挿入・欠落誤りを考慮した英文前置詞誤り修正支援システム

久保田　朗† 　太田　学†

† 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: †{kubota,ohta}@de.cs.okayama-u.ac.jp

あらまし　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．そこで，検討したい前置詞を含む英文から検索フレーズを生成して検索エンジンで検索し，検索結果中の前置詞の出現頻度から正解を判定する，英文前置詞誤り修正支援システムを開発してきた．本稿では前置詞の置換誤りに加え，欠落及び挿入誤りを修正対象に加えた修正支援システムを提案する．

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

A Support System to Correct English Preposition Errors Including Insertion and Deletion of Prepositions

Akira KUBOTA† and Manabu OHTA†

† Graduate School of Natural Science and Technology, Okayama University
3–1–1 Tsushimanaka, Kita-ku, Okayama-shi, Okayama, 700–8530 Japan
E-mail: †{kubota,ohta}@de.cs.okayama-u.ac.jp

Abstract　Japanese people are sometimes at a loss what preposition to use in English composition. Therefore, we developed a support system to correct preposition errors in English by using a search engine. The developed system automatically i) generates query phrases including a preposition, ii) collects summaries of the search results, iii) calculates appearance frequencies of prepositions in the collected summaries, and iv) presents applicable prepositions with their frequency information to a user. However, it can correct only substitution of prepositions. In this paper, we propose a method to correct insertion and deletion of prepositions to implement it into the developed system.

Key words　Search Engine, English Composition, Preposition, Error Detection, Error Correction

1. はじめに

普段英語を使わない日本人が英作文を行う際，前置詞の選択に迷うことがある．そのような場合，検索エンジンを用いて適切な前置詞を調べることがある．そこで有冨ら [1] は，検討したい英文を入力すると検索フレーズを自動生成して検索し，取得した検索結果のサマリにおける前置詞の出現率から修正候補を提示する，前置詞誤り修正支援システムを提案した．このシステムでは，適当な検索フレーズを自動生成して，適切な前置詞を含む検索結果のサマリを取得することが大変重要となる．そこで我々は，彼らの検索フレーズ生成法を改良し，前置詞誤りの検出と修正の性能を大幅に改善した [2]．しかしそれでも，入力された英文の前置詞に対して他の前置詞を修正候補として提示することしかできず、前置詞の置換誤りのみにしか対応していなかった．そこで本研究では，前置詞の欠落及び挿入誤りを対象とした誤り自動検出と修正を提案する．また，評価実験では英文校正サイト NativeChecker [8] との比較を行う．

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

前置詞の置換誤りを対象に，検索エンジンを用いて英文前置詞の誤りを自動検出，修正するシステムを，岡山大学の有冨ら [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，システムが検索フレーズを自動生成して検索し，その検索結果から前置詞の出現率を自動計算し，結果をユーザに提示する．また我々は有冨らのシステムを改良して複数の名詞に対応した検索フレーズ生成法を提案し，検出性能，修正性能の向上を確認した [2]．しかし，前置詞の欠落，挿入誤りには対応していない．

Gamon ら [4] は，英語学習者の誤りを含む英文とそれを正しく修正した英文から検索フレーズを生成し，検索エンジンに入力してその検索結果数を比較した．その結果，正しい前置詞を含む検索フレーズの検索結果数が，誤った前置詞を含む検索フレーズの検索結果数よりも，概ね多くなることを確認した．彼

らはまた前置詞の欠落や挿入，冠詞の誤用についても，誤った英文とそれを修正した英文を利用して，検索エンジンの返す検索結果数が正しい用法を支持することを実験により示している．彼らの実験は，検索エンジンが英語の誤りの発見に有効であることを示したが，誤りを含む英文と人が修正した英文の検索結果数を比較しているだけなので，実用的な誤り検出や修正はできない．一方本稿の提案は，人が修正した前置詞を用いることなく，誤りを含む可能性のある英文のみを入力として，前置詞誤りを検出し，さらにその修正を試みるものである．

前置詞誤り自動修正の研究は英語以外の言語でも行われており，例えば Hermet ら [5] はフランス語を対象とした前置詞誤りの修正を行った．彼らは，入力文に含まれる前置詞を “紛らわしい前置詞”に置き換えて検索し，その検索結果数を比較し，前置詞誤りの修正を試みた．評価実験では，第二外国語としてのフランス語学習者の誤りを含む文の修正を行い，約 7 割の精度で誤りを修正できることを確認した．

2.2 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，大鹿ら [3] が作成した．その機能の一部として前置詞の誤りの検出が実装されている．検討したいフレーズを入力すると，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て検索し，入力されたものとは異なる前置詞を取得する．また，英文翻訳サイトで入力した日本語を英訳し，用例として使用されている前置詞を取得する．その後，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し，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．欠点としては，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．

2.3 NativeChecker

NativeChecker [8] 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．検討したい英語の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，そのヒット件数が表示される．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合わ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ぶと，修正後のフレーズとそのヒット数が表示される．修正内容として，スペルミス，類義語，単複数形，to 不定詞，その他の表現，単語の除去，単語の順序の変更がある．

