

英作文採点・添削技術の開発

サービスイノベーション部

まつおか 松岡
むらかみ 村上ほうせい 保静
そういちろう 聡一朗なかむら 中村
さわやま 澤山としみつ 俊允
あつき 熱気

近年、日本の英語教育においては、より高い思考力、表現力が重要視され、「聞く」「読む」「書く」「話す」の4技能をバランスよく習得することが求められている。そのような中、ドコモは「書く」に相当する英作文の採点業務に着目し、解答を人工知能（AI）で自動採点および添削する技術を開発した。本技術により、これまで単語や文法の正誤で評価をしていたものが、文章全体の意味を捉えた採点および添削までできるようになった。本稿では、その技術の概要とその応用について解説する。

1. まえがき

近年、グローバル化の中にあって、日本の英語教育においては、より高い思考力、表現力が重要視され、「聞く」「読む」「書く」「話す」の4技能をバランスよく習得することが求められている。それに伴い、英語力を測定する各種試験において、英作文の設問が増えることは必至で、その採点・添削業務の負担を軽減する自動採点システムが注目されてきている。そのような背景の中、ドコモは、学習者の英作文をディープラーニング^{*1}で採点および添削する英作文採点・添削技術を開発した。この英作文採点・添削技術は、日本語とその対となる英語の対訳コーパス^{*2}を大量にディープラーニングで学習させること

で、学習データに無い日本語文章に対する英作文も採点できるようになる。この採点の過程では、問題文となる日本語文章の意味をベクトルとして表現し、そのベクトルから解答の英語文章を生成するという処理を行っている。問題文によっては、解答となりうる英語文章のパターンが幾通りもあるため、上記の日本語文章のベクトルから解答となる英語の文章を生成する過程では、学習者の英作文の表現に近い英語の文章を生成して採点するようにしている。これにより、模範解答と異なる表現でも意味が伝わる解答であれば高得点とすることが可能になる。

従来は英作文が模範解答とどれだけ同一か、つまり、模範解答と同じ構文（文法）で、同じ単語をどれだけ含んでいるかで採点されていたが、本技術を

^{*1} ディープラーニング：膨大なデータから、人間の知能を模擬できる能力をコンピュータに学習させる機械学習の手法の1つ。

^{*2} コーパス：テキストや発話を大量に収集してデータベース化した言語資料。

用いられれば、意識に対する採点も可能となる。

本稿では、英作文採点・添削技術の概要を解説する。

2. 従来の自動採点技術

英作文の自動採点に関する研究は、従来、文法や単語レベルでの模範解答にいかにか近いかを基準に採点を行う技術が一般的であった。自動採点システムは、1966年に開発されたPEG (Project Essay Grade)^{*3} [1] に始まり、現在は、ETS (Educational Testing Service) によって開発されたCriterion[®]^{*4} [2] など、さまざまなシステムがあるが、これらには前述の技術が用いられている。

英作文は大きく分類すると、和文英訳と自由英作文の2種類となる。和文英訳は、日本語を英語に置き換えるものであり、一方の自由英作文は、テーマについての自分の考えを自由に英語で作文するものである。前述の自動採点技術は、和文英訳にも自由英作文にも応用可能であるが、実際に評価指標として見ているのは、文法や語彙の相違、文長など、文章の形式から評価を推定するものがほとんどであり、文章全体の意味を正確に理解した判定は難しいと考えられる。そして、英語学習者の学習初期段階では、

自由英作文よりも和文英訳の方が学習の対象となる場合が多い。そのため、本稿で提案する採点・添削技術は、和文英訳に着目し、問題文の意味を表現できているかという指標を重視する。

3. 英作文採点・添削技術の提案

本技術では、図1で示すように、まず和文英訳の問題となる日本語文章と模範解答となる英語文章のペアのデータ（日英対訳コーパス）を大量に用意する。そして、これをディープラーニングによって学習させ、文章全体の意味を理解して英訳の解答の採点に使用できるモデルを構築する。話し言葉や書き言葉も合わせた数百万文の対訳コーパスを利用し、さまざまな文章に対応できるようにする。このモデルを利用し、問題文の意味を理解しながら、解答文の採点および添削を行う。

学習には、リカレント・ニューラルネットワーク^{*5}のエンコーダ・デコーダモデル^{*6}を利用し、エンコーダに問題文となる日本語文章を単語分割した単語列を入力し、デコーダに模範解答となる英語文章を単語分割した単語列を入力して学習を行う。

