

検索エンジンを用いた英文前置詞誤り修正のための 検索フレーズ生成法

久保田 朗^{†1} 太田 学^{†1}

英語を母語としない日本人が英作文を行うと、適切な前置詞の選択に迷うことがある。適切な前置詞であるかどうか判断するには、その前置詞を含む英文を検索フレーズとして検索エンジンで検索し、その検索結果数を調べる方法がある。そこで有富らは、検索フレーズの生成、検索結果のサマリの収集、サマリにおける前置詞の出現確率の自動計算、その結果のユーザへの提示を行うシステムを提案した。しかし、有富らの生成する検索フレーズには、いくつかの問題があった。そこで本稿ではそれらの問題の改善を含む、新たな検索フレーズ生成法を提案する。また、提案する検索フレーズ生成法による前置詞誤りの自動検出、自動修正性能を、有富らの手法ならびに NativeChecker と比較するために行った実験の結果について報告する。

A Query Phrase Generation Method for Correction of English Preposition Errors Using a Search Engine

AKIRA KUBOTA^{†1} and MANABU OHTA^{†1}

Japanese people are sometimes at a loss what preposition to use in English composition. To select an appropriate preposition, there is a way of confirming how appropriate used prepositions are by using a search engine. That is, we search for a query phrase including the preposition to be checked using a search engine and judge its appropriateness based on the number of returned search results. Aritomi et al. proposed a system that automatically i) generates query phrases including a preposition, ii) collects summaries of the search result, iii) calculates occurrence probabilities of prepositions in the collected summaries, and iv) presents the prepositions and their probabilities to a user. There are, however, several problems with their query phrase generation. Therefore, we propose a new method of generating query phrases by which their problems can be solved. We automatically detected and corrected preposition errors using the proposed method for experiment. Moreover, we compared the results to those of Aritomi's method and NativeChecker.

1. はじめに

我々日本人が英語で文書を作成する際、表現方法に迷うことがある。英語を母語としない日本人にとって、適切な前置詞の選択は特に難しいだろう。そこで有富ら¹⁾は、検索エンジンを利用して、Web上の文書コーパスにおける前置詞の出現頻度を調べ、それによって妥当性を判断し、結果をユーザに提示するシステムを提案した。検索エンジンを利用するためには、適当な検索フレーズを生成する必要があるが、有富らは、検討したい英文を入力すると検索フレーズを自動で生成し、検索を行う前置詞誤り修正支援システムを作成した。しかし、その検索フレーズ生成法には複合名詞や日付の表現をうまく扱えないといった問題が存在し、誤りの自動検出、修正ができない英文が存在した。そこで本稿では、新たな検索フレーズの生成法を提案し、評価実験により、有富らの検索フレーズ生成法と結果を比較する。

2. 関連研究

2.1 検索エンジンを用いた前置詞誤り修正支援

検索エンジンによる英文前置詞誤りの自動検出、修正を行うシステムを、岡山大学の有富ら¹⁾が提案している。検討したい前置詞を含む英文を入力として与えると、システムが検索フレーズを自動生成して検索を行う。次に、返ってきた検索結果から前置詞の出現確率を自動計算し、結果をユーザに提示する。しかし彼らの検索フレーズ生成法では、複合名詞等の扱いが考慮されておらず、名詞が複数並ぶ日付の表現や複合名詞等を含む英文に対して、適切な検出、修正ができなかった。

2.2 検索エンジンを用いた英文冠詞過りの検出

検索エンジンを使った英文誤り検出の研究は前置詞以外の品詞についても行われており、Lapataら²⁾によって、冠詞の誤り検出の研究が行われている。Lapataらは、単純な検索フレーズで検索を行い、誤り検出を行っている。彼らの手法は、まず構文解析を用いて名詞句を抽出する。続いて抽出した名詞の冠詞を { a/an, the, } に変化させ、3パターンの検索フレーズを生成し、それぞれ検索を行う。ここで、は無冠詞を表す。最後に、返ってきたヒット数を比較することで、冠詞誤りを検出する。また、早稲田大学の平野ら³⁾は、lapata

