

英語スピーキング採点技術の開発

サービスイノベーション部
株式会社みらい翻訳

さわやま あつき
澤山 熱気
まつおか ほうせい
松岡 保静

近年、日本の英語教育では、「聞く」「読む」「書く」「話す」の4技能をバランス良く習得することが求められている。特に、「書く」「話す」といった能動的なスキルの習得が重要視されている。しかし、個人で学習することは難しいため、ドコモでは、個人で「話す」練習が可能な、スピーキング採点技術を開発した。本技術により、面接試験をベースとしたスピーキング練習ならびに自動採点が可能になった。

本稿では、その技術の概要と応用について解説する。

1. まえがき

近年、日本の学校教育における英語学習では、「聞く」「読む」「書く」「話す」の4技能をバランス良く習得することが求められるようになってきている。中でも、「書く」「話す」といった、能動的スキルを生徒が習得できることに重点が置かれている。その結果、能動的なスキルを判定可能な資格取得が推奨されるようになり、英検[®]などの資格が授業単位として認定されたり、資格取得によって入学試験の一部免除が行われたりするなど、生徒にとって大きな変化が起きている。

そのような中で、生徒の能動的スキルを高める指

導のために教員の負担がさらに増大し、大きな問題となっている。例えば、外部試験・入試対策のための面接試験練習は、教員が放課後などの授業外の時間に指導しなければならないことが多い。その上、教員の時間には限りがあるため、生徒が十分な指導を受けられないなどの問題もある。

このような背景から、AI技術によって生徒の学習サポートや自動採点を実現することが、生徒の能動的スキルの向上や、教員の負担軽減に有効と考えている。

これまでドコモでは、これらの課題を解決するための最初のステップとして、AIを用いたライティングの採点・添削機能 [1] を開発し、その機能を中

©2023 NTT DOCOMO, INC.

本誌掲載記事の無断転載を禁じます。

本誌に掲載されている社名、製品およびソフトウェア、サービスなどの名称は、各社の商標または登録商標。

高生向け英語4技能の学習サービス「English 4skills」[2]におけるAIライティング機能として提供してきた。これにより、「書く」スキル学習における生徒の学習支援や教員の採点・指導の稼働削減に貢献してきた。

一方、学校教育の観点からは、「話す」スキルについても、取り組むべき喫緊の課題であると考えられる。「話す」スキル習得をサポートするには、生徒が発話した内容に対して、提供中のAIライティング機能と同様に、人手を介在せずに評価できる機能が必要となる。加えて、AIライティング機能と異なり、生徒の英訳文を採点するだけでなく、生徒が英語で話した状況説明や意見を採点できる必要がある。

そこでドコモは、「話す」スキル習得の練習が可能な、英語スピーキング採点技術を開発した。本技術は、「問題カードを用いた状況説明問題の採点技術」と「一問一答問題の採点技術」からなり、対面の面接試験を想定したスピーキング採点システムにこれらを組み込むことで「話す」スキル習得の練習が可能となった。

本稿では、開発した英語スピーキング採点技術について解説する。

2. 面接練習問題の概要

本技術を用いたスピーキング採点システムは、Webアプリケーションの形式で対面の面接試験を想定した練習ができるように設計されている。生徒は「試験官との挨拶」から始まり、「問題カードの受取り」「問題カードを用いた状況説明問題」「一問一答問題」「問題カードの返却」「試験官との挨拶」という一連の流れをスマートフォンやタブレットなどを使って練習できる（図1）。

本システムは、生徒がスマートフォンやタブレットなどのマイクに向かって行った発話を音声認識し、採点処理を行う。試験官との挨拶が終わると、一連の面接試験の採点結果が生徒に提示される。採点システムで実装した2種類の問題（問題カードを用いた状況説明問題、一問一答問題）について解説する。