3. 前置詞の置換・挿入誤りの検出と修正

本稿では，有冨らが作成した前置詞誤り修正支援システムのプロトタイプを改良した．検索エンジンを使った英文前置詞誤りの検出及び修正の概要を図 1 に示す．まず検討したい英文を入力すると，システムは英文を処理して検索フレーズを生成し，検索を行い，結果からサマリを取得する．得られたサマリにおける前置詞の出現回数から出現率を計算し，出現率の高いものを正解とする．検索フレーズの生成法は 3.3 節，サマリ取得と出現率の計算については 3.5 節で述べる．

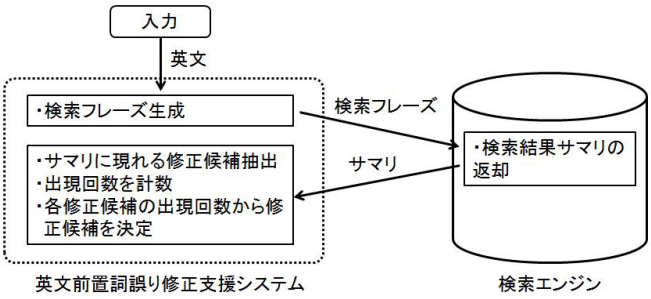

図 1 英文前置詞誤り修正支援の概要（置換・挿入誤り）

前置詞を含む英文を入力した場合，システムは前置詞の置換誤りと挿入誤りを検討する．そのため，入力された英文から，以下の手順で検索フレーズを生成する．

(1) 入力された英文がコンマを含む場合や，複数の前置詞を含む場合はそれらの前後で英文を分割する．

(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．タグ付けには Eric Brill の MontyTagger [6] 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 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，初期検索フレーズとする．

3.1 入力文の分割処理

入力された英文がコンマを含む場合は，その前後で分割する．また一文が複数の前置詞を含む場合は, [2] で述べたようにその前置詞の数だけ文を分割し，前置詞が一つだけ含まれる単語列を生成する．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．また，前置詞が連続して出現する場合，そこでは分割せず一つの前置詞として扱う．

3.2 単語の削除

分割処理後の単語列を，前置詞より前の単語列と後の単語列に分け，それぞれ以下のように処理する．

3.2.1 前置詞より前の単語列

まず，最後に現れる動詞より前の単語を全て削除した後，残った単語列中の名詞の有無で以下のように処理を分ける．

(1) 名詞が存在する場合

be 動詞以外の動詞，冠詞，名詞，人称代名詞所有格，目的格以外の単語を全て削除する．ここで，単語列中に 2 つ以上の連続する名詞（名詞群）が存在する場合は，名詞群を残す．名詞群や人称代名詞の目的格が複数存在する場合は，それぞれ最後に現れるもの一つを残す．また，動詞，冠詞，人称代名詞は，存在する場合のみ残す．

(2) 名詞が存在しない場合

名詞の代わりに最後に現れる形容詞または副詞 1 語を残し，他の単語を全て削除する．ただし，be 動詞以外の動詞，人称代名詞の目的格が存在していれば削除せずに残し，目的格が複数

存在する場合は，最後に現れるものを残す．ただし，形容詞及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が空になる場合は，本来削除されるべき単語であってもその単語を削除しない [2].

3.2.2 前置詞より後の単語列

(1) 動詞が存在する場合

to ＋動詞の原形，または前置詞の直後に動名詞が存在する場合は，動詞，動名詞より後の単語を全て削除する．動詞が存在しない場合，以下の (2)，もしくは (3) の処理を行う．

(2) 名詞が存在する場合

冠詞と，最初に現れる名詞群一つと人称代名詞の目的格，所有格を残し，他の単語をすべて削除する．冠詞と人称代名詞の目的格は存在する場合のみ残す．

(3) 名詞が存在しない場合

最後に現れる形容詞または副詞 1 語を残す．人称代名詞の目的格，または所有代名詞が存在する場合，これらの単語を残す．複数存在する場合は，最初に現れる 1 語だけを残す．

3.3 検索フレーズ

我々が [2] で提案した方法を用いて，初期検索フレーズ及び再検索フレーズを生成する．

3.3.1 初期検索フレーズ

3.2 節で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たものを初期検索フレーズとする．例として，“This country is very excellent in the information technology.”という英文からは，初期検索フレーズ “excellent * the information technology”が生成される．

3.3.2 再検索フレーズ

初期検索フレーズだけでは十分な数の検索結果が得られないことがしばしばある．そこで，その不足を補うため，初期検索フレーズに以下の (1)〜(3) の処理を個別に適用し，最大 3 種類の再検索フレーズを生成する．また，これらを再検索フレーズパターン (1)〜(3) と呼ぶ．

(1) 前置詞の前の単語列の最初の名詞 1 語

(2) 前置詞の後の単語列の最初の名詞 1 語

(3) 前置詞の前の単語列の動詞

先の例の初期検索フレーズ “excellent * the information technology”からは，再検索フレーズパターン (2) の処理により名詞 “information”が削除され，“excellent * the technology” という再検索フレーズが生成される．

3.3.3 前置詞が文頭または文末に存在する場合

前置詞が文頭及び文末にある場合は前置詞より前，または後の単語列が必ず空になるので，検索の手がかりとなる単語が少ないため，適切な誤り検出，修正ができないことが多い．そこで，我々が [2] で提案したようになるべく多くの語を残しながら検索フレーズを生成する．

3.4 検　　索

検索エンジンには Yahoo!デベロッパーネットワーク [7] で提供されている Yahoo!検索 web API を使用し，3.3.1 項で生成した初期検索フレーズ，及び 3.3.2 項の再検索フレーズそれぞれで検索を行う．本システムで検索を行う際は，語順が保てるようフレーズ検索を用いる．

