼす影響を検証した。

まず、初期に設定したプロンプトでは、条件を正しく満たす正答に対して誤答と判断するフィードバックや、逆に誤答に対して正答と誤判定することが度々発生し、その平均評価値は0.7~0.8であった。このことから、プロンプトが不十分であると考え、以下の修正を加えた。

- 条件適合性の判定や語数カウントなど、誤ったフィードバックの多く発生していた項目に対して、プロンプト内の該当箇所「※」を付与し、単語、文字のカウントの定義を設定し、指示を強調した。
- 問題文における条件部分を正しく認識できるよう、問題文リストにtext(問題文)とconditions(条件)というタグ付けをして、これに基づいて条件適合性におけるプロンプト指示を修正した。
- “指摘”, “提供”などの言葉の表現に対して、“フィードバック”という用語に統一した。
- 条件の指定部分の英文を“”で囲み、プロンプト内で明確に定義した。

これらの修正により、以下の平均評価値の結果が得られ、改善が見られた。

表 4 プロンプト検討結果

| | 初期平均評価値 | 修正後平均評価値 |
|----|---------|----------|
| 初級 | 0.89 | 1.00 |
| 中級 | 0.72 | 0.94 |
| 上級 | 0.83 | 0.83 |

修正を重ね、最終的に設定したプロンプトは図 2のとおりである。レベル別にプロンプトは異なっており、初級レベルのプロンプトでは項目4, 5, 6, 7がなく、中級レベルでは項目6, 7が存在しない。

次に解答パターンを用いた精度検証の実験を行う。本研究では、先に示した解答パターン一覧に基づき「文法」、「論理構成」、「条件適合性」の3項目と、語彙の適正レベルによって定量的評価が可能な「高度な語彙の提案」に焦点を当てて実験を実施した。正答と誤答を含む解答パターンを自学自習システムに入力し、そのフィードバックの適正さを評価した。なお、ChatGPT APIが必ずしも同じフィードバックを返すわけではないため、すべてのフィードバック項目について試行を3回ずつ繰り返した。以下に試行回数をまとめる。

なお、「高度な語彙の提案」に関しては、本研究のシステムが対象とする中学生用に、英検準2級程度の単語を条件とすることをプロンプトで指示した。フィードバックされた語彙を準2級程度と3級以上に分類し、指示から逸脱したレベルの語彙の割合でこれを評価する。

この解答に対して、以下の評価項目・基準に従って上級者向けにフィードバックをお願いします。評価項目は次の8つです。

- 1.文法の評価: 文法的な誤りを的確に指摘し、訂正箇所について正しい修正例を提示してください。
 - 2.論理構成の評価: 論理の流れの矛盾、文と文のつながり、問題に対しての文章の意味が矛盾していないかなどを正確に評価し、問題点を指摘して改善策を提示してください。
 - 3.例文の提案
 - 4.条件の適合度の評価: 問題文と条件を正確に把握し、解答と照らし合わせ、すべての条件が満たされているかどうかを根拠付きで詳細に説明してください。
 - 5.高度な語彙の提案: 日本英語検定準二級程度の単語を提案してください。
 - 6.自然な表現の例文の提案
 - 7.表現のニュアンスの解説: 単語や表現のニュアンスの違いを詳細に解説し、ユーザが適切な表現を選べるようにサポートしてください。
 - 8.応用問題の提供: ユーザの考えを深め、他の文脈で応用できるような質問や提案を含んでいる。
- 修正例では、条件を満たした形で文法的にも正しい例文を示してください。

※なお、ここでの○文以上という条件の場合は、コンマで区切った文を1文とカウントし、評価すること。※○語以上○語未満の条件については、単語一つを一語としてカウントし、(., は語数としてカウントしない)評価してください。※○語程度という条件の場合は、条件の語数から最低8割以上、最高1.2割以下の語数を満たす解答を正解としてフィードバックしてください。※問題文リストの”conditions”より、「」で指定の英文が条件されている場合は、解答が指定の英文と合致しているかも評価し、フィードバックを行ってください。

