

日本人の英作文誤りに着目した修飾語表現選択のための例文検索

Example Sentence Retrieval for Selection of Modifier Expression that Focuses on Errors in English Composition of Japanese

西川 新之助^{*1}, 國近 秀信^{*2}, 竹内 章^{*2}Shinnosuke NISHIKAWA^{*1}, Hidenobu KUNICHIKA^{*2}, Akira TAKEUCHI^{*2}^{*1}九州工業大学情報工学部知能情報工学科^{*1}Department of Artificial Intelligence, Computer Science and Systems Engineering,
Kyushu Institute of Technology^{*2}九州工業大学情報工学研究院^{*2}Computer Science and Systems Engineering, Kyushu Institute of Technology

Email: s.nishikawa@minnie.ai.kyutech.ac.jp

あらまし: 英語初学者が, 自然な英文を書くためには英語母語話者が書いた英文を手本として英作文を行うことが有効である. 学習者に有用な例文は学習者の誤りに依存すると考え, 誤りに対応した構文構造を検索パターンとして用いる例文検索方法を提案する. 本研究では英語初学者の英作文における副詞および形容詞の誤りに着目し, 構文構造を検索パターンとして用いた副詞および形容詞の語彙選択, 位置選択のための例文検索機能を実現した.

キーワード: 英作文支援, 例文検索, 構文木検索, 大規模コーパス

1. はじめに

稲沢は, 英語初学者の英作文支援には, 一般に日本人学習者が頻繁に犯す誤りに対応できる例文検索機能が有効であると考え, 誤りの種類ごとに構文木検索パターンを用意し, 例文検索システムを実現した⁽¹⁾. しかし, 対応した誤りは語彙に関するものだけであり, 語彙選択においても副詞や形容詞のように冠詞や前置詞と比較して語彙の種類が多いものは, 検索された例文の中から学習者が必要とする例文を探し出すのに手間が掛かった. そこで, 本研究では先行研究で未対応の誤りへの対応を目指す.

2. 誤り調査

小島は, 日本人学習者は強意副詞を適切に使用できていないと指摘している⁽²⁾. そこで, 例文検索用いて強意副詞による修飾表現選択の支援が出来ないかと考え, 学習者コーパス NICE2 を用いて強意副詞の誤りの種類の調査を行った. 調査の結果, 誤りの種類は文の表現の変更に伴う削除, 過剰, 同一意味範囲内の語彙選択, 意味カテゴリー間での語彙選択, 文中での修飾語の位置選択であった.

3. 副詞・形容詞の誤りに対応した例文検索

上記の誤りのうち, 同一意味範囲内の語彙選択および文中での修飾語の位置選択は例文検索による作文支援が有効であると考えた. 語彙選択では, 特定の意味範囲に属する修飾語と学習者が指定した被修飾語の係り受け関係ごとに検索された例文とその数から, 位置選択では, 文中での修飾語の位置ごとに検索された例文とその数から, 適切な修飾表現を判断できると考えた.

また, これらの誤りにおいては強意副詞に限らず

副詞と形容詞でも同様に例文検索による作文支援が有効であると考え, 副詞と形容詞も対象とした. 以下に誤りに対する例文検索方法を述べる.

1) 同一意味範囲内の語彙選択

修飾語と被修飾語が兄弟の関係にある構文構造をあらかじめ用意する. 学習者が修飾語の意味範囲と被修飾語を指定すると, 指定された意味範囲に属する修飾語ごとに被修飾語と兄弟関係にある構文木検索パターンを作成し, 構文解析コーパスに対し構文木探索を行い例文を取り出す.

2) 文中での修飾語の位置選択

学習者が書いた英文中のある単語を修飾したい時, 修飾語を挿入する位置の判断の参考になるのは, 学習者が書いた英文の構文と類似する構文をもつ例文だと考え, 学習者が書いた英文から検索に用いる構文構造を抽出する. 修飾語の位置は英文構造や修飾語や被修飾語の形態, 用法, 意味にも依存するが本研究では英文構造のみ考慮する. 例えば, 学習者が書いた英文中の被修飾語が名詞の場合, 学習者が書いた英文の構文木上で, 修飾語が被修飾語の左および右の兄弟関係にある構文構造を生成する. このような生成規則により生成された構文構造を用いて, 構文解析コーパスに対し構文木検索を行い例文を取り出す.

これらの例文検索方法によって, 構文知識が不足した英語初学者でも, 英作文に有用な例文を構文構造に依存して検索が出来る.

4. システム概要

副詞・形容詞の語彙選択と位置選択のための例文検索システムを実現した. システム構成図を図 1 に示す.