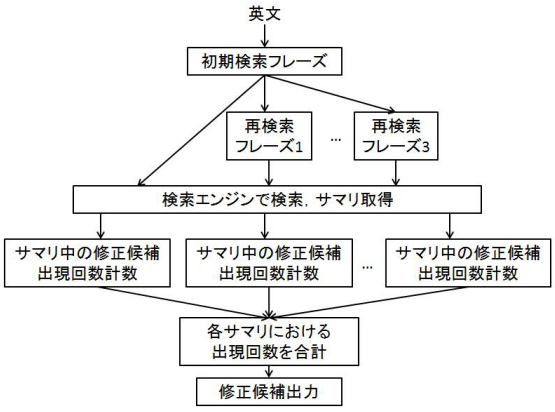

図 2　置換・挿入誤り修正のための検索フレーズ生成の流れ

3.5 修正候補の出現率

サマリの取得方法，システムが提示する修正候補の出現率の計算法及びサマリの重み設定について説明する．

3.5.1 修 正 候 補

提案システムでは，サマリから MontyTagger により前置詞とタグ付けされた単語及びφを抽出する．また，本稿では前置詞及びφをまとめて “修正候補”と呼ぶ．システムが出現率の高い修正候補として前置詞を提示した場合は置換誤り，φを提示した場合は挿入誤りをそれぞれ示唆している．

3.5.2 サマリ取得

3.3.1 項の方法で生成した初期検索フレーズから得た検索結果，及び 3.3.2 項の最大 3 種類の再検索フレーズによる検索から得た検索結果より，それぞれ上位から最大 100 件ずつサマリを取得し，合計最大 400 件のサマリを用いて修正候補の出現率を計算する．この処理を図 2 に示す．ただし，前置詞が文頭または文末にある場合は，3.3.3 項で説明した検索フレーズで検索し，検索結果のサマリを最大 100 件取得する．検索フレーズのワイルドカード部分には二つ以上の単語が入る場合もあり，前置詞やφ以外もあり得る．そのため，φの出現回数が前置詞に比べ相対的に少なくなりやすいので，ワイルドカード部分に冠詞が入る場合は，φとしてカウントする．また，取得したサマリに修正候補が含まれない場合は，修正候補を含むサマリを取得できるまで 100 件ずつ繰り返し取得する．ただし，サマリを 100 件ずつ 10 回取得しても修正候補が含まれるサマリを取得できなかった場合は，サマリ取得を打ち切る．これは，Yahoo!検索 Web API の結果返却先頭位置を 1000 を超えて指定することができないためである．

3.5.3 修正候補の出現率計算

取得した全て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修正候補を抽出し，各修正候補の出現回数をカウントする．ただし，カウントの際に各サマリの重みを考慮する．重みの設定方法については，3.5.4 項で説明する．その後，抽出した修正候補それぞれに対して，式 (1) で修正候補の出現率 R_i を求める．ここで，3.3.1 項の初期検索フレーズの検索結果のサマリにおける修正候補 $cand_i$ の出現回数を $Count_i$，全ての修正候補の出現回数を N とする．同様に，3.3.2 項の再検索

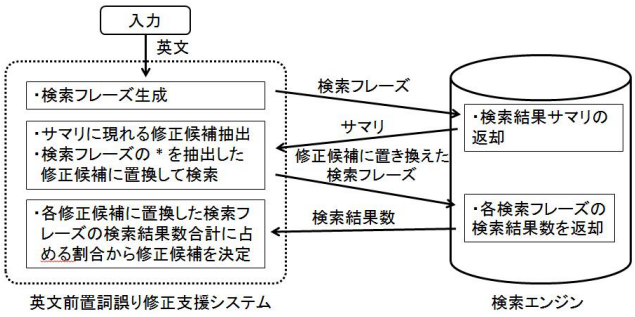

図 3　英文前置詞の欠落誤りの検出・修正の概要

フレーズパターン j の検索結果のサマリにおける修正候補 i の出現回数を $Count_i^{(j)}$，全修正候補の出現回数を $N^{(j)}$，重みを $w^{(j)}$ とする．また，初期検索フレーズの重み w は，式 (2) に示すように重みの合計が 1 になるように設定する．

$$R_i = \frac{wCount_i + \sum_{j=1}^{3} w^{(j)} Count_i^{(j)}}{wN + \sum_{j=1}^{3} w^{(j)} N^{(j)}} \times 100(\%) \qquad (1)$$

$$w + \sum_{j=1}^{3} w^{(j)} = 1 \qquad (2)$$

3.5.4　サマリの重み

初期検索フレーズの検索結果に，再検索フレーズパターン (1)〜(3) の結果を加えた場合を比較してそれぞれの修正候補の正解数の比を基に，(1)〜(3) のサマリの重みを決定した [2]．具体的には，再検索フレーズのサマリの重みを $w^{(1)} = 0.104$, $w^{(2)} = 0.113$, $w^{(3)} = 0.116$ とし，式 (2) から，初期検索フレーズの重み w を 0.667 に設定した．

4.　英文前置詞の欠落誤りの検出と修正

本節では前置詞の欠落誤りの検出及び修正について説明する．提案手法では，自動詞と共に用いられるべき前置詞の欠落誤りの検出と修正を行う．具体的には，入力された英文に動詞と名詞が連続して出現する場合，それらの動詞名詞間に前置詞の欠落誤りがないか検討する．