この学習したモデルを用いて採点する方式を図2に示す。例として、日本語の問題文が「明日は晴れ

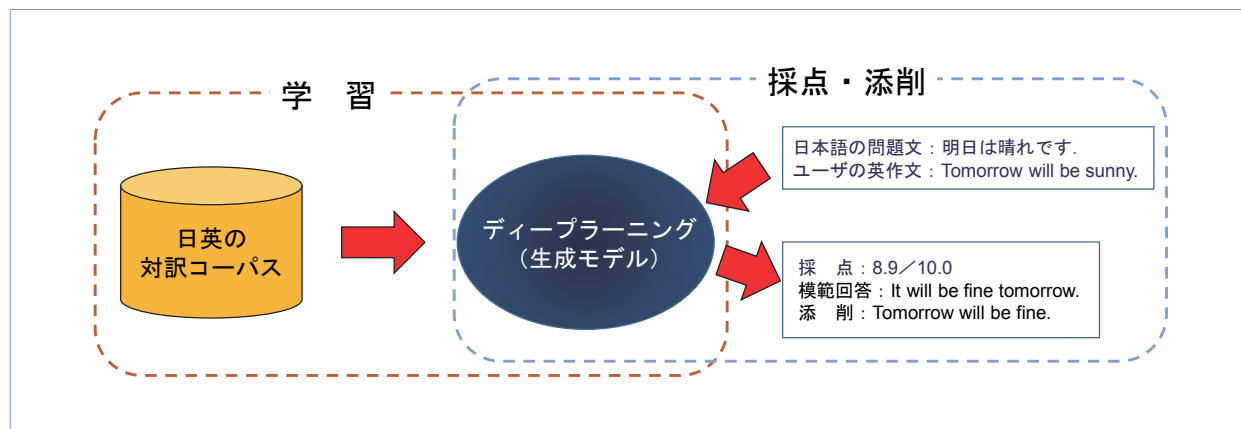


図1 ディープラーニングの応用

*3 PEG[®]：Measurement Incorporationの登録商標。
 *4 Criterion[®]：エデュケーション・テスト・サービスの登録商標。
 *5 リカレント・ニューラルネットワーク：ニューラルネットワークの手法の1つで、中間層の出力が、時系列的に次のステップの入力となる構造の再帰型ネットワーク。

