

検索エンジンを用いた英文前置詞誤り修正支援

有富 隼[†] 太田 学[‡]

[†] 岡山大学工学部情報工学科 〒700-8530 岡山県岡山市北区津島中 3-1-1

[‡] 岡山大学大学院自然科学研究科 〒700-8530 岡山県岡山市北区津島中 3-1-1

E-mail: [†] aritomi@de.cs.okayama-u.ac.jp, [‡] ohta@de.cs.okayama-u.ac.jp

あらまし 英作文を行う際、前置詞の選択に迷うことはないだろうか。使用する前置詞の妥当性を調べるには、その前置詞を含む英文のフレーズを検索エンジンで検索し、返ってきた検索結果数を利用してその妥当性を判断するという方法がある。本研究では、この方法に基づく英文の前置詞誤り修正支援システムを提案する。具体的に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分かりやすい提示を行う。実験では提案システムを使って前置詞誤りを自動で検出、修正したところ、誤り検出はF値0.85、誤り修正精度は0.82だった。さらにWeb上で英文修正サービスを提供するNativeCheckerと比較実験を行い、前置詞誤りの検出能力、修正精度ともにそれを上回ることを示した。

キーワード 検索エンジン, 英作文, 前置詞, 誤り検出, 誤り修正

Supporting Correction of Preposition Errors in English Using a Search Engine

Shun ARITOMI[†] Manabu OHTA[‡]

[†]Department of Information Technology, Faculty of Engineering, Okayama University

3-1-1 Tsushimanaka, Kita-ku, Okayama-shi, Okayama, 700-8530, Japan

[‡]Graduate School of Natural Science and Technology, Okayama University

3-1-1 Tsushimanaka, Kita-ku, Okayama-shi, Okayama, 700-8530, Japan

E-mail: [†] aritomi@de.cs.okayama-u.ac.jp, [‡] ohta@de.cs.okayama-u.ac.jp

Abstract Japanese people are sometimes at a loss what preposition to use in English composition. There is a way of confirming how appropriate used prepositions are by using a search engine. That is, we often 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. Based on this idea, this paper proposes a system for supporting correction of preposition errors in English by using a search engine. The proposed system automatically i) generates query phrases including a preposition, ii) collects summaries of the search results, iii) calculates occurrence probabilities of prepositions in the collected summaries, and iv) presents such information as prepositions and their probabilities to a user. We automatically detected and corrected preposition errors using the proposed system for experiment. The experimental results showed that F-measure was 0.85 in automatic error detection and the accuracy of automatic error correction was 0.82. Moreover, we also compared the proposed system to NativeChecker. The results showed that the proposed system outperformed NativeChecker in both detection and correction of preposition errors.

Keyword Search Engine, English Composition, Preposition, Error Detection, Error Correction

1. はじめに

我々日本人が英語の文章を作成する際、どのような表現を使えば良いか悩むことはよくある。英語を母国語としない日本人にはとりわけ前置詞の選択は、困難な場合がある。そのため通常は、辞書で調べたり、ネイティブスピーカーに質問したりすることでこのような問題に対処する。しかし、辞書には期待するような例文が存在しない場合もあるし、身近にネイティブス

ピーカーがいない場合も多い。

一方Web上には、英語を母国語とする人々が発信した膨大な文書コーパスが存在し、このコーパスから検討したい前置詞の出現頻度を調べれば、前置詞の妥当性を判断する一つの根拠になる。この作業は通常検索エンジンを利用することで可能となるが、適当な検索フレーズをその都度生成したり、多数の検索結果を参照したりするのはユーザにとって大きな手間である。

そこで本稿では、検索エンジンを利用した英文の前置詞誤り修正支援システムを提案する。提案システムは、検討したい英文を入力するだけで自動的に前置詞を含む適切な検索フレーズを生成、コーパス文書を収集、前置詞の出現確率の計算を行って結果をユーザに分かりやすく提示する。

本稿では 2 節で関連研究について述べる。3 節では提案システムについて説明し、4 節で提案システムの実装について動作例を示しながら具体的に述べる。5 節で評価実験と考察を行い、6 節でまとめる。

