

## 英語ライティング学習支援のための前置詞誤り解説生成

永田 亮†,††

† 甲南大学 †† 国立研究開発法人科学技術振興機構, さきがけ

E-mail: †nagata-nlp2019@ hyogo-u.ac.jp.

## 1. はじめに

本稿では、解説文生成というタスクを導入し、その解法を提案する。解説文生成とは、図1に示すように、与えられた文書に対してライティングに関するヒントや説明を生成するタスクのことである(図1は、提案手法で実際に生成した解説文である)。解説文生成は、どこが悪いのか、どのように改善したらよいかを学習者が理解する助けになり、ライティング能力の向上に繋がるのが期待される。実際、Bitchenerら[3]は、誤り訂正と共に解説を学習者に与えることで、ライティングの正確さが向上することを示している(注1)。

しかしながら、このような応用面での重要性にもかかわらず、2.で述べるように解説文生成の研究例は非常に少ない。解説文生成が難しい問題であることに加えて、研究に必要となるデータが公開されていないことが理由として挙げられる。

そこで、本稿では、解説文付き学習者コーパスの構築/公開と共に、解説文生成手法を提案する。まず、従来、解説文生成が一つのタスクとして明示的に取り扱われてこなかった背景を考慮し、タスクの定式化を行う。その定式化に基づき、解説文生成手法を提案する。提案手法では、解説箇所の検出と適切な解説文の検索というマルチタスク学習として解説文生成問題を解く。更に、構築した解説文付きデータを用いて、提案手法が前置詞の用法に関する解説文の生成に有効であることを示す。なお、前置詞の用法を対象にするのは、頻出誤りであることと、比較可能な従来手法[12]が存在することを考慮してのことである。

## 2. 関連研究

文法誤り訂正は、解説文生成に関連が深いタスクである。最近では、深層学習を適用し、文法誤り訂正を翻訳問題として解く[10]ことが多い。その結果、目覚ましい性能改善がみられる。一方で、文法誤り訂正では、誤りに関する説明ができないため、学習支援としては改善の余地がある。

解説文生成の研究例は非常に少ない。一部の研究者(例えば、Kakegawaら[11])は、人手で作成した規則により文法誤りを診断することを試みている。また、Nagataら[12]は、自動獲得した格フレームを用いて、解説文と共に前置詞誤り

を訂正する手法を提案している。しかしながら、いずれの手法とも、解説できる誤りは非常に限られている。

ソースコードのコメント生成も関連するタスクである。例えば、Hu[7]らは、ソースコードをコメントに翻訳するという翻訳問題として同タスクを解くことを提案している。このことは、解説文生成も翻訳問題として解けることを示唆する。しかしながら、純粋な翻訳問題として解説文生成を解くことには困難が伴う。図1の(S4)のように、一文中に複数の解説箇所が存在する可能性があるが、通常、翻訳では、原文一文に対して翻訳文一文を仮定する。逆に、(S3)のように、一文中に解説箇所が一つもない事例も多数ある。このような事例に対しては、定型的な解説文(例:「誤りはありません。」)を出力するように訓練することで対処可能である。しかしながら、定型的な解説文に偏った訓練となる可能性がある。実際、5.で示すように、翻訳器を解説文生成に単純に適用することはうまくいかない。

## 3. 解説文生成

## 3.1 タスク定義

解説文生成の入力は文書である。出力は、オフセットと解説文をペアにしたものの集合である。オフセットとは、ペアになった解説文が文書中のどの単語に対するものかを示す整数値の範囲(0スタート)である。解説文は、学習者のライティング能力向上に繋がると思われる内容を記述した文字列である。多くの場合、文法誤りや不自然な表現に対する解説であるが、より良くするためのアドバイスや書き手のモチベーションを上げるための励ましも含む(詳細な内容については、公開データ中のガイドラインに詳しい)。

以上が、タスク定義であるが、以降では、前置詞に対する解説文生成に取り組むことを踏まえ、緩和した定義を利用する。まず、各オフセットの範囲は重複しないとする。また、単語の抜けの誤りの場合(例えば、図1の(S1))、左隣の単語位置をオフセットとする。これにより、単語の抜けに対する解説かその単語自身に対する解説かの間で曖昧性が生じるが、大抵の場合、解説文の内容からどちらのケースにあたるか判断できるため(特に前置詞に関する解説文生成では)問題となることは少ない。更に、オフセットを次式により、範囲から単一の整数に変換して扱う：
$$\lceil \frac{\text{end offset} - \text{start offset}}{2} \rceil$$
。例えば、最初の単語から二番目の単語というオフセットは、 $0 (= \lceil \frac{1-0}{2} \rceil)$ に変換する。緩和した定義を用いる理由は、解