4.1　概　　要

検索エンジンを使った英文前置詞の欠落誤りの検出と修正の概要を図 3 に示す．検討したい英文を入力すると，システムは 4.2 節で述べる動詞と名詞からなるフレーズを一つだけ含むように英文を分割する．次に分割された英文を処理して，検索フレーズを生成し，検索する．その後検索結果のサマリを取得し，サマリ中に現れる修正候補を抽出する．得られた修正候補を使用した検索フレーズで検索し，得られた検索結果数から各修正候補の出現率を計算する．この検索フレーズ生成法は 4.3 節で説明する．最後に，出現率の高いものを修正候補として，ユーザに提示する．

4.2　動詞名詞フレーズ

be 動詞以外の動詞 (活用形を含む) と名詞群，動名詞または人称代名詞の目的格が連続して現れる箇所を，本稿では “動詞名詞フレーズ” と呼ぶ．ただし，動詞と名詞の間に冠詞，形容詞，副詞，人称代名詞の所有格がある場合も動詞名詞フレーズとする．

```
For i = 1 to N
    If i = 1 Then
        S_i = 文頭から 2 番目の動詞名詞フレーズの動詞まで
    Else If i = N Then
        S_i = N − 1 番目の動詞名詞フレーズの動詞の直後の
        単語から文末まで
    Else
        S_i = i − 1 番目の動詞名詞フレーズの動詞の直後の
        単語から i + 1 番目の動詞名詞フレーズの動詞まで
    End If
End For
```

図 4　英文分割処理（欠落誤り修正）

4.3　検索フレーズ生成

入力された英文から以下の手順で検索フレーズを生成する．

(1)　入力された英文がコンマを含む場合，それらの前後で英文を分割する．

(2)　(1) の処理後，英文が複数の動詞名詞フレーズを含む場合は動詞名詞フレーズを一つだけ含むように文を分割する．この処理については 4.3.1 項でさらに詳しく説明する．

(3)　分割処理後の英文に対して，MontyTagger [6] により品詞のタグ付けを行う．

(4)　付けられた品詞タグに基づき，英文から 4.2 節で定義した動詞名詞フレーズを抽出する．

(5)　抽出した動詞名詞フレーズの動詞の直後にワイルドカードを挿入し，欠落誤り検出及び修正のための初期検索フレーズとする．この初期検索フレーズについては，例を交えて 4.3.2 項で説明する．

4.3.1　入力文の分割処理

入力文がコンマ，ピリオドを含む場合はその前後で分割する．次に，分割した単語列に動詞名詞フレーズが複数含まれる場合は動詞名詞フレーズをひとつだけ含み，次の動詞名詞フレーズの動詞までを抽出する．分割の手順を図 4 に示す．動詞名詞フレーズを N 個含む入力を分割して抽出される単語列 S_i は，動詞名詞フレーズ vnp_i を一つだけ含み，一つ前に現れる動詞名詞フレーズ vnp_{i-1} を構成する動詞の次の単語から，次に現れる動詞名詞フレーズ vnp_{i+1} を構成する動詞までとする．動詞の前後の単語から単語列を構成するので，分割後の英文には重複する部分がある．例えば，“TV often insist controlling TV programs.” という入力からは，動詞 “insist” と名詞 (動名詞) “controlling” を含む “TV often insist controlling” と，動詞 (動名詞) “controlling” と名詞群 “TV programs” を含む “controlling TV programs” の二つの単語列が生成される．

4.3.2　初期検索フレーズ

4.3.1 項の方法で分割された英文から，動詞名詞フレーズを抽出する．抽出された動詞名詞フレーズを構成する動詞の直後にワイルドカードを挿入したものを初期検索フレーズとする．

例えば，“We can swim many places.”という英文からは，動詞名詞フレーズ “swim many places”が抽出され，動詞 “swim”の直後に直後ワイルドカードを挿入し，検索フレーズ “swim * many places”が生成される．

4.3.3 再検索フレーズ

初期検索フレーズで検索した結果，検索結果数が 0 件である場合や，検索結果数が膨大で，取得したサマリの中に 4.5 節で述べるような修正候補となる単語が存在しないことがある．そのような場合には初期検索フレーズを修正して再検索を行う．

a） 検索結果数が 0 件だった場合

初期検索フレーズの検索結果数が 0 件で，動詞より後ろの単語が冠詞を除いて 2 語以上存在する場合，動詞より後ろの単語列の先頭の単語を削除することで検索結果数を増やす．例えば，初期検索フレーズ “swim * many places”の場合，その検索結果が 0 件であれば，“swim * places”という再検索フレーズを生成する．

b） 修正候補を含むサマリを取得できなかった場合

初期検索フレーズの検索結果が膨大で，取得したサマリから修正候補となる前置詞やφが得られない場合は，検索フレーズを構成する単語を増やすことで検索対象を絞り込む．4.3.1 項で生成した単語列において，動詞より前の単語列中に最後に現れる名詞もしくは人称代名詞 1 語を初期検索フレーズの先頭に追加する．例えば，初期検索フレーズ “swim * many places”で検索し，サマリを取得したが修正候補となる単語を得られなかった場合，動詞より前の単語列 “We can”の最後の名詞 “We”をフレーズの先頭に追加し，再検索フレーズ “We swim * many places”を生成する．