語彙選択のための例文検索システムでは学習者が英文または英単語を入力する。英文・英単語判別機能で入力に対して形態素解析を行い、品詞情報を付与する。形態素解析には Tree Tagger を用いた。入力が英文か英単語かの判定は、形態素解析で単語数を数えることで判定した。被修飾語選択機能では、入力が英文の場合は、品詞情報が付与された英文の中から学習者に被修飾語を選択させる。入力が単語の場合は、形態素解析によって付与された品詞情報の信用性が低いため学習者に品詞を選択させる。修飾語意味範疇選択機能では被修飾語の品詞により、修飾語の品詞を決定した後に、学習者に修飾語の意味範疇を選択させる。意味範疇が選択されると、あらかじめ意味分類した修飾語から学習者が選択した意味範疇に属する修飾語を取り出す。検索パターン合成機能では、あらかじめ用意した構文構造に被修飾語とその品詞、修飾語を合成し、構文木検索パターンを作成する。

位置選択のための例文検索システムでは、学習者が英文を入力すると、構文解析機能で入力文に対して品詞情報と構文情報を付与する。構文解析には Stanford Parser を用いた。被修飾語選択機能では、品詞情報が付与された英文の中から学習者に被修飾語を選択させる。修飾語入力機能では、被修飾語の品詞から修飾語の品詞を決め、学習者に修飾語を入力させる。構文木生成機能では、前項で述べたような生成規則に従い、入力文の構文情報、被修飾語とその品詞、修飾語から構文木検索パターンを生成する。

検索には構文木検索ツール Stanford Tregex を用いた。検索の対象は BNC の構文解析コーパス(約 375 万文)である。

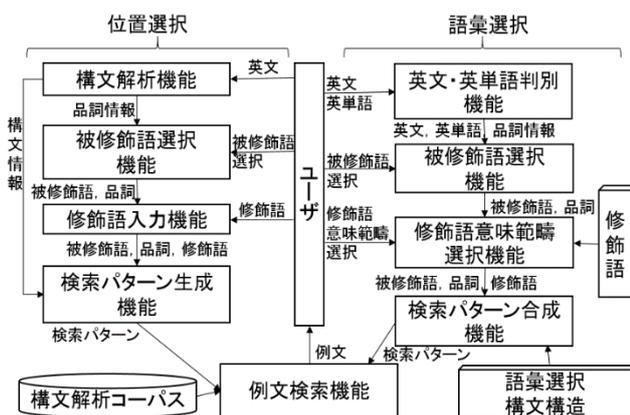


図1 システム構成図

5. 評価

BNC 構文解析コーパスを対象に、指定した文字列を含む文を取り出す文字列検索、語彙選択のための例文検索、位置選択のための例文検索を用いて取り出した例文数を比較し、どの程度余分な例文を除外できるのか調査した。被修飾語を has(JACET8000 中で出現頻度 15 位)、被修飾語を「確信がない」こと

を表す副詞 perhaps(278 位)、probably(313 位)、maybe(326 位)、possibly(1401 位)とした場合、検索方法と修飾語の種類ごとに検索された例文数を比較した結果を表 1 に示す。文字列検索では 2 語が係り受け関係にない文が含まれるが、語彙選択、位置選択では 2 語が係り受け関係にない文は除外された。位置選択では 2 語の修飾関係だけでなく入力文の構文構造も検索に用いられるため、語彙選択よりもさらに取り出された例文数が減少している。また、maybe のように has とそもそも共起しにくい副詞や possibly のように出現頻度の低い副詞は位置選択でヒットする例文がないこともある。

表 1 検索方法と修飾語の種類による例文数の比較

	perhaps	probably	maybe	possibly
文字列検索	1189 文	1036 文	112 文	314 文
語彙選択	101 文	203 文	15 文	11 文
位置選択	2 文	3 文	0 文	0 文

実現したシステムを用いた語彙、位置選択が適切であるか調査するため、英語正誤辞典^③から副詞の語彙、位置に関する禁止を表す注意事項を無作為に 10 個ずつ選び、システムを用いて各注意事項に従った修飾表現の選択ができるか確認した。語彙、位置は例文数が最も多いものを選択した。選んだ注意事項の内、語彙選択では、7 つは適切な選択ができたが、2 つは適切な選択ができなかった、1 つはヒットする例文がなく選択ができなかった。位置選択では、6 つは適切な選択ができたが、4 つはヒットする例文がなく選択ができなかった。この調査では、語彙選択、位置選択のどちらも過半数の注意事項に従うことができたが、注意事項に従うことができたということは選択した修飾表現が誤りでないことを示すが、英語母語話者が日常的に用いる修飾表現であるとは限らないことに注意する必要がある。

6. おわりに

本研究で、副詞と形容詞において先行研究では対応していない位置選択の誤りに対応した。また、副詞の語彙選択においては利用者による副詞の意味範疇の指定に対応した。課題としては、形容詞の語彙選択についても副詞と同様の機能を持たせることがある。また、位置選択では検索で用いる構文構造が入力文の影響を強く受けて、例文を取り出せないことがあるため、検索方法の改善も必要である。

参考文献

- (1) 稲澤恭輔, 國近秀信, 竹内章: “英作文における誤り修正のための大規模コーパスを用いた例文検索”, 人工知能学会 ALST 研究会, 70, 61-66 (2014)
- (2) 小島ますみ: “強意副詞と形容詞のコロケーションの分析”, 投野由紀夫, 金子朝子, 杉浦正利, 和泉絵美: “学習者コーパス活用ハンドブック”, 大修館書店, pp.78-84
- (3) 荒木一雄編: “英語正誤辞典”, 研究出版社 (1986)