図 2 上級用プロンプト文

表 5 試行回数一覧

| | 文法 | 論理構成 | 条件適合性 |
|--------------|-------------------|-----------------------|-------------------|
| 試行内容 | 4種の誤答+正答 | 誤答+正答 | 3種の誤答+正答 |
| 試行回数 | 誤答4種×3回 +正答3回 | 誤答1種×3回 +正答3回 | 誤答3種×3回 +正答3回 |
| レベル別 試行回数 | 15回×3レベル | 6回 ×3レベル | 12回×3レベル |
| 合計回数 | 450回 (10問×45回) | 180回 (10問 ×18回) | 360回 (10問×36回) |

4.2 実験結果と考察

実際に解答パターンをそれぞれ自学自習システムに入力し、フィードバック結果を評価する。評価方法は、評価値を使用し、「文法」、「論理構成」、「条件適合性」では、10問分の評価値の平均値を算出した。「高度な語彙の提案」では、フィードバック内容に含まれる提案された単語を検索し⁷、その単語が英検準2級または3級の語彙に該当する割合を算出した。具体的には、該当単語数を全提案単語数で割った値を用いた。項目「例文」については内容には踏み込まず、フィードバックで例文が提案されたか否かで評価する。「自然な表現の例文」、「ニュアンス解説」、「応用問題の提供」については評価方法が確立しないため、今回は評価対象外とした。

⁶ 生成 AI のシステムに与える入力指示

⁷ weblio 英和辞典・和英辞典 <https://ejje.weblio.jp/>

表 6 文法に関するフィードバックの平均評価値

| | 誤答1: 3 単元 の“s” | 誤答2: 不定詞・ 動名詞 | 誤答3: 前置詞 | 誤答4: 時制の 混乱 | 正答 |
|-----|----------------------|---------------------|-------------|-------------------|-------------|
| 問1 | 1.00 | 1.00 | 0.88 | 0.88 | 1.00 |
| 問2 | 0.61 | 1.00 | 0.77 | 1.00 | 1.00 |
| 問3 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 問4 | 1.00 | 1.00 | 0.61 | 1.00 | 1.00 |
| 問5 | 1.00 | 1.00 | 1.00 | 0.88 | 1.00 |
| 問6 | 0.62 | 0.88 | 0.22 | 1.00 | 1.00 |
| 問7 | 0.83 | 1.00 | 0.88 | 1.00 | 1.00 |
| 問8 | 0.94 | 1.00 | 1.00 | 1.00 | 1.00 |
| 問9 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 問10 | 0.92 | 1.00 | 0.50 | 1.00 | 1.00 |
| 平均 | 0.89 | 0.98 | 0.78 | 0.97 | 1.00 |
| 全平均 | 0.92 | | | | |

表 6 のとおり、「文法」の平均評価値については、全体の評価値が0.92と比較的高い結果となり、不定詞・動名詞に関する誤答では特に精度が高かった(0.98)。一方で、前置詞に関する誤答は精度がやや低かった(0.78)が、一般的な文法チェックソフトでは対応できない誤答パターンにも対応できているため十分意味のある結果といえる。

表 7 条件適合性の平均評価値

| | 平均評価値 |
|-----|-------------|
| 初級 | 0.88 |
| 中級 | 0.88 |
| 上級 | 0.83 |
| 全平均 | 0.86 |

次に「条件適合性」の平均評価値の結果を表 7 に示す。各レベルの平均評価値はいずれも0.8を超えており、高い精度が得られた。一方で、上級レベルではやや精度が低下しており、これは上級レベルの問題が有する条件の複雑性による影響が考えられる。また、正答のみの平均評価値は0.86、誤答のみの平均評価値は0.89となった。

この項目で第1問のときの初級レベルにおいて、誤答を送った場合の実際のChatGPT APIからのフィードバックの例を図 3に表示する。

ユーザの解答(誤答パターン2):
I like playing sports. Because It's fun.