2. 関連研究

2.1. 検索エンジンを用いた英文冠詞誤りの検出

検索エンジンを用いた英文誤り検出の研究は前置詞だけでなく、冠詞でも行われている。Lapata ら[1]が 2005 年に提案した手法では、単純な検索フレーズで検索を行うことで冠詞誤りの検出を実現している。彼らの提案手法では、まず入力された英文から構文解析を用いて名詞句を抽出する。抽出した名詞の冠詞を {a/an, the, ϕ } に変化させて 3 つの検索フレーズを生成する。ここで ϕ は無冠詞を表している。この 3 つの検索フレーズのヒット数を検索エンジンにより取得し、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[2]の手法は、Lapata らの検索フレーズを拡張し、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、検出精度を向上させている。

2.2. 検索エンジンを用いた英文前置詞誤りの検出

検索エンジンを用いた英文前置詞誤りの検出は、早稲田大学の太田ら[3]が英作文支援システムの一部として実装している。まず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で置き換えて検索を行う。これにより、入力された前置詞とは別の前置詞を取得することができる。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、検索を行う。最後に、検索エンジンから返ってきた検索結果数を表示することで、ユーザに前置詞の妥当性を判断してもらう。欠点としては、応答時間が長い点、英文全体ではなく特定のフレーズをユーザ自身が考えて入力する必要がある点が挙げられる。

2.3. NativeChecker

NativeChecker[6]とは、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の修正を行うことができる Web サービスである。まず修正したい英語フレーズを入力するとそのまま入力フレーズが検索フレーズとなり、検索結果数を表示する。修正を行いたい場合、修正したい単語をポイントすることにより修正方法が提

示され、適切な修正方法を選ぶことで検索フレーズの修正を行うことができる。そして再度検索を行うことで、修正されたフレーズの検索結果数を得ることができる。修正できる内容は、スペルミス、類義語、その他表現、単複数形、to 不定詞、動名詞、過去分詞、単語の順序変更、単語の除去である。

2.4. 本研究の位置づけ

検討する前置詞を含む適切な検索フレーズをユーザが考えて入力することは面倒であり、場合によっては困難な作業となる。本研究では、入力として与えられた英文中のすべての前置詞について、自動的に適切な検索フレーズを生成し、その検索結果に基づきユーザに有用な情報を提示する。

3. 提案システム

本節では、提案する前置詞誤り修正支援システムについて述べる。以下 3.1 節ではシステムの概要を述べ、3.2 節で検索フレーズの生成方法について述べる。次に 3.3 節で利用する検索エンジンについて述べ、最後に 3.4 節で前置詞の出現確率の求め方を説明する。

3.1. システムの概要

提案システムの簡単な処理の流れを図 1 に示す。まず入力として検討したい英文を与える。次に、入力された英文を処理して検索フレーズを生成し、検索を行う。検索結果が 0 件の場合、検索フレーズを修正して再検索を行う。こうして検索エンジンの検索結果上位 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前置詞を並べ表示する。

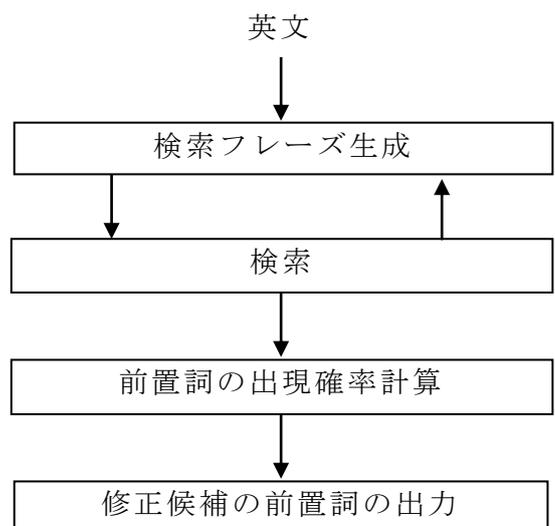


図 1 提案する修正支援システムにおける処理

3.2. 検索フレーズ生成

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。以下の(1)から(5)の手順で処理を行う。

- (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。この分割方法は3.2.1項で詳しく説明する。
- (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。品詞のタグ付けには Eric Brill の Monty Tagger[4]を使用する。
- (3)品詞タグに基づいて、英文から主語、助動詞をすべて削除する。
- (4)特定の形容詞、副詞、名詞のみが残るように他の単語を削除する。この単語の削除方法については(3)とあわせて3.2.2項で詳しく説明する。
- (5)前置詞部分をワイルドカードに置き換えて検索フレーズとする。