4.4 検　　索

置換誤りと挿入誤りの修正同様，Yahoo!検索 Web API を使用し，フレーズ検索を用いて検索する．まず初期検索フレーズで検索し，サマリを 100 件取得する．初期検索の結果に応じて，4.3.3 項のいずれかの処理により再検索フレーズを生成し，再検索を行う．

4.5 修正候補の出現率

検索結果のサマリから，前置詞またはφを抽出する．欠落誤りの検出と修正においても前置詞とφを同列に扱い，これらをまとめて “修正候補”と呼ぶ．検索フレーズのワイルドカード部分を，抽出した修正候補に置き換え，その検索結果数を取得する．抽出した各修正候補 $cand_i$ に置き換えた検索フレーズの検索結果数 $\#r_i$ を計数し，式 (3) により修正候補の出現率 R_i を計算する．ここで N は獲得した修正候補の数を表す．また，検索から出現率計算の一連の処理を表したものが図 5 である．

$$R_i = \frac{\#r_i}{\#r_{all}} \times 100(\%) \tag{3}$$

$$\#r_{all} = \sum_{i=1}^{N} \#r_i \tag{4}$$

5. 実装システム

図 6 は英文前置詞誤り修正支援システムの実行画面の例である．最上部のボックスが英文入力部であり，ここに検討したい英文を入力する．英文に前置詞が含まれる場合，前置詞の置換誤り及び挿入誤りの検出と修正を試みる．その下のボタン二つは，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．下部の結果表示部分には，3.1 節の分割処理後の英文を表示し，その前置詞の下に，出現率の高い修正候補を順番に表示する．また，挿入誤りを意味する修正候補φを表す場合は，挿入誤りである前置詞の位置に下線を表示する．その下に初期検索フレーズと，最大 4 回の検索結果数の合計を表示する．修正候補の前に表示される表 1 に示す 4 段階評価は，出現率と出現回数に基づいて決定している．また，最も出現率の高い修正候補を赤く表示し，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．システムが出現率が最も高いと判断する修正候補と入力文中の前置詞が異なる場合，入力文の先頭に “[!]”をつけて注意を喚起する．図 6 では，“I laid to dolls on the sofa.”という前置詞誤りを含む英文を入力している．この文では正しくは “to”が不要であるが，図 6 でも，φの出現率が最も高いこと，即ち前置詞の挿入誤りであるということを示している．また，入力文中の “on”の用法は正しく，システムも修正候補最上位として “on”を提示し，置換誤り及び挿入誤りがないことを示している．

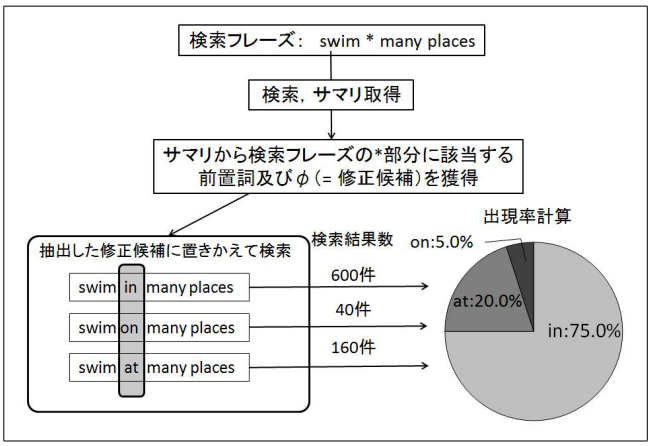

図 5　出現率計算（欠落誤り修正）

一方，動詞名詞フレーズを含む英文が入力された場合，前置詞の欠落誤りがないか検討する．図 7 はその実行画面である．表 2 に示す欠落誤りの 4 段階評価は，各検索結果数が全検索結果の合計に占める割合によって決定する．最も検索結果数の多い修正候補を赤く表示し，検索結果数が全体の 1 %に満たないものは薄く表示する．図 7 では，“She has just graduated a college.” という前置詞の欠落誤りを含む英文が入力されている．この文では “graduated”と “a college”の間に前置詞 “from”が必要であり，この図 7 でも，“from”が修正候補として示されていることがわかる．

6. 評 価 実 験

本稿で提案する手法を用いた英文前置詞誤り修正支援システムの有用性を示すため，評価実験により前置詞の置換誤り，挿入誤り及び欠落誤りの自動検出性能，自動修正精度を検証する．