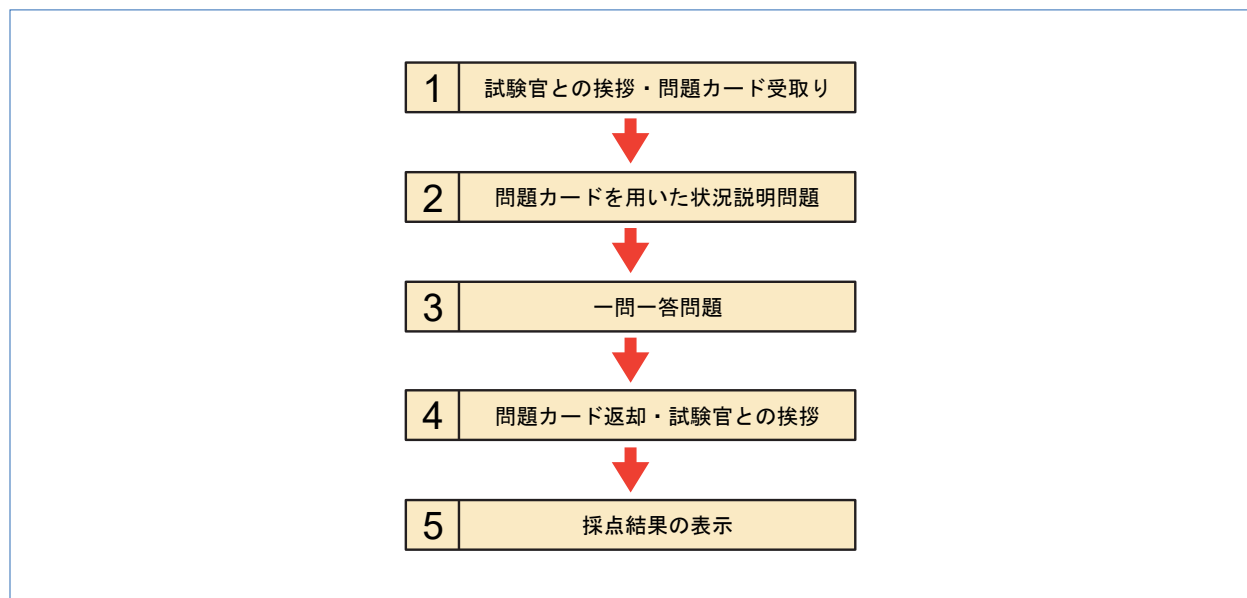


図1 面接試験練習の流れ

2.1 問題カードを用いた状況説明問題

この問題では、問題カードに記載されたパッセージやイメージに関する内容について出題があり、生徒は記載内容の状況説明や、説明についての理由を述べる。問題カードに記載されるイメージの例を図2に示す。問題図に描かれている人物それぞれが行っている行動について状況説明する問題などが試

験官から出題される。

2.2 一問一答問題

この問題では、問題カードなどの参照資料を使わず、トピックに関する出題がなされる。生徒は自分自身の立場や意見を回答する(図3)。試験官は回答した生徒の意見を踏まえて、理由などについて追



図2 問題カードに記載されるイメージの例



図3 一問一答問題のイメージ

加で質問する。

3. スピーキング採点技術の概要

本システムが提供する「問題カードを用いた状況説明問題」「一問一答問題」の2形式に用いられる採点技術について解説する。

3.1 問題カードを用いた状況説明問題の採点

問題カードを用いた状況説明問題の採点は、図4に示すように、(1)回答からの人物・状況の推定、(2)人物・状況の模範解答集合からの模範解答選択と(3)英作文添削技術を活用した回答採点・添削の3つのステップで行う。

最初に(1)のステップでは、事前に用意した、問題図に複数描かれた人物それぞれの状況を説明した模範解答と、今回入力された生徒の回答を比較し、生徒の回答が問題図のどの人物・状況に該当するかを推定する。

続いて、(2)のステップでは、推定された人物・状況に対応する模範解答集合の中で、生徒の回答と最も類似する模範解答を選択し、選択された模範解答とペアデータになっている模範解答の日本語文を出力する。

その後、(3)のステップにより、選択された模範解答に対応する日本語文を英作文添削技術 [1] を用

いて英訳し、生徒の回答と比較することで、意味の観点で生徒の回答の採点・添削を行う。

これら3つのステップにより、生徒の回答が、問題図のどの人物や状況に関する説明なのかを推定しつつ、生徒の回答の意味的な添削まで行うことを可能としている。

(1)回答からの人物・状況の推定

入力された生徒の回答に類似する問題図の人物・状況を推定するにあたり、あらかじめ問題図に描かれた人物それぞれの状況を説明した複数の英文と日本語文のペアデータをまとめた模範解答集合を用意しておく。この模範解答集合は構造化されており、1階層目は、問題図中のどの人物や状況を示すかの情報が、2階層目は、各人物・状況を表現するさまざまな模範解答が格納されている (図5左)。