検索フレーズ生成において動詞を残す理由としては、動詞は前置詞との結びつきが強く、前置詞の前後の単語が同じでも動詞によって適切な前置詞が異なる場合が考えられるためである。また入力された英文もしくはそれが分割されたものが3単語以下の場合、単語の削除は行わず、前置詞をワイルドカードに置き換えるだけで検索フレーズとする。これは短いフレーズをさらに短くすると検索結果が膨大となり、期待する結果を得られないことが多いためである。

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、コンマの前後に分割する。また入力された英文が複数の前置詞を含む場合、その前置詞の数だけ英文を分割する。つまり入力された英文がコンマを含まず N 個の前置詞を含む場合、図2のようにして元の英文を分割して、検索フレーズの基になる単語列を抽出する。ここで S_1, \dots, S_N は入力文から抽出する単語列、 i は整数、 N は2以上の整数である。例えば、“He graduated from college in 1995.” という文からは、“He graduated from college”、“college in 1995” の2つの単語列が抽出される。ただし前置詞が2つ連続で出現する場合は、その2つの前置詞では分割せずにその他を上記の方法で分割する。

```
FOR i = 1 to N
  IF i = 1
     $S_i$  = 文頭から
      2番目の前置詞の直前の単語までの単語列
  ELSE IF  $1 < i < N$ 
     $S_i$  =  $i-1$ 番目の前置詞の直後の単語から
       $i+1$ 番目の前置詞の直前の単語までの単語列
  ELSE IF  $i = N$ 
     $S_N$  =  $N-1$ 番目の前置詞の直後の単語から
      文末までの単語列
  END IF
END FOR
```

図2 英文の分割処理

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。ただし連続する2つの前置詞をここでは1つとみなす。前置詞の前の単語列については、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。こうして生成された前後2つの単語列それぞれについて以下の処理を行う。

・前置詞の前の単語列

(1)名詞が含まれている場合

動詞と名詞と冠詞以外の単語をすべて削除し、動詞＋冠詞＋名詞1語とする。動詞と冠詞は、単語列に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、最後に出現する名詞以外は削除する。

(2)名詞が含まれていない場合

動詞と最後に出現する形容詞もしくは副詞を残し、動詞＋形容詞1語、もしくは動詞＋副詞1語となるようにする。動詞は含まれている場合のみ残す。

・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

動詞以外の単語をすべて削除する。動詞が前置詞の直後に存在しない場合は、以下の(2)(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、冠詞＋名詞1語となるようにする。冠詞は単語列に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、最後に出現する名詞を残す。

(3)名詞が含まれていない場合

最後に出現する形容詞もしくは副詞以外をすべて削除し、形容詞もしくは副詞1語となるようにする。

これらの処理を行った2つ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。

例えば、“Your viewpoint is totally different from mine.” という文からの検索フレーズ生成は次のようになる。まず前置詞の前後での分割により “Your viewpoint is totally different” と “mine” の2つの単語列を生成する。次いで、前置詞の前の単語列について動詞より前の単語をすべて削除し、“is totally different” と “mine” とする。前置詞の前の単語列には副詞 “totally” と形容詞 “different” が存在するので、最後に出現する “different” のみを残す。そして最後に2つの単語列の間にワイルドカードを挿入し、検索フレーズとする。つまり “is different * mine” が検索フレーズとなる。ただし連続する2つの前置詞を含む場合は、1つ目の前置詞と2つ目の前置詞のそれぞれをワイルドカード

で置き換えた 2 パターンの検索フレーズを生成する。例えば，“I went along with her.”からは“went * with her”と“went along * her”の 2 つの検索フレーズが生成される。

3.2.3. 検索フレーズの修正

3.2 節の手順で生成したフレーズを検索エンジンに入力しても、検索結果数が 0 件になってしまう場合がある。その場合、前置詞の出現確率の計算が出来ない。よってその場合は検索フレーズから動詞を削除して、新たな検索フレーズとすることで検索結果数を増やす。