^{†1} 岡山大学大学院自然科学研究科

Graduate School of Natural Science and Technology, Okayama University

らの手法を改良し、名詞を単数形と複数形に変化させた検索フレーズを加えることで、検出率、検出精度共に向上させている。

2.3 検索エンジン用いた英作文支援

英作文した文章の前置詞、冠詞、多義語などが適当であるか検索エンジンを用いて検討するシステムを、早稲田大学の大鹿ら⁴⁾らが作成している。機能の一部として前置詞の誤りの検出も実装されている。検討したいフレーズを入力として与えると、システムがフレーズ内の前置詞部分をワイルドカードに置き換え、検索を行い、入力されたものとは異なる前置詞を取得する。次に、入力した前置詞と取得した前置詞それぞれを含む検索フレーズで検索を行い、検索結果数を表示する。この結果数から、どの前置詞が適切かユーザに判断してもらう。欠点として、何度も検索を行うため応答時間が長い点、英文全体でなく、ユーザ自身が適当なフレーズを考えて入力する必要がある点が挙げられる。

2.4 NativeChecker

NativeChecker⁷⁾は、英語のフレーズを入力すると、そのフレーズの様々な項目を修正することができる Web サービスである。検討したい英語フレーズを入力すると、入力がそのまま検索フレーズとなり、その検索結果数が表示される。修正を行う場合、修正したい単語の上にマウスのカーソルをのせると修正方法が提示され、ユーザが適当な修正方法を選び、修正を行う。修正後、再検索を行うことで検索結果数を得ることができる。修正できる内容として、スペルミス、類義語、単複数形、to 不定詞、その他の表現、単語の削除、単語の順序の変更がある。

3. 英文前置詞修正支援システム

3.1 英文前置詞修正支援システムの概要

本研究では有富らが作成した前置詞誤り修正システムのユーザーインターフェースを利用する。その処理の流れを図 1 に示す¹⁾。まず検討したい英文を入力として与える。次に入力した英文を処理し、検索フレーズを生成し、検索を行う。検索結果が不足している場合は、検索フレーズを修正して再検索を行う。このようにして得られた検索結果上位 100 件までのサマリを取得し、その中に含まれる前置詞の出現確率を計算する。最後に、出現確率の高い順に並べて前置詞を表示する。

3.2 検索フレーズの生成

本節では、入力された英文から検索フレーズを生成する処理について説明する。

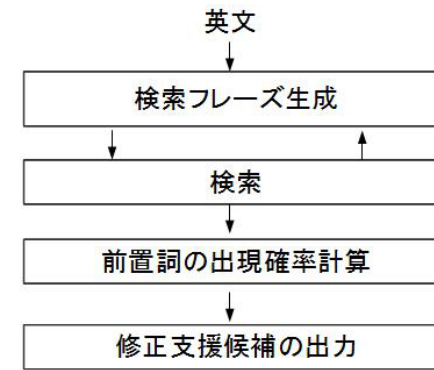


図 1 英文前置詞修正支援システムの流れ

- (1) 入力された英文がコンマを含む場合、もしくは複数の前置詞を含む場合は英文を分割する。分割処理の方法については 3.2.1 節で述べる。
- (2) 入力された英文もしくは分割された英文に対して、品詞のタグ付けを行う。タグ付けには Eric Brill の MontyTagger⁵⁾ を用いる。
- (3) 品詞タグに基づき、英文から主語と助動詞を全て削除する。
- (4) 特定の品詞だけが残るように、不要な単語を削除する。この単語の削除については 3.2.2 節で詳しく述べる。
- (5) 前置詞部分をワイルドカードに置き換え、検索フレーズとする。この方法については、例を交えて 3.2.3 節および 3.3 節で述べる。

3.2.1 英文の分割処理

入力された英文がコンマを含む場合は、コンマの前後で分割する。入力された英文に前置詞が複数含まれる場合は、その前置詞の数だけ英文を分割する。前置詞の前後の単語列が検索フレーズを構成するため、分割後の英文には重複する部分が存在する可能性がある。例えば、“I waited for his call until midnight.”という文からは、“I waited for his call”、“his call until midnight”の 2 つの単語列が抽出される。また、前置詞が連続して出現する場合は、そこでは分割しない。

3.2.2 単語の削除

分割処理の済んだ、前置詞を1つだけ含む単語列をその前置詞の前後で分割する。ただし、ここでは連続する2つの前置詞は1つとみなす。前置詞の前の単語列に動詞が存在する場合、動詞より前に存在する主語、助動詞をすべて削除する。このようにして生成された前置詞の前後2つの単語列に、次の処理を行う。