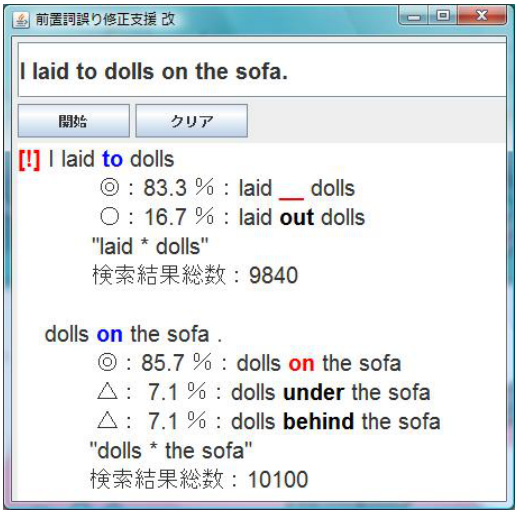

図 6　実行画面の例（置換・挿入誤りの修正）

表 1　置換・挿入誤りのための修正候補の 4 段階評価

	基準
◎	出現率が 50 %以上
○	出現率が 10 %以上 50 %未満
△	出現率が 10 %未満かつ出現回数が 2 回以上
—	出現回数が 1 回

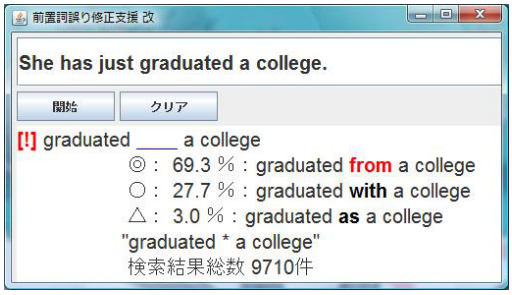

図 7　実行画面の例（欠落誤りの修正）

表 2　欠落誤りのための修正候補の 4 段階評価

	基準
◎	出現率が 50 %以上
○	出現率が 10 %以上 50 %未満
△	出現率が 1 %以上 10 %未満
—	出現率が 1 %未満

また，6.1 節で述べる前置詞の正解判定に検索結果数を用いる手法，及び NativeChecker と結果を比較する．

6.1　検索結果数を用いる手法

提案手法が取得したサマリにおける前置詞の出現頻度を用いるのに対し，[3], [5] の手法と同じように検索結果数から正解を判定する．3.3.1 項および 3.3.2 項で生成した検索フレーズのワイルドカードの前置詞を他の前置詞に置き換えて検索し，その検索結果数が多いものを正解とする．置き換える前置詞は，NewYorkTimes [9] の記事 906 文 (前置詞数 2606) から抽出した 71 種類の前置詞を使用する．

6.2　NativeChecker

NativeChecker による前置詞誤りの自動検出，修正を行い，提案手法を用いたシステムと結果を比較する．本稿の実験では，NativeChecker の修正項目 “その他の表現”及び “単語の除去” を入力文の前置詞に対して使用する．他の前置詞に置き換えたフレーズと前置詞を削除したフレーズのヒット数を提示し，入力文のヒット数と比較することで置換誤りや挿入誤りがないか判定する．また，NativeChecker への入力は，提案手法により生成された初期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とする．

6.3　前置詞の置換誤りの検出と修正

一定割合の前置詞誤りを含む英文を作成し，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 Times [9] 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 個の前置詞を他の前置詞に置換することで置換誤りとし，この英文をシステムへの入力とする．

6.3.1　置換誤りの自動検出および自動修正の正誤判定

提案手法，検索結果数を用いた手法及び NativeChecker で，前置詞誤りの自動検出と自動修正の正誤判定を以下のように行う．

- 提案手法

入力した英文中の前置詞とシステムが提示する出現率の最も高い前置詞が一致しないとき，誤りとして検出したとみなす．誤りでない前置詞を検出した場合は誤検出となる．また，システムが提示した前置詞の出現率上位 3 件以内，かつ 4 段階評価が ‘○’ 以上のものの中に正解が含まれる場合，修正できたとみなす．

- 検索結果数を用いる手法

入力された英文中の前置詞が，検索結果数最大の検索フレーズの前置詞と一致しない場合，誤りとして検出したとみなす．また，正解前置詞が検索結果数上位 3 件の検索フレーズに含まれる前置詞と一致すれば，修正できたとみなす．

- NativeChecker

修正項目 “その他の表現”によって提示された修正候補最上位のヒット数が入力フレーズのヒット数を上回る場合，誤りとして検出したとみなす．また，修正候補の上位 3 件以内に，正解前置詞を含むフレーズが含まれる場合，誤りを修正できたとみなす．

6.3.2　置換誤りの自動検出

前置詞の置換誤りの自動検出の正誤判定の結果および自動検出性能を表 3，表 4 にまとめる．提案手法を用いた場合，自動検出の検出率 0.990，検出精度は 0.818，F 値は 0.896 であった．NativeChecker は提案手法に近い性能を示しているが，NativeChecker を使う際の様々な制約を考慮すると，実質的には提案手法が優れていると考えられる．71 種類の前置詞を総当りで試す検索結果数を用いる手法と比較すると，検出率では下回り，検出精度では上回ったが，いずれも僅差であり，検出性能の差はほとんどないと言える．しかし，提案手法が 1 フレーズの検出と修正を平均 3〜4 秒で実行可能であるのに対し，検索結果数を用いる手法は 30 秒近くかかることもある．よってこの点では提案手法の方が優れていると言える．

表 3　置換誤りの自動検出結果

	置換誤りあり		置換誤りなし	
	検出	非検出	非検出	誤検出
提案手法	99	1	78	22
NativeChecker	97	3	77	23
検索結果数を用いる手法	100	0	77	23

表 4　置換誤りの自動検出性能

	検出率	検出精度	F 値
提案手法	0.990	0.818	0.896
NativeChecker	0.970	0.808	0.882
検索結果数を用いる手法	1.000	0.813	0.897