3.3. 検索

本研究では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可能となる。また検索を行う際、言語とドメインの指定を行う。これは、英語を母国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。そのため言語は英語を指定し、検索対象ドメインについてもトップレベルドメインの一部を指定する(表 1)。そして結果として、ヒット数と検索結果上位 100 件のサマリを取得する。検索エンジンには Yahoo!デベロッパーネットワーク[5]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	.org	.biz	.mil
.net	.info	.gov	.edu

3.4. 前置詞の出現確率の計算

取得した 100 件の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、その前置詞の出現数をそれぞれカウントする。ワイルドカード部分には、前置詞でないものも含まれることもあるので、Monty Tagger を用いて前置詞の判定を行い、前置詞のみを抽出し、カウントする。そして抽出した前置詞それぞれに対して、式(1)で前置詞 i の出現確率 P_i (%) を求める。ここでは、 N を最大 100 サマリ中からこのようにして抽出した前置詞の総数、 $Prep_i$ をそのうちの前置詞 i の出現回数としている。

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

4. 実装システム

図 3 は実装システムの画面例である。最上部のボックスが英文を入力する部分である。その下の 2 つのボタンは、左が前置詞誤りの検出開始ボタン、右側がシステムを初期画面に戻すクリアボタンとなっている。最下部の結果表示部分には、検索フレーズ等を表示し、注目する前置詞部分の下に、出現確率の高い前置詞を

順番に表示している。加えて、表 2 のように 4 段階評価も表示することでユーザの修正判断がしやすいようにした。この 4 段階評価は、出現確率と出現回数に基づいて決定している。また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、“In this reason, all computers can perform floating-point arithmetic.”という前置詞誤りを含む英文を入力した時の実行結果画面である。この文では文頭の“In”の使用が不適切であり、正解は“For”である。提案システムでも、“For”の出現確率が最も高いことを示している。

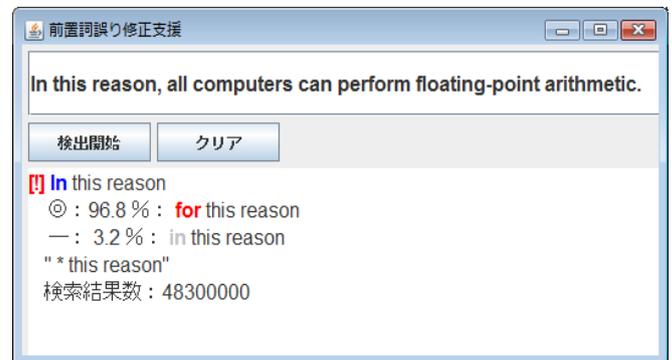


図 3 実行結果画面例

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

本節では、提案したシステムに対する評価実験について述べる。以下 4 つの項目について実験を行い、提案システムを評価する。

- (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
- (3) NativeChecker との比較
- (4) 専門英語テキスト[9]の英文訂正問題による実験

5.1. 修正候補の適切性の評価

提案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し、評価する。テストデータは、英文法書の総合英語 Forest[7]の例文 100 文を用い、こ

の中の 125 個の前置詞を評価対象とする。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、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可という基準では、87%の正解率が得られた。一方、うまく正しい前置詞を示せなかった原因としては以下の 2 つが挙げられる。

- ・文法的には正しいが意味が異なる
- ・検索フレーズの不備

文法的には正しいものとは、例えば、“My hamster appeared from under the desk.”という文の“under”の部分は“behind”でも正しいといった場合である。このような間違いは、英文の日本語訳など意味を与えてもらわないと修正は難しい。一方検索フレーズの不備では、年月日などを考慮した処理を行っていない事が原因と考えられる事例があった。具体的には、検索フレーズは前置詞の前後で名詞を 1 語ずつしか含むことができないので、“July 7”のような月日をうまく検索フレーズに取り込むことができない。このような事例については、検索フレーズの生成方法に改良の余地があると考えられる。

表 3 正解前置詞検出の結果

	正解数	誤検出	正解率
出現確率が最も高いものと一致	95	30	0.76
○以上の前置詞と一致	109	16	0.87

5.2. 前置詞誤りの自動検出・修正の精度

一定割合の前置詞誤りを含む英文を生成してシステムに与え、その誤りを自動的に検出、修正できるかを実験により評価する。テストデータとしては、New York Times[8]の記事中の 50 文を用いる。この中に含まれる前置詞 166 個の内 83 個の前置詞を無作為に選んだ他の前置詞に変更して誤りとし、これを提案システムへの入力文とする。