*6 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。

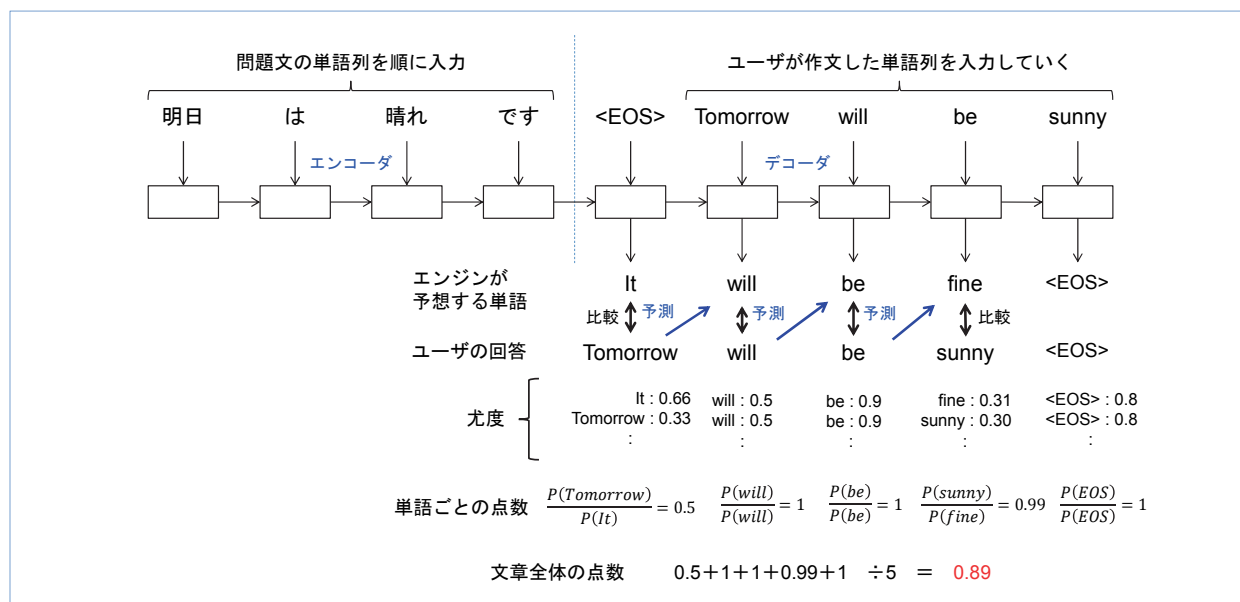


図2 採点方式

です」という場合、まずこの文章を単語分割し、「明日」「は」「晴れ」「です」に分け、これをエンコーダ・デコーダモデルのエンコーダに順番に入力していく。次にデコーダの最初に入力文の終了を示すEOS（End of Sentence）の記号を入力する。そうすると、問題文の英訳が始まり、予測する英文の最初の単語がデコーダから出力される。デコーダでは、ユーザーが作成した英文の単語をデコーダへ1つの入力するごとに、予測した単語との比較で採点を行い、さらに次の予測単語を1つ出力する。

最初は、〈EOS〉入力を契機として、「明日は晴れです」の英訳結果として出力されうる「It will be fine tomorrow」の最初の単語の「It」が出力される。ここで、ユーザーの解答が「Tomorrow will be sunny」だったとすると、この解答文の最初の単語である「Tomorrow」が、デコーダの最初で出力される可能性を計算する。デコーダの最初の出力単語の尤度^{*7}が、「It」が0.66、「Tomorrow」が0.33だとすると、「Tomorrow」の尤度を、最も高い「It」の尤度で除算した値0.5をこの単語の点数とする。

次に、ユーザーの解答の最初の入力である「Tomorrow」を次のデコーダの入力として処理する。デコーダは、最初の単語を「Tomorrow」に固定して、次の単語を予測し、次の単語は「will」が最も尤度が高い単語となる。これは、ユーザーの解答と一致するため、「will」の尤度を「will」の尤度で除算した1がこの単語の点数となる。同様に最後の単語まで計算を行い、すべての単語の点数の平均を文章全体の採点結果の点数とする。

これは、ユーザーの解答文をデコーダに1単語ずつ入力し、ユーザーの解答文を使いながら、次に来る最も良い単語を予測し、その単語とユーザーの解答の次の単語を比較することで採点していることになる。したがって、ユーザーの解答が模範解答と異なる表現であっても、ユーザーが使用した表現に基づいてデコーダはそれに続く英文を作成し、その英文とユーザーの解答を比較していくことになるため、多様な表現に対応できるようになる。

さらに、ユーザーが途中でしか作文できなかったとしても、その続きをデコーダが作文することも可

*7 尤度：ある結果を推測する際に、そのもっともらしさを表す数値。

能である。ユーザの解答文に対して、誤りのある単語部分を最も尤度の高い単語に置き換えることや、その後に続く文章を提案することもでき、ユーザの解答文に対して、よりよい文章になるような添削が可能になる。

ディープラーニングを用い、膨大な対訳コーパスを学習させることで、日本語の問題文の意味を理解しながら、英文を評価できるようになり、問題文と解答文のペアのデータが多くあるほど、より精度の高い採点や添削が可能になる。また、問題文によっては解答例が複数ある場合もあるが、この場合は、解答例ごとに対訳ペアを作成し、ディープラーニングで英訳モデルに学習させることで、さまざまな解答例にも対応できるようになる。また、学習データに無い問題文に対しても採点や添削が可能であり、学習者の自学自習にも有用であると考えられる。

4. 採点指標と採点例

本技術の英作文採点においては、デコーダが採点したスコア0~1を10倍し、10点満点で示すようにしている。実際にいくつかの問題と解答例を点数化し、これらの点数から作成したおおまかな指標を表1に示す。

実際に本技術を用いて採点した例を示す。問題文の和文と、模範解答の英文は下記の通りである。

問題文：このバスに乗れば、駅に着きます。

模範解答：If you take this bus, you will get to the station.

この文章に対して、さまざまな解答例とそのスコアを図3に示す。

①の「If you take this bus, you will get to the station.」は、模範解答通りなので10点となり5という評価になる。②の「This bus will take you to the station.」は、バスが主語になった場合の言い方であるが意味が合っているため、9.53点で5という高評価になる。このように文章の構造が全く変わっていても意味が同じであれば高得点に採点できる。③の「You can get to the station with this bus.」も言回しは違うが意味は合っているため9.27点で高評価5。④の「You can get to the station by taking this bus.」は、意味は間違っていないが、「by taking this bus」という言い方はネイティブから見て流暢な文とは言えないため8.39点で評価4。⑤も同様にくだい言い方になっているため7.84点で評価3になる。⑥の「You can walk to the station.」は、「バス」ではなく「徒歩」という意味になっていて、意味が違っているため、5.78点で評価2。⑦の「This bus you can go station.」は、模範解答で使用されている単語を適当に並べてみただけで文章として成り立っていないため、評価1となる。このように模範解答と同じ単語を使用していても、意味が伝わらなければ低い評価となるため、文章の構造や意味を