(注1) :ただし、彼らの研究では、誤り訂正は人手で紙面上で行われ、また、解説は口頭で行われていることに注意が必要である。

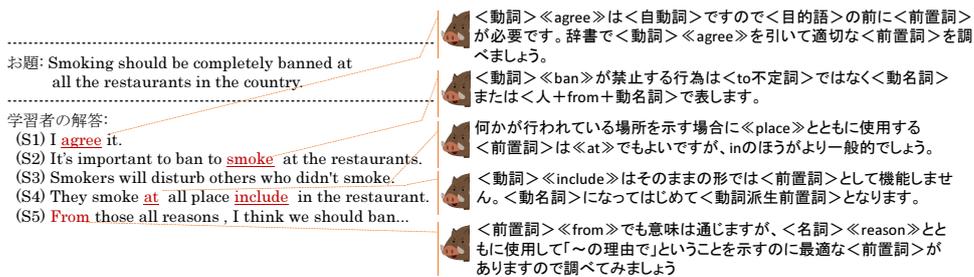


図 1: 提案手法で生成した前置詞の用法に関する解説文の例。

説箇所の検出を 2 値の系列ラベリング問題として解くことができるためである。

### 3.2 解説文生成手法

提案手法は、次の六つの手順から成る：(0) 前処理, (I) 解説文ベクトルの生成, (II) 解説箇所の検出と文脈ベクトルの抽出, (III) ベクトルのマッピング, (IV) 解説文生成 (検索), (V) 出力。主な手順である (I)–(V) の流れを図 2 に示す。

(0) 前処理は、訓練時及び生成時に、入力文書に対して行う。入力文書に対して、文分割、トークン分割、小文字化の処理を施す。また、閾値未満の頻度の単語については <unk> に置き換える。同様の処理を訓練データ中の解説文に対しても行う (ただし、文分割は行わない)。

(I) 解説文ベクトルの生成では、LSTM ベースの言語モデル (LSTMLM) を用いて、訓練データ中の解説文をベクトルに変換する (図 2 (I))。この LSTM は、入力単語から次の単語を予測するように訓練する。訓練済みの LSTMLM に、解説文を入力することで、対応する解説文ベクトルを得る。すなわち、LSTM の最後の隠れ状態ベクトルを解説文ベクトルとする。ただし、表現力を高めるため、前向きと後ろ向きの 2 種類の LSTMLM を用意し、両者の隠れ状態ベクトルを連結したものを解説文ベクトルとする。

(II) 解説箇所の検出と文脈ベクトルの抽出は、二つのサブタスクから成る。最初のサブタスクでは、解説する箇所を決定する (なお、以降では、一文ずつ処理することを想定する)。3.1 で導入した緩和したタスク定義により、解説箇所は、0/1 (解説文なし/あり) の系列で表すことができる。したがって、2 値の系列ラベリングとして解ける。提案手法では、BiLSTM に基づいた系列ラベリングモデルを用いる (図 2 (II))。具体的には、各トークンを分散表現に変換し、順次 BiLSTM に与える。出力層は、softmax 層とし、0/1 の確率を推定する。訓練時の損失関数は、交差エントロピー損失関数を用いる。

二番目のサブタスクでは、解説箇所の文脈を文脈ベクトルとして抽象的に表現する。これには、系列ラベリングで 1 が出力されたトークンに対応する BiLSTM の隠れ状態ベクトルを利用する。BiLSTM の隠れ状態ベクトルは、各タイムステップで、前後の単語の情報を蓄積しているため、文脈ベクトルに適している。