人物・状況の模範解答集合の推定ステップの動作例を述べる。生徒の回答「A man wearing suit.」が入力されると単語単位に分割する。次に、回答文中の単語をWord2Vec^{*1} [3] を用いて単語ベクトルに変換する。その後、単語ベクトルを足し合わせ、文ベクトルを作成する [4]。同様に、問題図に描かれた人物・状況の模範解答集合の各模範解答文についても、あらかじめ文ベクトル化しておき、生徒の回答の文ベクトルと、それぞれの模範解答集合に含まれる各文の文ベクトルとのコサイン類似度^{*2}を算出し、生徒の回答と模範解答集合との平均類似度か

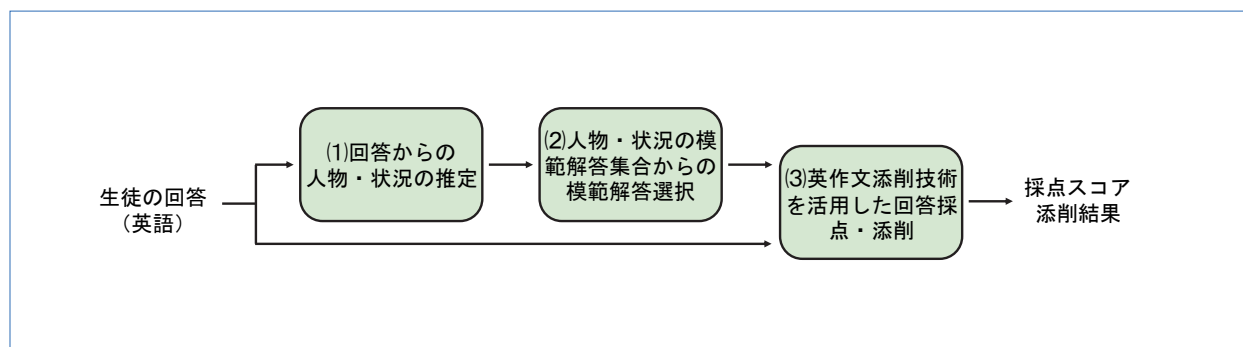


図4 問題カードを用いた状況説明問題の採点フロー

*1 Word2Vec：テキストデータを解析し、各単語の意味をベクトル表現化する手法の1つ。

*2 コサイン類似度：2つのベクトルの向きがどの程度近いかを数値化したもの。

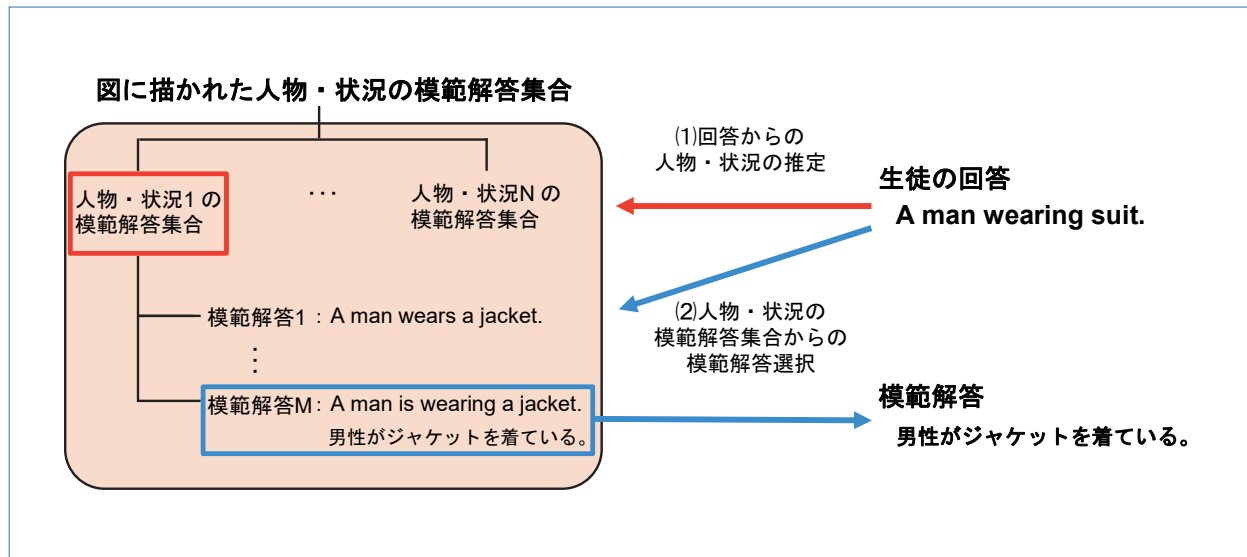


図5 模範解答集合の構造と「回答からの人物・状況の推定ステップ」の流れ

ら、最も類似する模範解答集合を選択する。

(2)人物・状況の模範解答集合からの模範解答選択

生徒の回答の文ベクトルと(1)で選択された模範解答集合に含まれる模範解答の文ベクトルを比較し、最も生徒の回答と類似する模範解答を選択する。図5では「A man wearing suit.」に最も類似する「A man is wearing a jacket.」が選択される。このとき、後段の英作文添削技術を活用した生徒の回答の採点・添削には、選択した模範解答の日本語文「男性がジャケットを着ている。」が用いられる。