フィードバック:

4. 条件の適合度の評価

- **条件 1**:
「1 文目は、あなたはスポーツを見ることとすることのどちらが好きかについて、I like に続けて書きなさい。」

- 解答は「I like playing sports.」という形で条件を満たしていますが、条件に「見ること」と「すること」のどちらかを選ぶ必要があります。したがって、適切な選択がされていません。

- **条件 2**:
「2 文目以降はその理由がわかるように、2 文以上で書きなさい。」

- 解答は理由を 1 文でしか述べておらず、2 文以上とする条件を満たしていません。

図 3 誤答パターンのフィードバック

条件1は満たしているものの、条件2を満たしていない(下線部)解答パターンを入力した際、誤った結果を返してしまうケースがあることが確認できる。

続いて、「論理構成」の評価値の結果を示す。得られた評価値は0.69にとどまり、精度が十分に高いとはいえない。そのため、論理構成の評価精度を向上させる目的で、「論理構成」に関するプロンプトを修正し、再度評価を行った。しかし、修正後の結果も0.74にとどまり、大幅な改善には至らなかった。

最後に、「高度な語彙の提案」の結果を示す。中級・上級レベルともに5~6割程度にとどまり、精度は高いとはいえなかった。英語力向上の観点から、提案される語彙の質と多様性をさらに向上させる必要がある。なお、「例文の提示」の要件はすべてのフィードバックで満たされていた。

表 8 高度な語彙の提案の結果

| | 総提案語数 | 準2級程度の語数 | 割合 |
|----|-------|----------|-------|
| 中級 | 33 | 20 | 61.2% |
| 上級 | 28 | 14 | 52.8% |
| 全体 | 61 | 36 | 57.0% |

本研究の結果、提案した自学自習システムは、条件適合性、文法、論理構成、高度な語彙提案の各観点で一定の評価精度を達成できることが確認された。一方で、上級レベルや論理構成の評価精度向上にはさらなる改善が必要であることが示唆された。

5. むすびに

本研究では、ChatGPT APIを活用した自学自習システムを開発し、条件英作文のフィードバック精度を評価した。実験の結果、文法や条件適合性のフィードバックは高い精度を達成し、特に文法と条件適合性は一貫した評価が可能であり、学習者のレベルに応じた指導が可能であることを確認できた。しかし、論理構成の評価精度向上や語彙提案の改善が今後の課題となる。

参考文献

- [1] レアジョブ英会話 English Lab, 英作文の勉強法まとめ <https://www.rarejob.com/englishlab/column/20220922/#1>, (参照 2024-12-11).
- [2] 横野健二, “生徒の視点から見た英作文添削”, 高校教育研究(70), pp71-80, (参照 2024-12-11).
- [3] kinimi ブログ, オンライン英会話の市場規模, <https://kinimi.online/blog/archives/30866>, (参照 2024-12-11).
- [4] 教育総合ポータル, ラーニングスケルトン AI, <https://www.media-5.co.jp/hp/learningskeleton/english/> (参照 2024-6-25).
- [5] 桐原書店, 桐原 AI エッセイライティング, <https://www.kirihara.co.jp/-/special/aecs/>, (参照 2024-6-25).
- [6] 古川渉一・萩原優衣, ChatGPT API×Python ではじめる対話型 AI 実装入門, 株式会社インプレス, 2023.
- [7] 日本英語検定協会, ライティングテストの採点に関する観点と注意点, https://www.eiken.or.jp/eiken/exam/2016scoring_w_info.html, (参照 2024-10-30).
- [8] Toshin.com, 2024 年 全国公立高校入試解答速報, https://www.toshin.com/koukou_nyushi/, (参照 2024-12-18).