表 5　置換誤りの自動修正精度

	修正可	修正不可	修正精度
提案手法	83	16	0.838
NativeChecker	70	27	0.722
検索結果数を用いる手法	80	20	0.800

6.3.3　置換誤りの自動修正

6.3.2 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．表 3 から，提案手法は 99 個，NativeChecker は 97 個，検索結果数を用いる手法では 100 個の前置詞がそれぞれ対象となる．自動修正実験の結果を表 5 にまとめる．提案手法は，他の手法に比べ，良い結果を示している．ただし修正できなかった事例の中には，システムが提示した前置詞が正解とは異なるが，不自然な英文ではない場合があった．この場合は “修正できていない” と判定している．英文の和訳を与えるなどしない限り，このような事例の自動修正は難しい．また，入力英文やそれを 3.1 節の方法で分割した後の英文が極端に短い場合も，適切な修正が困難な場合があった．

6.4　前置詞の挿入誤りの検出と修正

前置詞の誤挿入を含む英文を入力として与え，システムが正しくその前置詞が不要であると判定できるかを実験により評価する．テストデータとして，日本人英語学習者コーパス NICE [10] から例文を 10 件取得した．さらに，“ゼロからスタート英文法” [11] に紛らわしい他動詞として掲載されている動詞をオンライン百科事典 Weblio [12] で調べ，その動詞を使った例文を 1 件ずつ，計 7 文取得した．この例文の動詞の直後に前置詞を挿入したものを挿入誤りとみなす．これらを合わせ，22 個の前置詞を含む 17 件の英文をテストデータとして使用する．このうち，11 個の前置詞が挿入誤りであり，これに対して自動検出，自動修正を試み，NativeChecker と結果を比較する．なお，提案手法及び NativeChecker の自動検出の方法と自動修正の正誤判定を，以下のように行う．

- 提案手法

入力した英文中の前置詞と，システムが提示する出現率の最も高い前置詞が一致しないとき，誤りとして検出したとみなす．誤りでない前置詞を検出した場合は，誤検出となる．また，システムが提示した前置詞の出現率上位 3 件以内，かつ 4 段階評価が ‘○’ 以上のものの中にφが含まれていれば，修正できたとみなす．

表 6　前置詞挿入誤りの自動検出結果

	挿入誤りあり		挿入誤りなし	
	検出	非検出	非検出	誤検出
提案手法	10	1	7	4
NativeChecker(検索フレーズ)	11	0	8	3
NativeChecker(分割後の英文)	11	0	5	6

表 7　前置詞挿入誤りの自動検出性能

	検出率	検出精度	F 値
提案手法	0.909	0.714	0.800
NativeChecker(検索フレーズ)	1.000	0.786	0.880
NativeChecker(分割後の英文)	1.000	0.647	0.772

- NativeChecker

入力されたフレーズの前置詞に対して，修正項目 “単語の除去” 及び “その他の表現” を使用する．提示された修正後のフレーズのヒット数を調べ，入力フレーズが，最もヒット数の多い修正候補と一致しない場合，誤りとして検出したとみなす．誤りでない前置詞を検出した場合は誤検出となる．また，“単語の除去” によって得られたフレーズのヒット数が上位 3 件以内であれば，挿入誤りを修正できたとみなす．

挿入誤りの検出及び修正の評価実験において，NativeChecker への入力は，3.3.1 項の初期検索フレーズのワイルドカードを元の前置詞に戻したフレーズと，3.1 節の分割処理後の前置詞を一つだけ含む英文，の二つとする．分割処理後の英文を用いて実験をするのは，提案手法と NativeChecker に同じ入力を与えて性能を比較するためである．

6.4.1　挿入誤りの自動検出

表 6，表 7 に示すように，挿入誤りの自動検出性能において，提案手法は検索フレーズを入力した NativeChecker よりも劣る結果となった．これらの表では，NativeChecker に提案手法の検索フレーズを入力した結果を “NativeChecker(検索フレーズ)”，分割後の英文を用いた結果を “NativeChecker(分割後の英文)” と表記している．提案手法では，挿入誤りを含む英文が前置詞を含んでも不自然な英文にならない場合，誤りを検出できなかった．また，提案手法において誤りのない例文を誤検出する理由として，検索フレーズを構成する単語数が少ないことが挙げられる．このような場合，検索結果数が膨大になり，取得したサマリの中に正解となる修正候補が含まれないことが多い．

一方，NativeChecker(検索フレーズ) は提案手法に比べ結果が良いが，NativeChecker にはもともと，入力文から検索フレーズを自動生成する機能が無いため，実験では提案手法の検索フレーズを入力として使用している．NativeChecker への入力として分割処理後の英文を用いた場合，即ち提案手法と NativeChecker にほぼ同じ入力を与えるという条件では，提案手法が検出精度，検出 F 値において NativeChecker を上回っている．

6.4.2　挿入誤りの自動修正

挿入誤りの自動修正においても，提案手法が Na-

tiveChecker(検索フレーズ) には及ばなかった．実験結果を表 8 にまとめる．提案手法では，挿入誤りの修正を行うために，修正候補としてφを抽出する必要があり，ワイルドカード部分にφが相当するようなテキストを含むサマリの取得が必須である．しかし，そのようなサマリを取得できない，またはφの出現回数が少ないことが多かった．この原因の一つに，φを含む検索結果は得られても，それが下位の検索結果である，ということが挙げられる．Yahoo!検索 Web API を用いて検索結果を取得する場合，最大 1000 件までしか取得できないので，検索結果数が 1000 件以上の場合には下位の検索結果からサマリを取得できない．このため，Web 上にワイルドカード部分にφが当てはまる文書が存在している場合でも，必ずしもφが存在するサマリを取得できるとは限らない．一方，NativeChecker は，修正項目 “単語の除去”により，必ずφを修正候補に含めることができる．また，NativeChecker は検索結果のサマリではなく，各前置詞を挿入した場合と前置詞を挿入しない場合の “検索結果数”を取得する．提案手法では，サマリは上位 1000 件以内の検索結果からしか取得できないが，NativeChecker には検索結果数を利用するためこのような制約は関係ない．これが，提案手法が NativeChecker に及ばなかった理由の一つと考えられる．ただし，修正精度でも，提案手法は NativeChecker(分割後の英文) は上回っている．