(3)英作文添削技術を活用した回答の採点・添削

続いて、(2)で選択された模範解答に対応する日本語文を基に英作文添削技術を用いて生徒の回答の採点と添削を行う。この技術は、日英翻訳モデルをベースとした英文の採点ができ、日本語文の模範解答から英文を生成する過程で、生徒の回答の言回しに近い文章が生成できる。これによって、英語で書かれた模範解答と生徒の回答との比較だけでは難しい、模範解答と意味が似ているが、言回しが異なる生徒の回答も採点・添削できる。

採点の例として、日本語の模範解答を「男性が

ジャケットを着ている.」、生徒の英語回答を「A man wearing suit.」とした場合の動作を示す。まず、日本語の模範解答を単語分割し、日英翻訳モデルをベースとしたエンコーダ・デコーダモデル*3のエンコーダに入力する。そうすると、模範解答文の英訳が始まり、予測単語が推定される。予測単語と生徒の回答した単語から、単語ごとの点数と、添削結果が出力される。すべての単語の点数と添削結果とが出力されると、文としての点数が計算される。このようにして、生徒の回答の採点・添削結果が表示される（図6では添削結果のみ表示）。

3.2 一問一答問題の採点

一問一答問題の採点技術では、回答がシステムに入力されると、採点システムは質問と生徒の回答のペアに一貫性があるか（回答が質問に答えられているか）を評価し、点数を出力する。これによって、提示された質問に対し、回答がどの程度正しく回答できているかを採点することが可能である。

生徒の回答が質問に答えられているかを採点するにあたり、あらかじめ質問と質問に正しく答えてい

*3 エンコーダ・デコーダモデル：ある時系列データの入力から時系列データを生成するリカレント・ニューラルネットワークの構造。ディープラーニングの一種。

る回答のペア（正例）と、質問と質問に答えていない回答のペア（負例）を大量に用意した。この際、人手作成した正例・負例のペアのみを学習に用いるのではなく、正例ペアの質問と回答をランダムに組み替えて疑似的な負例データを作成し、学習データに追加した。その後、単語分割を行った質問と回答のペアを分割トークン「<SEP>」で結合したものを入力系列とし、ディープラーニングの一種である、構造化アテンションモデル [5] に学習させた。これによって、質問文と回答文の意味を理解して、質問に答えられている回答かどうかを採点できる。

質問文を「Do you like dogs?」、回答文を「Yes,

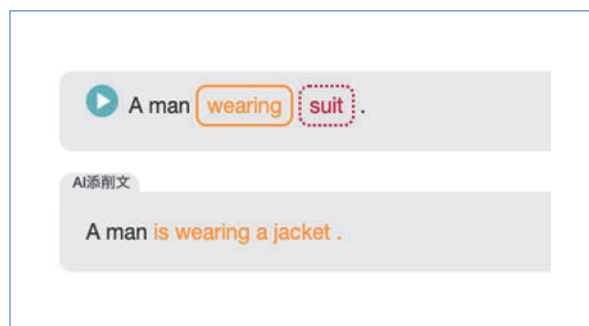


図6 問題カードを用いた状況説明問題の採点例
(図2の左下の男性に対する説明)

I do.」として、採点モデルに入力した場合の採点の動作を図7に示す。まず、上記のように作成された質問・回答ペアを採点モデルに入力する。その後、各入力の隠れ層*4を結合し、アテンション層*5に入力することで、質問文と回答文のどの部分が関連しているかを計算する。その後、質問回答ペアの一貫性の点数が0から1の範囲の値で出力され、出力値を10点満点に正規化する。最後に、質問文に対して、明らかに短い回答がされている場合は、質問に対して内容が不十分であるとし、質問文と生徒の回答文の文長の差をペナルティー（Brevity penalty）として与え、最終的な生徒の回答の点数とする。

生徒の回答の採点例を図8に示す。質問に答えている文章は高い点数を出力しつつ、質問文に使われている単語が含まれていても、質問に答えられていない文章には低い点数を出力する。