表1 採点指標

5段階評価	説明	採点による点数
5	意味が正しく伝わる。ネイティブから見ても流暢な文。	9.0~10.0
4	意味が正しく伝わる。ネイティブから見ると多少改善点がある。	8.0~9.0
3	意味はほとんど正しく伝わる。改善点がある。	7.0~8.0
2	意味が正しく伝わらない場合がある。不自然な点が多い。	5.0~7.0
1	意味がほとんど伝わらない。文として成立していない。	0.0~5.0

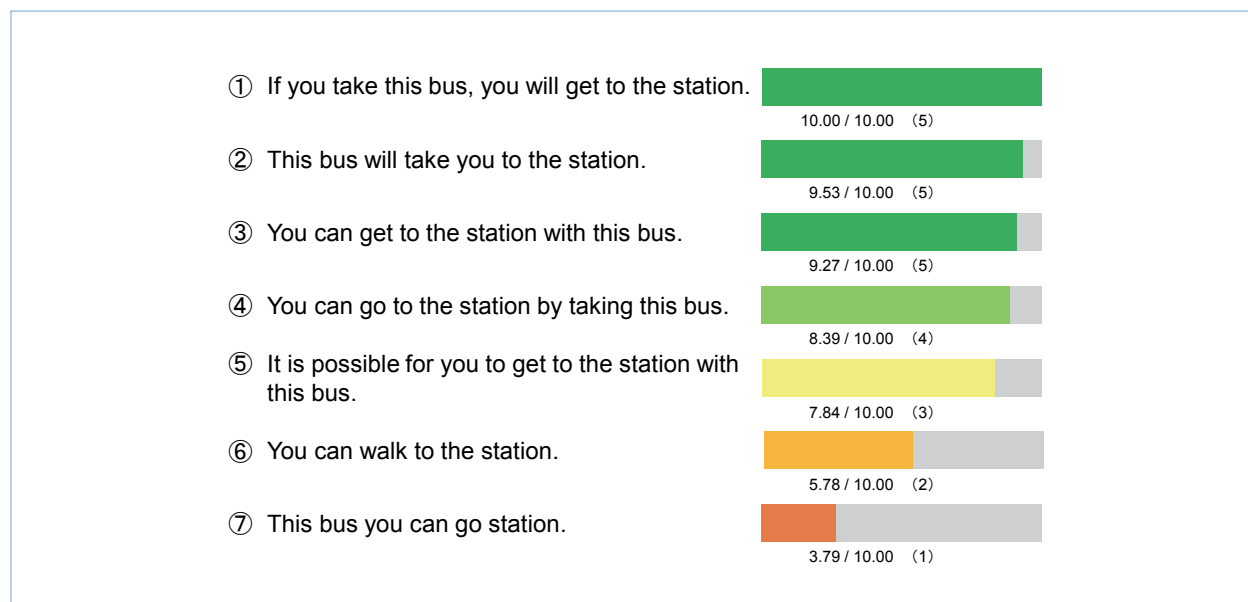


図3 採点例

理解して採点できている。

このように和文英訳の作文において、従来の英語教育では、模範解答通りか否かの○×での採点が多いと聞くと、本技術では、意味が正しく伝わるか、流暢かどうかを数値化して評価できるため、多少間違っていてどれくらい意味が伝わるかを測る手段として有用であると考えられる。また、ユーザの解答に基づいてよりよい解答文を作るための添削やフィードバックもできるため、学習者が自学自習する上でも大いに役立つと考えられる。

5. あとがき

本稿では、和文英訳の意味を重視した採点・添削

ができる技術を開発した。実際に、英作文の上達のために和文英訳を数多く練習する場合においては、AIによる採点や添削は有効であると考え、今後は、英検やTOEFLなどで実施されているエッセイなどの自由英作文にも対応可能な技術も開発していきたい。

文 献

- [1] E. B. Page : "The Imminence of Grading Essays by Computer," Phi Delta Kappan, Vol.47, No.5, pp.238-243, 1966.
- [2] J. Burstein, M. Chodorow and C. Leacock : "CriterionSM Online Essay Evaluation : An Application for Automated Evaluation of Student Essays," IAAI, pp.3-10, 2003.