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

提案システムを利用した英文前置詞誤りの自動検出の方法と自動修正の正誤判定について説明する。

- ・自動検出の方法
入力した英文中の前置詞と、システムが提示する出

現確率の最も高い前置詞が一致しないとき、誤りとして検出する。よって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 4 段階評価の○以上として提示した前置詞の中に、正解前置詞が含まれている場合を修正できたとみなす。

以上の基準で実験を行う。

5.2.2. 前置詞誤り検出精度

結果を表 4, 表 5 に示す。自動検出の検出率は 0.95, 検出精度は 0.77, F 値は 0.85 という結果が得られた。うまく検出できなかった原因の 1 つとして、連続する名詞を考慮せずに検索フレーズを生成している点が挙げられる。提案システムは、前置詞の前後で名詞を 1 語ずつしか検索フレーズに含めない。そのため複合名詞が含まれる場合は、適切な検索結果を取得することができない。例えば“Some members of the Federal Reserve suggested the bank may have to expand measures to help mortgage markets.”という文の前置詞“of”を含む検索フレーズは“members * the Reserve”となってしまう。この問題に対処するには、固有名詞や複合名詞の扱いについて検討する必要がある。

表 4 前置詞誤りの自動検出結果

誤っている前置詞(83)		誤りのない前置詞(83)	
検出	非検出	非検出	誤検出
79	4	59	24

表 5 前置詞誤りの自動検出性能

検出率	検出精度	F 値
0.95	0.77	0.85

5.2.3. 前置詞誤り修正精度

5.2.2 項の実験で、前置詞誤りとして検出できた 79 の前置詞に対して、自動修正を試みた。結果は表 6 に示す通り、0.82 という修正精度だった。うまく修正できなかった原因としては以下の 3 つが挙げられる。

- ・検索フレーズの不備
- ・文法的には正しいが意味が異なる
- ・自動修正の判定や表 2 の 4 段階評価の妥当性の問題

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
65	14	0.82

1つ目の検索フレーズの不備については、5.2.2項でも述べたように固有名詞や複合名詞の扱いについて検討し、検索フレーズの生成法を改良することで対応していきたい。2つ目の文法的には正しい場合については、5.1節でも説明したように英文の日本語訳を与えてもらうなどしないと修正は難しい。3つ目は、自動修正の判定とも関係するが、4段階評価の△に正解が含まれている場合が存在した。現在4段階評価は単純に前置詞の出現確率などによって決定しているが、この決定基準についてもさらに詳しく検討する必要がある。

5.3. NativeChecker との比較

NativeChecker を使えば、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の修正を行うことができる。ここでは、NativeChecker と提案システムで前置詞誤りの自動検出と自動修正について比較実験を行う。NativeChecker は、英文全体を入力として与えようと大きく修正を行うことが出来ないため、提案システムで生成した検索フレーズを入力として与える。テストデータは5.2節の実験で使用した New York Times のデータを用いる。NativeChecker で前置詞誤りの検出と修正を行うためには、NativeChecker の修正方法の1つである“その他表現”を用いる。この修正方法により、入力フレーズの前置詞を他の前置詞で置き換えた結果を表示することが可能となる。

提案システムの自動検出の方法と自動修正の正誤判定は5.2.1項で示した通りである。一方NativeChecker におけ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する。

- ・自動検出の方法

入力フレーズ中の前置詞と修正候補の最上位の前置詞が一致しないとき、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

NativeChecker が提示した修正候補上位3件の前置詞の中に、正解前置詞が含まれている場合を修正できたとみなす。

誤り検出の比較結果を表7、表8に示す。表8から、自動検出のF値は提案システムが0.1上回っていることが確認できる。この理由の1つは、NativeChecker では文頭と文末の前置詞に対して、修正を行うことができない場合があるためである。

誤りの自動修正の比較結果は表9に示す。わずかに提案システムの方が高い修正精度であることが確認できる。他の相違点を挙げれば、提案システムでは検索フレーズを自動生成しており、修正箇所を指定する必要

表7 前置詞誤りの自動検出の比較

	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
NativeChecker	74	9	52	31
提案システム	79	4	59	24