6.5 前置詞の欠落誤りの検出と修正

修正支援システムに前置詞の欠落誤りが存在する英文を与え，欠落箇所の検出の可否を実験により評価する．テストデータは，挿入誤り修正実験と同様，NICE 及び “ゼロからスタート英文法を用いて作成する．NICE からは合計 11 の動詞名詞フレーズを含む例文 9 文を選出した．また，“ゼロからスタート英文法”で紹介されている紛らわしい自動詞を含む例文を Weblio から 5 件取得し，その例文から前置詞を削除したものを欠落誤りとした．よって，計 16 件の動詞名詞フレーズ（うち 10 件が欠落誤り）を対象に実験を行った．提案手法の自動検出および自動修正の正誤判定を以下のように定める．

- 欠落誤りの自動検出

動詞名詞間において，システムが最も出現率の高い修正候補としてφでなく前置詞を提示すれば，欠落誤りを検出したとみなす．ただし，それらの間に前置詞を含まない表現が正解である場合は，誤検出となる．

- 欠落誤りの自動修正

自動検出実験において，誤りとして検出した動詞名詞フレーズを含む英文をテストデータとして使用する．システムが修正候補として示した上位 3 件以内に正解前置詞が含まれていれば，欠落誤りを修正できたとみなす．

誤り自動検出の結果を表 9，表 10 に，誤り自動修正の結果を表 11 にまとめる．検出 F 値は 0.947，修正精度は 0.889 であった．提案手法が提示した前置詞が例文とは異なるが不自然な英文ではない事例が 1 件あり，その欠落誤りは修正できなかった．

表 8　前置詞挿入誤りの自動修正精度

	修正可	修正不可	修正精度
提案手法	6	4	0.600
NativeChecker(検索フレーズ)	11	0	1.000
NativeChecker(分割後の英文)	6	5	0.545

表 9　前置詞欠落誤りの自動検出結果

	欠落誤りあり		欠落誤りなし	
	検出	非検出	非検出	誤検出
提案手法	9	1	6	0

表 10　前置詞欠落誤りの自動検出性能

	検出率	検出精度	F 値
提案手法	0.900	1.000	0.947

表 11　前置詞欠落誤りの自動修正精度

	修正可	修正不可	修正精度
提案手法	8	1	0.889

7. ま と め

本稿では有冨らの英文前置詞誤り修正支援システムで対象としていなかった，前置詞の挿入誤りおよび欠落誤りの検出と修正の方法を提案し，これを実装した．実験結果は挿入誤りの検出 F 値が 0.783，修正精度が 0.600，また欠落誤りの検出 F 値は 0.947，修正精度は 0.889 であった．挿入・欠落誤りの検出及び修正の実験より，提案手法のように Web の検索結果のサマリを取得し，その中の前置詞の出現頻度に基づき修正候補を抽出する場合は，検索フレーズを構成する単語を減らし検索結果数を増やすことは必ずしも効果的ではなく，反対に検索フレーズを構成する単語を増やし，検索結果を絞り込むことが有効である場合もあることが分かった．一方，本稿で実装した欠落誤りに対する修正機能は，動詞と名詞の間の前置詞の欠落にしか対応していないため，他の品詞間での欠落への対応が今後の課題と言える．欠落誤りの検出，修正は比較的性能が良かったので，検索結果のサマリにおける出現頻度のみではなく，検索エンジンの返す検索結果数も有効に利用する方法を検討したい．無論，前置詞以外の品詞の誤りの修正も課題として挙げられる．

文　献

[1] 有冨隼，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] 久保田朗，太田学: “検索エンジンを用いた英文前置詞誤りの自動検出と修正”，情報研報 (DBS), Vol.2011-DBS-153 No.2 , 2011.

[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] Michael Gamon and Claudia Leacock: “Search right and thou shalt find ... Using Web Queries for Learner Error Detection”，Proseedings of the NAACL HLT 2010 Fifth Workshop on Innovative Use of NLP for Building Educational Application，pp.37-44，2010.

[5] Matthieu Hermet, Alain Desilets, and Stan Szpakowicz. “Using the Web as a Linguistic Resource to Automatically CorrectLexico-syntactic Errors” Proceedings of the 6th Conference on Language Resources and Evaluation(LREC), pp.874-878, 2008.
[6] Monty Tagger
http://web.media.mit.edu/ hugo/montytagger/
[7] Yahoo!デベロッパーネットワーク
http://developer.yahoo.co.jp/
[8] NativeChecker
http://native-checker.com/native-checker/
[9] The New York Times
http://www.nytimes.com/
[10] NICE
http://sugiura5.gsid.nagoya-u.ac.jp/ sakaue/nice/features.html
[11] 安河内哲也：“ゼロからスタート英文法”，J リサーチ出版，2003.
[12] Weblio
http://www.weblio.jp/