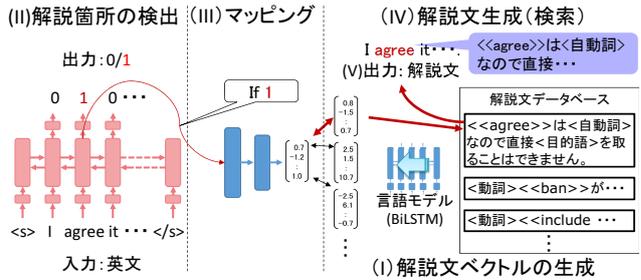


図 2: 解説文生成のための提案モデル。

(III) ベクトルのマッピングでは、FFNN を用いて (II) で得られた文脈ベクトルを解説ベクトルの空間にマッピングする (図 2 (III))。この理由は、文脈ベクトルから、その文脈に適した解説文を検索できるようにするためである。損失関数としては二乗誤差関数を用いる。ただし、FFNN の訓練は、(II) で解説箇所として検出された場合のみ行う。また、実際には、(II) と (III) の訓練は同時に行う (すなわち、解説箇所検出とベクトルマッピングからなるマルチタスク学習である)。ネットワーク全体の損失関数は、 $(1 - \beta)l_s + \beta l_m$  を用いる ( $l_s$ ,  $l_m$ ,  $\beta$  は、それぞれ、系列ラベリングの損失関数、ベクトルマッピングの損失関数、両者のバランスを調整するハイパパラメタである)。

(IV) 解説文生成 (検索) および以降の処理は、生成時のみ適用する。(II) で得られた文脈ベクトルは、(III) により、対応する解説文のベクトルの近くに位置するように、解説文ベクトルの空間上にマッピングされる。したがって、この空間上で、近傍に存在する (訓練データ中の) 解説文は適切なものである可能性が高い。このアイデアに基づき、(IV) では、マッピングされた文脈ベクトルと訓練データ中の解説文ベクトルとの間の余弦類似度を計算する。この結果が生成の基盤となる。

更に性能を高めるため、追加処理を行う。まず、解説箇所の前置詞 (例えば、“at”) と解説内容の前置詞 (例えば、“in”) が一致しない場合がある。このような解説文は、類似度がどれだけ高くとも、不適切である可能性が高い (この例の場合、“at” に対して “in” に関する解説を行うことになる)。したがって、このような解説文は除外する。また、余弦類似度が閾値未満となる解説文も除外する。言い換えれば、類似した解説文が見つからない場合、解説文の生成は行わない。

(V) 出力で、生成結果を出力する。(IV)の結果、最も類似すると判定された解説文を解説箇所の情報と共に出力する。

## 4. データ構築

評価実験のため前置詞に関する解説文を人手で付与した(Web上<sup>(注2)</sup>でデータを公開している)。基礎となるコーパスとして、ICNALE [8]のwritten essayを選択した。ICNALEでは、ライティングのトピックが「大学生のアルバイト」と「レストランでの喫煙」の2種類に統制されている(以降、それぞれ、「アルバイト」、「喫煙」と表記する)。このトピックの統制により、語学学習で頻繁にみられる状況を想定することができ、評価環境として好ましい。すなわち、授業内のライティング課題や資格試験など、全ての学習者が同一のトピックについて書くという状況を想定した評価が行える。

ICNALEから1,040の文書をサンプリングし、日本語で解説文の付与を行った。2名のプロのアノテータを雇用し、いずれか1名が一つの文書を担当した。解説対象は、全ての前置詞誤りに加え、解説を加えたい箇所(例えば、より良くするための助言やモチベーションを向上させる励ましなど)とした。付与の際には、MS-Wordのコメント機能を用いた。なお、前処理として、Stanford Statistical Natural Language Parser (ver.2.0.3)を用いて文分割とトークン分割を行った。このデータの統計量は表1に示すとおりである。

## 5. 評価実験

### 5.1 実験条件と実験手順

構築したデータを用いて、提案手法を実装した。英単語の分散表現は、事前訓練により、次のコーパスから得た: CEEJUS<sup>(注3)</sup>, ETS [4], ICLE [6], CLC FCE [14], NICE [13] (以上、学習者コーパス), EWT [2] (母語話者コーパス)。また、この分散表現を、解説文中に出現する英単語の分散表現としても使用した。その他の日本語単語については、ランダムな値で初期化した。日本語の形態素解析器としてMeCab<sup>(注4)</sup>を用いた。評価実験用に、構築データを8:1:1の比率で訓練、開発、評価データに文書単位で分割した。訓練は、「アルバイト」と「喫煙」別々に行い、2種類の訓練結果を得た。ハイパパラメータは開発データで決定した<sup>(注5)</sup>。その結果を対応する評価データで評価した。訓練を5

(注2) : [http://nlp.ii.konan-u.ac.jp/fc\\_corpus.html](http://nlp.ii.konan-u.ac.jp/fc_corpus.html)