● 前置詞の前の単語列

(1) 名詞が存在する場合

動詞、名詞、冠詞以外の単語すべてを削除し、単語列を動詞 + 冠詞 + 名詞とする。ただし、動詞と冠詞は存在する場合のみ残す。また名詞が複数含まれている場合は、最後に現れる名詞群を1つ残す。ここで、名詞群とは連続して現れる複数の名詞のことであり、複合名詞に限ったものではない。有富らは、最後に現れる名詞1つしか残さないため、複合名詞をうまく扱えないなどの問題があった。

(2) 名詞が存在しない場合

有富らは、動詞と最後に現れる形容詞か副詞を1つだけ残して他の単語を削除し、動詞 + 形容詞1語、または副詞1語とする。動詞は単語列に含まれている場合のみ残す。提案手法では、有富らと同様に動詞、形容詞または副詞の選択をするが、動詞、形容詞および副詞が単語列に存在せず、その他の単語の削除を行うと単語列に何も残らない場合は、その単語を削除しない。

● 前置詞の後の単語列

(1) 動詞が前置詞の直後に存在する場合

動詞以外の単語を全て削除し、単語列を動詞1語とする。ここでいう動詞には、動名詞も含まれる。動詞が存在しない場合、次の(2)または(3)の処理を行う。

(2) 名詞が含まれている場合

名詞と冠詞以外の単語をすべて削除し、単語列を冠詞 + 名詞とする。名詞が複数含まれている場合は、最初に登場する名詞群を1つ残す。有富らは、ここでも最後に登場する名詞1つしか残さない。

(3) 名詞が含まれていない場合

有富らは、最後に現れる形容詞または副詞以外を全て削除し、形容詞1語または副詞1語とする。提案手法では、同様に削除を行うが、形容詞、副詞が存在せず、

“me”, “yours”といった人称代名詞の目的格、所有代名詞が存在する場合は、それらを残す。

有富らは前置詞の後の単語列において、“最後”の名詞を残すのに対し、提案手法では、前置詞の後の単語列の“最初”の名詞群を残す。名詞ではなく、名詞群を残すのは、前置詞の前の単語列と同じく、複合名詞等を扱うためである。また、適切な前置詞の選択において、前置詞の直後の単語との関係が重要と考え、前置詞の後の単語列に複数の名詞群が現れた場合、最も前置詞に近い名詞を残すため、“最後”でなく、“最初”の名詞群を残す仕様になっている。

また、提案手法では前置詞の前後の単語列が、なるべく空にならないようにするため、前置詞の前後の単語列において名詞が存在しない場合、前置詞の前の単語列にはその他の単語を、後の単語列には人称代名詞を残す仕様になっている。これは、単語列が空になると検索結果が膨大になることが多く、前置詞誤りをうまく自動検出、修正できないからである。

3.3 検索フレーズ

先に述べた単語の削除を行った後、前置詞の前後の単語列の間にワイルドカードを挿入して、検索フレーズとする。例として、“This country is very excellent in the information technology.”という英文から検索フレーズを生成してみる。まず、前置詞“in”の前後で分割し、“This country is very excellent”と“the information technology”という2つの単語列を得る。次に、前置詞前の単語列から動詞の前に存在する主語、助動詞を削除し、“is very excellent”を残す。この単語列には名詞が存在せず、副詞“very”と形容詞“excellent”が含まれるので、最後に登場する形容詞である“excellent”だけを残す。このように、前置詞の前の単語列からは“is excellent”という単語列が残る。そして、前置詞の後の単語列には名詞が存在するので、冠詞と、最初に登場する名詞群を残し、“the information technology”とする。この時、有富らの手法では冠詞と、最後に登場する名詞だけを残すので、“the technology”となる。こうして得られた2つの単語列の間にワイルドカードを挿入し、以下の検索フレーズを生成する。

- 提案手法
“is excellent * the information technology”
- 有富らの手法
“is excellent * the technology”

ここで、連続する2つの前置詞を含む場合は、片方の前置詞だけをワイルドカードに置き換えた2パターンの検索フレーズを生成する。例えば、“My hamster appeared from under the desk.”という英文からは、“appeared * under the desk”と“appeared from * the desk”の2つの検索フレーズを生成し、それぞれで検索を行う。

3.4 検索フレーズの修正

先に述べた手順で生成した検索フレーズで検索を行い、検索結果数が極端に少ない場合は、検索フレーズから単語を削除し、それを新たな検索フレーズとして、再検索を行う。具体的には、以下の条件で検索フレーズの修正、再検索を行う。