表8 前置詞誤りの自動検出性能の比較

	検出率	検出精度	F 値
NativeChecker	0.89	0.65	0.75
提案システム	0.95	0.77	0.85

表9 前置詞誤りの自動修正精度の比較

	修正できた	修正できない	修正精度
NativeChecker	60	14	0.81
提案システム	65	14	0.82

がないため NativeChecker と比べて、ユーザの負担が少ない。

5.4. 専門英語テキストの英文訂正問題による実験

岡山大学工学部情報工学科で使用されている専門英語テキストにおいて出題された英文訂正問題を用いて実験を行う。実験に使用するデータは、このテキストの問題文のうち前置詞のみに誤りを含む12文を使用する。12文の中には26個の前置詞が含まれており、その内14個が誤りである。このデータをシステムに入力として与え、自動検出と自動修正の実験を行う。なお自動検出の方法と自動修正の正誤判定は5.2.1項で示した通りである。

自動検出実験の結果を表10、表11に示す。14個の前置詞誤りに関しては、すべて検出することができた。しかし、12個の正しい前置詞のうち、3件を誤検出した。この3件の原因を以下に示す。

- ・文法的には正しいが意味が異なる
- ・数式が含まれたため、検索結果が得られなかった
- ・検索フレーズのワイルドカード部分に該当する前置詞が100件のサマリ中に数個しかない

1つ目は、意味をユーザに入力してもらうなどしなければ検出は難しい。2つ目は、使用頻度が少ない単語の扱いについて今後詳しく検討していきたい。3つ目は、サマリ取得件数を現在の100件より増やせば、対応できると考える。しかし、取得件数を多くしすぎると処理時間が長くなってしまいうので、適切な取得件数について検討する必要がある。

自動修正実験の結果を表12に示す。14個の前置詞誤りに関して、修正できたのが9件、修正できなかったのが5件であった。この5件の原因を以下に示す。

- ・関係代名詞を含む文のため，うまく検索フレーズが生成できなかった
- ・使用頻度の低い語が含まれているため，十分な検索結果が得られなかった
- ・サマリより抽出した前置詞の種類が多すぎて，正解前置詞の出現確率が10%以上(表2で○以上)とならなかった

1つ目の原因は，構文解析などを利用して検索フレーズを生成することで対応できると考える．2つ目は，先の自動検出の考察でも述べたが，使用頻度が少ない単語の扱いについて詳しく検討する必要がある．3つ目は，検索フレーズの基になる単語列の文頭が前置詞であり，しかもその前置詞に続く単語が1語のみというフレーズで発生した．そのため抽出される前置詞の種類が多くなり，意図する結果を得ることができなかった．この問題に対処するには，文頭の扱いについて検討する必要がある．

表10 専門英語テキストの問題文における前置詞誤りの自動検出

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
14	0	9	3

表11 専門英語テキストの問題文における前置詞誤りの自動検出性能

検出率	検出精度	F値
1.00	0.82	0.90

表12 専門英語テキストの問題文における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
9	5	0.64

6. まとめ

本稿では，一般の日本人には判別が難しいと思われる英語の前置詞誤りの修正を支援するために，検索エンジンを利用して前置詞の妥当性を判断するための情報収集を行い，その結果を分かり易く提示するシステムを提案した．実験では，自動誤り検出のF値は0.85，自動修正精度は0.82だった．今後の課題は，検索フレーズの改良による自動検出と自動修正の性能の向上と，前置詞以外の誤りへの対応である．また将来的には，提案する修正支援機能を，エディタの英文文章校正機能やWebブラウザの拡張機能として提供したいと考えている．

参考文献

- [1] 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.
- [2] 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.
- [3] 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.
- [4] Monty Tagger
<http://web.media.mit.edu/~hugo/montytagger/>
- [5] Yahoo!デベロッパーネットワーク
<http://developer.yahoo.co.jp/>
- [6] NativeChecker
<http://native-checker.com/native-checker/>
- [7] 奥タカユキ, “総合英語 Forest”, 桐原書店, 2006.
- [8] The New York Times
<http://www.nytimes.com/>
- [9] 金谷健一, “専門英語テキスト—理数系の英語文献を読むために—”, 岡山大学大学院自然科学研究科, 2009.