(注3) : <http://language.sakura.ne.jp/icnale/modules.html>

(注4) : <http://taku910.github.io/mecab/>

(注5) : LSTMLM : LSTMの層数2; 隠れ状態の次元600。BiLSTM : LSTMの層数1; 隠れ状態の次元1,200。FFNN : 層数1 (次元1,200)。共通ハイパパラメータ : 単語分散表現の次元600; 未知語(<unk>)の頻度閾値 : 英単語5未満, 日本語単語2未満; バッチサイズ64; early stopping : エポック数200, patience 10; ドロップアウトレート0.5; 最適化アルゴリズムAdam (step size: 0.01, first/second moment: 0.9/0.999)。余弦類似度の閾値0.971 (アルバイト)と0.976 (喫煙)。損失関数の重み $\beta = 0.9$ 。

表1: 構築した解説文データの統計量。

トピック	アルバイト	喫煙
文書数	521	519
トークン数	131,947	130,154
解説文数	1,408	1,426

回行い、各生成結果の和集合を最終的な生成結果とした(同じ個所に複数の解説文が生成された場合、余弦類似度が最も高いものを選択した)。

比較手法として、2種類の手法を実装した。一つ目は、格フレームに基づいた手法[12]である。訓練データとして次のコーパスを用いた: BNC [5], EDR コーパス [9] (母語話者コーパス); 構築した訓練データと開発データ(学習者コーパス)<sup>(注6)</sup>。二つ目は、注意機構付きエンコーダデコーダニューラルネットワーク[1] (以降、enc-dec+attと表記する)である。すなわち、学習者の書いた英文を解説文に翻訳するという翻訳問題として解く手法である。訓練には、基本的に、提案手法と同じ条件を用いた。ただし、enc-dec+attは、解説文が付与された英文のみを用いて訓練と評価を行った<sup>(注7)</sup>。

解説文付与に携わった一人が生成結果を主観評価した。各生成結果を、適切、部分的に適切、不適切の3種類に分類した。部分的に適切とは、生成結果の一部を修正することで適切な解説となる場合を指す。評価尺度として、recall, precision,  $F$  値を用いた。「適切」のみを正しい生成とみなした厳密評価と「部分的に適切」も正しい生成とみなした緩和評価の2種類を用いた。

### 5.2 評価結果

表2に評価結果を示す。表2から、提案手法は、比較手法の性能を大きく改善していることがわかる。enc-dec+attの結果は、解説文生成を純粋な翻訳問題として解くことは有効でないことを示す(その理由は、2.で議論した通りである)。格フレームに基づいた手法でも、そのrecallは非常に低い。一方、提案手法は、同程度のprecisionを達成しつつ、recallを2~5倍も改善している。この結果は、解説対象の英文と解説文とをベクトルとして抽象的に表現し、検索問題として解説文生成を解く提案手法の有効性を示すものである。

一方で、表2から、前置詞の用法に限ったとしても解説文生成は難しいタスクであることもわかる。次節では、提案手法で、どのような問題が解けるのか(すなわち、どのような誤りに対して解説文が生成できるのか)を考察する。

## 6. 考察

提案手法の生成結果を人手で分析したところ、典型的な誤りに対して適切な解説文が生成できることを確認した。具体

(注6) : この手法は開発データを必要としないため、開発データも訓練データとして使用した。

(注7) : 解説文が付与されていない英文を訓練に用いると、何も出力しないネットワークが得られたためである。定型文を出力させた場合も同様である。

表 2: 生成性能 (カッコ内の数値は緩和した評価尺度に対応).

手法	アルバイト (解説文総数: 190)			喫煙 (解説文総数: 184)		
	Recall	Precision	F 値	Recall	Precision	F 値
提案手法	<b>0.232 (0.253)</b>	0.611 ( <b>0.667</b> )	<b>0.336 (0.366)</b>	<b>0.283 (0.304)</b>	0.722 ( <b>0.778</b> )	<b>0.406 (0.438)</b>
格フレーム	0.095 (0.095)	<b>0.621</b> (0.621)	0.164 (0.164)	0.049 (0.04)	<b>0.750</b> (0.750)	0.092 (0.092)
Enc-dec+att	0.063 (0.074)	0.069 (0.080)	0.066 (0.077)	0.098 (0.125)	0.108 (0.131)	0.103 (0.131)

的には、トピックに典型的な解説 (図 1 の (S2)) やエッセイライティングに典型的な解説 ((S1), (S5)) などである。これら典型的な誤りに対しては、人手で作成した規則でも、ある程度、対応可能であるかもしれない。しかしながら、そのためには、英語の専門家と言語処理の専門家が必要となる。まず、英語の専門家は、各トピックにおいて典型的な誤りとは何かを考慮しながら、誤り検出と解説文生成の基盤となる規則を考えなければならない。その後、言語処理の専門家が、その規則を計算機で扱える形で実装しなければならない。一方、提案手法では、(何が典型的かということに特に意識する必要はなく)、英語の専門家は、与えられたデータに解説文を付与するだけでよく、より効率的かつ効果的である。