● 提案手法

検索結果数が5件未満のとき、検索結果が不十分とみなし、再検索を行う。

最初に不定冠詞を全て削除した後、以下の優先順位に従って、条件に当てはまる単語を順に削除して検索フレーズを修正する。ただし、一回の検索フレーズ修正において、削除する単語は1つだけとし、最初に削除する不定冠詞はこの削除数に含まない。また、検索結果は5件未満だが、それ以上単語の削除ができなくなった場合は、得られた5件未満の検索結果を使用する。

- (1) 前置詞の前の単語列の be 動詞を削除する。
- (2) 前置詞の前の単語列に2つ以上名詞が存在するならば、最初に現れる名詞1語を削除する。
- (3) 前置詞の後の単語列に2つ以上名詞が存在するならば、最初に現れる名詞1語を削除する。
- (4) 前置詞の前の単語列の動詞を削除する。

検索フレーズ修正の例を示す。先の例のフレーズ“is excellent * the information technology”で検索結果が不十分な場合、(1)の条件より“is”を削除し、“excellent * the information technology”という検索フレーズを生成し、再検索を行う。それでも検索結果が不十分だった場合は、(3)の条件より“information”を削除し、“excellent * the technology”という検索フレーズを生成し、再検索を行う。

● 有富らの手法

検索結果数が0の場合、動詞を削除し、一度だけ再検索する。例えば、“is excellent * the technology”で検索結果が不十分な場合、動詞“is”を削除し、“excellent * the technology”という検索フレーズで再検索を行う。

3.5 検 索

本研究では、有富らと同様にフレーズ検索を用いる。これにより語順を保ったままの検索を行うことができる。また、検索を行う際、言語とドメイン指定を行う。使用するドメインを、表1に示す。これは、英語を母語とする人が書いたと推測される英文を主要な検索対象とするためである。検索エンジンにはYahoo!デベロッパーネットワーク⁶⁾で提供されているYahoo!検索APIを使用する。

表1 使用するドメイン

.com	.net	.org	.info
.biz	.gov	.mil	.edu

3.6 前置詞の出現確率の計算

得られた検索結果から上位100件までのサマリを取得する。本研究で利用する英文修正支援システムでは、検索結果が不足した場合はサマリを取得せず、最後に行った検索の結果からサマリを取得する。取得したサマリから、検索フレーズのワイルドカード部分に相当する前置詞を抽出し、その前置詞の出現回数をそれぞれカウントする。その後、抽出した前置詞それぞれに対して、式(1)で前置詞の出現確率 P_i を求める。ここでは、最大100のサマリからこのようにして抽出した前置詞の総数を N 、そのうち前置詞 i の出現回数を $Prep_i$ とする。

$$P_i = \frac{Prep_i}{N} \times 100(\%) \quad (1)$$

4. 実装システム

図2は有富らが作成した英文修正支援システムのGUIに提案手法を実装して実行した画面例である。最上部のボックスに英文を入力する。その下のボタン2つは、左が前置詞誤り検出を開始するボタン、右側がシステムを初期画面に戻すクリアボタンである。下部には結

果を表示する．結果表示部分には，分割処理が済んだ英文を表示し，その中の前置詞の下に，出現確率の高い前置詞を順番に表示，その下に検索フレーズと検索結果数を表示する．この時あわせて表示する，表 2 に示す 4 段階評価は，出現確率と出現回数の 2 つに基づいて決定している．最も出現確率の高い前置詞を赤く表示し，逆にサマリ中に 1 回しか出現しないものは薄く表示する．出現確率が最も高い前置詞と入力文中の前置詞が違う場合，“[!]”をつけて注意を喚起する．また最下部に表示されている検索結果数は，生成した検索フレーズを検索エンジンで検索した時のヒット数を示している．この例は，“This burger is made of soybeans.” という前置詞誤りを含む英文を入力した時の実行例である．この文では“of”の使用が不適切であり，正解は“from”である．実装したシステムでも，“from”の出現確率が最も高いことを示している．