4. あとがき

本稿では「話す」スキルの練習が可能な英語スピーキング採点技術を解説した。面接試験練習用として開発した採点技術を学習アプリに組み込むこと

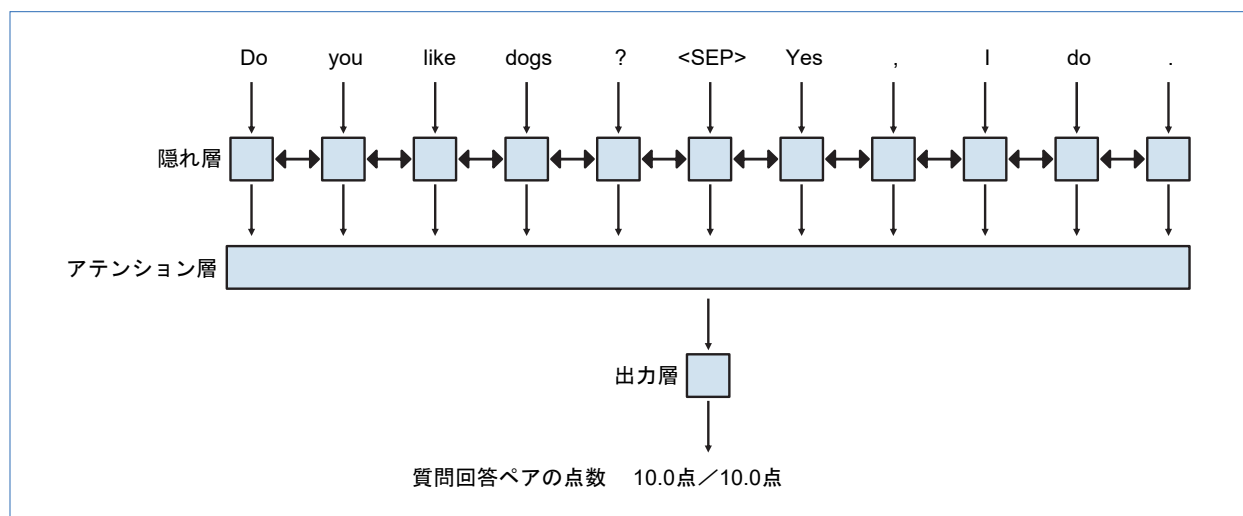


図7 一問一答問題のモデル採点概要

*4 隠れ層：ある入力があった際に、学習されている重みなどに基づき、内部的な値を計算し、出力層に伝播させる層。

*5 アテンション層：ある出力をする際に、入力データ中のどの部分に着目するか（重きをおくか）を出力計算する層。今回の採点では、セルフアテンションと呼ばれる、入力された文中の、単語と単語などの相対的な関係性をとらえることができる手法を用いている。

面接官の質問 : What usually do you do on Sunday ?

回答例1 : I usually go shopping on Sunday.



回答例2 : I usually spend my free time in playing the guitar.



回答例3 : Sunday is the best day of the week.



図8 一問一答問題の採点例

で、これまで個人での練習が難しかった、リアルな面接試験のシーンを想定した、自由な回答を許容するスピーキング練習が可能となった。本技術は、ドコモが提供する、英語4技能の学習サービス「English 4skills」内で「AIスピーキング」として実装されており、実際に教育現場で活用され始めている。今後は、より高度な自由会話の採点が可能な技術も開発していきたい。

文献

- [1] 松岡, ほか: “英作文採点・添削技術の開発,” 本誌, Vol.27, No.4, PP.56-60, Jan. 2020.
 [2] NTTドコモ: “English 4skills,” Apr. 2018.
<https://e4skills.com/>

- [3] T. Mikolov, I. Sutskever, K. Chen, G. Corrado and J. Dean: “Distributed representations of words and phrases and their compositionality,” Advances in neural information processing systems 26, Oct. 2013.
 [4] D. Shen, G. Wang, W. Wang, M. R. Min, Q. Su, Y. Zhang, C. Li, R. Henao and L. Carin: “Baseline needs more love: On simple word-embedding-based models and associated pooling mechanisms,” arXiv preprint arXiv : 1805.09843, May 2018.
 [5] Z. Lin, M. Feng, C. Nogueira dos Santos, M. Yu, B. Xiang, B. Zhou and Y. Bengio: “A structured self-attentive sentence embedding,” arXiv preprint arXiv : 1703.03130, Mar. 2017.