更に、典型的な誤りではあるが、従来手法では対処が困難な言語現象に対しても解説文が生成できることを確認した。具体例として、図 1 (S4) の “include” を挙げることができる。この誤りに対して適切な解説文を生成するためには、“smoke” と “include” が、それぞれ動詞と前置詞として使用されていることを認識しなければならない。通常、単語の関係の認識には構文解析が用いられるが、見た目上、両単語とも動詞の形をしているため解析に失敗する可能性が高い (したがって、解説文の生成にも失敗する)。構文解析に成功するためには、“include” が前置詞誤りであることを認識しなければならない。このように、構文解析が先か誤りの認識が先かというジレンマに陥る。実際、構文解析を使用する格フレームに基づいた手法では、この例に対して解説文を生成できなかった。一方、提案手法では、類似した誤りから適切な解説文を選択していた。同様に、構文解析のミスを引き起こしそうな誤りに対して適切な解説文を生成できた例は、少なくとも 5 例あることを分析により確認した。

以上の分析により、提案手法は、生成性能に改善の余地はあるものの、従来手法より幅広い解説文を効率的に生成できることがわかる。特に、典型的な誤りに対して適切な解説文を生成できることは学習支援としては有益である (学習支援の場面では、学習者が頻繁に間違える典型例から、まず解説すべきであろう)。更に、典型的な誤りではあるが、従来手法では対処が困難である言語現象に対しても解説文が生成可能となったことは、今後の解説文生成研究を考える上で重要である。

## 7. おわりに

本稿では、解説文生成タスクを導入し、その解法を示した。具体的には、解説箇所検出 (系列ラベリング) と類似解説

文の検索 (ベクトルのマッピング) というマルチタスク学習として解いた。評価実験により、前置詞の用法に対して、従来手法より性能が高いかつ、従来手法では生成が困難である言語現象に対して解説文が生成できることを示した。

今後は、提案手法で生成された解説文を、より適切なもの書き換える手法の考案に取り組む予定である (翻訳問題として解けると予想している)。また、前置詞だけでなく、一般的な解説文の生成にも取り組む予定である (一般的な解説文を付与したコーパスも構築済みである)。

## 謝 辞

本研究の一部は、JST、さきがけ、JPMJPR1758 の支援を受けたものである。

## 参考文献

- [1] Bahdanau et. al., “Neural machine translation by jointly learning to align and translate,” arXiv e-prints, 2014.
- [2] Bies et al., “English Web Treebank LDC2012T13. Web Download,” 2012.
- [3] Bitchener et. al., “The effect of different types of corrective feedback on ESL student writing,” *Journal of Second Language Writing*, vol.14, no.3, pp.191–205, 2005.
- [4] Blanchard et al., “ETS corpus of non-native written English, LDC2014T06. web download,” 2014.
- [5] Burnard, *Users Reference Guide for the British National Corpus. version 1.0*, Oxford University Computing Services, Oxford, 1995.
- [6] Granger, “The international corpus of learner English,” in *English language corpora: Design, analysis and exploitation*, pp.57–69, Rodopi, 1993.
- [7] Hu et. al., “Deep code comment generation,” *Proc. of 26th Conference on Program Comprehension*, pp.200–210, 2018.
- [8] Ishikawa, *The ICNALE and Sophisticated Contrastive Interlanguage Analysis of Asian learners of English*, pp.91–118, Kobe University, Kobe, 2013.
- [9] Japan electronic dictionary research institute Ltd, “EDR electronic dictionary specifications guide,” 1993.
- [10] Junczys-Dowmunt et. al., “Approaching neural grammatical error correction as a low-resource machine translation task,” *Proc. of 2018 NAACL*, pp.595–606, 2018.
- [11] Kakegawa et. al., “Diagnostic processing of Japanese for computer-assisted second language learning,” *Proc. of 38th ACL*, pp.537–546, 2000.
- [12] Nagata et. al., “Correcting preposition errors in learner English using error case frames and feedback messages,” *Proc. of 52nd ACL*, pp.754–764, 2014.
- [13] Sugiura et. al., “A discriminant analysis of non-native speakers and native speakers of English,” *Proc. of CL2007*, pp.84–89, 2007.
- [14] Yannakoudakis et. al., “A new dataset and method for automatically grading ESOL texts,” *Proc. of 49th ACL*, pp.180–189, 2011.