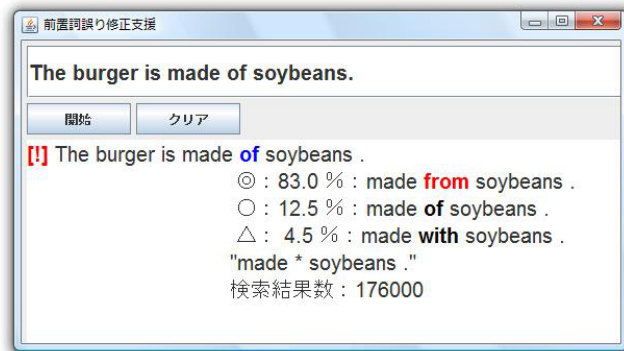


図 2 実行画面の例

表 2 表示する 4 段階評価

基準
出現確率が 50 % 以上
出現確率が 10 % 以上 50 % 未満
出現確率が 10 % 未満かつ出現回数が 2 回以上
出現回数が 1 回

5. 評価実験

以下の 2 つの評価項目について，有富らの手法と提案手法を比較するための実験を行った．

- (1) 提示する修正候補の適切性の評価
- (2) 前置詞誤りの自動検出性能と自動修正精度

5.1 前置詞誤り修正候補の適切性の評価

英文修正システムに前置詞を含む誤りのない英文を入力として与え，そこに含まれる前置詞を修正候補として提示できるかを実験により評価する．テストデータには英文法書の総合英語 Forest⁸⁾ の例文 100 文を用い，この中に含まれる前置詞 127 個を評価の対象とする．評価基準は，“システムが最も出現確率が高いと提示した前置詞と入力した英文中の前置詞が一致するか”と，“システムが 4 段階評価の ‘ ’ 以上として提示した前置詞のいずれかと入力した英文中の前置詞が一致するか”の 2 つとする．

結果を表 3 に示す．この表から，‘ ’ 以上の前置詞の中に正解が含まれれば良いという基準では，正解率は有富らの手法は 0.78，提案手法は 0.82 だった．

提案手法が上回った要因として，以下の検索フレーズ生成規則が有効だったことが挙げられる．

- 前置詞の前後の単語列を空にしない
- 検索フレーズに名詞を複数含むことができる

例えば，“How about having dinner on Christmas Eve.” という英文がある．まず前置詞 “about” の検討において，有富らの手法は “How” を削除して，“* having” という検索フレーズを生成し，この “about” を誤りとして誤検出してしまう．一方，提案手法では “How * having” という検索フレーズを生成し，“about” を誤検出しない．有富らの手法は，動詞，冠詞，名詞，副詞および形容詞以外の品詞は検索フレーズに含まない．よってこれらのいずれの品詞にも当てはまらない “How” は削除するが，これを削除すると前置詞の前の単語列が空になるので，提案手法ではこれを削除しない．こうして，“How * having” という検索フレーズを生成する．また，“on” の検討において，有富らの手法では前置詞の後の単語列に最後の名詞 1 つを残し，“having dinner * eve” という検索フレーズを生成し，結果的に “on” を

誤りとして誤検出する。これは、“Christmas eve”という複合名詞を検索フレーズに含められなかったのが原因である。一方、提案手法では前置詞の後の単語列に最初に出現する名詞群を残すので、“having dinner * Christmas eve”という検索フレーズが生成され、システムは“on”が正解であると正しく判定する。また、先の例にはないが、“July 7 2003”といった日付の表現も名詞群として扱え、複合名詞と同様に検索フレーズに含むことができる。

表 3 提示する修正候補の適切性

手法	条件	正解数	誤検出	正解率
提案手法	出現確率が最も高いものと一致	88	39	0.69
	以上の前置詞と一致	104	23	0.82
有富らの手法	出現確率が最も高いものと一致	77	50	0.61
	以上の前置詞と一致	99	28	0.78

一方、有富らの手法では正しく判定できるが、提案手法では誤検出する英文もあった。原因として、以下のものが挙げられる。

- 提案手法では検索結果数が不十分だと判断した 5 件未満の検索結果から有富らの手法が正しく判定し、再検索を行った提案手法では誤検出してしまった
- 検索フレーズ修正の際に不定冠詞を全て削除すると、それまで検索されていたページが検索されなくなる場合がある

最初の原因は、検索結果数が 5 件未満だった場合に起こりうる。提案手法では検索結果が 5 件未満の場合、検索結果を増やすため、特定の単語を削除して再検索を行うが、この時、得られた 5 件未満の検索結果を全て削除し、検索フレーズを修正後、再検索して得られた結果のみに基づいて判定を行う。一方、有富らの手法では、1 件でも検索結果が得られれば、その結果から判定を行う。確かに有富らの手法の方が、下の英文に近い検索フレーズから得られた結果のため信頼性が高いが、提案手法では再検索を行い、より多くの検索結果に基づいて判定することにしている。

もう 1 つの原因として、不定冠詞を削除したことが考えられる。提案手法では、再検索を行う場合、不定冠詞を削除する。本研究で利用する英文修正支援システムはフレーズ検索を用いているため、単語列中の途中の単語が削除されると、それまで検索されていたページが検索できなくなる。このような理由で、不定冠詞を削除しない有富らの手法では検索さ

れていたページが、提案手法では検索できなくなることがある。こうして、有富らの手法では正しく正誤判定ができていたものが、不定冠詞を削除したため、提案手法ではそれができなくなる場合があった。しかし逆に、不定冠詞を削除する提案手法でのみ正しく前置詞の正誤判定ができる英文も存在するため、不定冠詞削除の功罪については判断が難しいが、本稿の実験では、功の部分为上回った。

その他、正しい前置詞を示せなかった原因としては以下のものが挙げられる。

- システムが示した前置詞は例文とは異なるが不自然な英文にはならない
- 前置詞が文頭、文末にある

上記 2 つは有富らの手法と共通する問題である¹⁾。不自然な英文でないとは、例えば、“The book under the desk is not mine.”という文の“under”の部分は“on”でも正しいといった場合である。これは、英文の日本語訳など意味を与えないと修正は難しい。

文頭、文末に前置詞が存在する場合は、必然的に前置詞の前後のどちらか一方の単語列が空になり、検索結果が膨大になりやすく、うまく修正できない場合が多い。改善策としては、文頭に前置詞がある場合は前置詞の直後、文末の場合は直前の単語数個をそのまま抜き出して検索フレーズとするなど、提案手法とは別の方法で検索フレーズを作成することが考えられる。

5.2 前置詞誤りの自動検出と自動修正の精度

一定割合の前置詞誤りを含む英文を人工的に作成し、それをシステムに入力として与え、それぞれの手法を用いてどの程度誤りを自動検出、自動修正できるかを実験により評価する。テストデータには、New York Times⁹⁾の記事中の 50 文を用いる。この中に含まれる 200 個の前置詞のうち、100 個の前置詞を無作為に選んだ他の前置詞に置換することで誤りとし、この英文をシステムへの入力とする。

5.2.1 前置詞誤りの自動検出の方法と自動修正の正誤判定

有富らと我々の手法で、図 2 の修正支援システムを利用して、英文前置詞誤りの自動検出、および自動修正の正誤判定を以下のようにして行った。

- 自動検出の方法
システムが提示する出現確率上位 3 件の中に、入力した英文中の前置詞が含まれなけ

れば、誤りとして検出する。誤りでない前置詞を検出した場合は、誤検出となる。

● 自動修正の正誤判定

システムが提示した前置詞の、出現確率上位 3 件以内、かつ 4 段階評価が ‘ ’ 以上のものの中に正解の前置詞が含まれていれば、修正できたとみなす。

5.2.2 NativeChecker

NativeChecker⁷⁾ を使うことで、英語のフレーズの様々な項目について修正が可能である。そこで、NativeChecker による前置詞誤りの自動検出、修正を行い、提案システムと結果を比較する。NativeChecker は、英文をそのまま入力とする適切に前置詞の修正ができないので、提案手法により生成された検索フレーズのワイルドカードを、元の前置詞に置き換えたフレーズを入力とする。前置詞誤りの修正、検出を行うには、NativeChecker の修正方法のうち、“その他の表現”を利用する。そうして、入力されたフレーズの前置詞を、他の前置詞に置き換えた結果を表示する。NativeChecker による自動検出の方法と、自動修正の正誤判定は以下のように行う。

● 自動検出の方法

入力フレーズ中の前置詞が、修正候補の上位 3 件までに含まれない場合、誤りとして検出する。誤りでない前置詞を検出した場合は誤検出となる。また、修正方法“その他の表現”が利用できない場合は、前置詞の誤りの有無に関わらず非検出とする。

● 自動修正の正誤判定

NativeChecker が提示した修正候補の上位 3 件までの前置詞に、正解前置詞が含まれている場合、修正できたとみなす。

5.2.3 前置詞誤りの検出精度

前置詞誤りの自動検出の結果を表 4、表 5 に示す。提案手法を用いた場合は、自動検出の検出率 0.96、検出精度は 0.89、F 値は 0.92 であり、全てにおいて提案手法が有富らの手法を上回った。この主な理由は、提案手法では、検索フレーズに名詞群を含められることである。これにより 5.1 節で述べたように、名詞が複数並ぶ複合名詞や固有名詞、日付の表現などを適切に扱うことができる。検出精度は NativeChecker を下回ったが、検出率では大きく上回り、F 値でも上回っている。また、NativeChecker では修正方法“その他の表現”が使用できず、非検出としてカウントした例が、誤りありの英文で 12 件、誤りなしの英文で

も 12 件存在した。

表 4 前置詞誤りの自動検出結果

	誤っている前置詞		誤りのない前置詞	
	検出	非検出	非検出	誤検出
提案手法	96	4	88	12
有富らの手法	95	5	72	28
NativeChecker	68	32	94	6

表 5 前置詞誤りの自動検出性能

	検出率	検出精度	F 値
提案手法	0.96	0.89	0.92
有富らの手法	0.95	0.77	0.85
NativeChecker	0.68	0.92	0.78

5.2.4 前置詞誤り修正精度

5.2.3 項の実験で、前置詞誤りとして検出できた前置詞を対象に、自動修正を行った。提案手法では 96 個、有富らの手法では 95 個、NativeChecker では 68 個の前置詞が対象となる。結果を表 6 に示す。修正精度は、提案手法は 0.82、有富らの手法が 0.79、NativeChecker は 0.84 だった。この修正精度のみで判断すると NativeChecker が最良であるが、各手法で修正対象となる検出できた前置詞の数が異なることに注意すべきである。すなわち、前置詞誤りの検出から修正までを考慮すると、表 6 の“修正できた”前置詞数が性能判断の 1 つの基準となる。

提案手法を利用した場合に、うまく修正できなかった事例には次のものがある。

- システムが提示した前置詞は英文記事とは異なるが不自然な英文ではない
- 前置詞が文頭、文末にある

5.1 節で述べたのと同様の理由で、英文の和訳を与えるなどしない限り、不自然でない英文の自動修正は難しい。また、前置詞が文頭、文末にある場合も 5.1 節で述べたように、前置詞の前後の単語列のどちらか一方が必ず空になるため検索フレーズの単語数が少なくなりやすく、検索結果が膨大になり、うまく修正できないことが多い。

表 6 前置詞誤りの自動修正精度

	修正できた	修正できない	修正精度
提案手法	79	17	0.82
有富らの手法	74	21	0.79
NativeChecker	57	11	0.84

6. ま と め

本稿では、有富らが提案した英文前置詞誤り修正支援システムにおいて、前置詞誤りの自動検出、修正性能を決定づける検索フレーズ生成の改良方法を提案した。さらに評価実験により、提案手法の検出、修正性能を有富らの手法や NativeChecker のそれと比較した。実験では、前置詞誤り自動検出の F 値が有富ら 0.85, NativeChecker 0.78 に対し、提案手法は 0.92, 自動修正精度は有富ら 0.79, NativeChecker 0.84 に対し、提案手法が 0.82 であった。今後の課題として、検索フレーズ生成法の改良による、検出、修正性能のさらなる向上と、前置詞以外の品詞の誤りの検出と修正が挙げられる。

参 考 文 献

- 1) 有富隼, 太田学: “検索エンジンによる英文前置詞誤り修正支援”, DBSJ Journal vol.9 No.1, pp. 70-75, 2010.
- 2) M.Lapata, and F.Keller, “Web-based models for natural language processing”, ACM Trans. Speech and Language Processing, Vol.2, No.1, pp.1-31, Feb. 2005.
- 3) 平野孝佳, 平手勇宇, 山名早人, “検索エンジンを用いた英文冠詞誤りの検出”, 情報研報 (DBS), Vol.2007, No.65, pp.139-144, 2007.
- 4) 大鹿広憲, 佐藤学, 安藤進, 山名早人: “Google を活用した英作文支援システムの構築”, DEWS2005, 4B-i8, 2005.
- 5) Monty Tagger
<http://web.media.mit.edu/hugo/montytagger/>
- 6) Yahoo!デベロッパネットワーク
<http://developer.yahoo.co.jp/>
- 7) NativeChecker
<http://native-checker.com/native-checker/>
- 8) 塙タカユキ: “総合英語 Forest”, 桐原書店, 2006.
- 9) The New York Times
<http://www.nytimes.com/>