

Web 資源を利用した冠詞の用法に基づく冠詞誤り自動修正

尾崎 弘明[†] 太田 学[‡]

[†] 岡山大学工学部 〒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: [†], [‡] {osaki,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語を使用する際によくある誤りの一つに、冠詞の用法の誤りがある。本稿では、Web 検索やオンライン辞書などの Web 資源を用いて、冠詞の用法に基づいて冠詞の誤りを自動修正するシステムを提案する。本研究では、24 の冠詞の用法をシステムが判断し、さらに検索エンジンから取得した検索結果数を考慮して入力された英文の冠詞を修正する。評価実験で、提案手法と検索結果のみを用いて修正を行う方法と修正精度を比較した結果、一般的な文章では精度に差は出なかったが、論文のように比較的使用頻度の低い単語が多く使われる文章においては、提案手法の精度が上回り、提案手法の有効性を示せた。

キーワード Web 資源, 冠詞, 誤り検出, 誤り修正

Automatic Correction of Article Errors Based on the Usage of Articles Using Web Resources

Hiroaki OSAKI[†] Manabu OHTA[‡]

[†] Faculty of Engineering, Okayama University, 3-1-1 Tsusima-naka, Kita-ku, Okayama, 700-8530 Japan

[‡] Graduate School of Natural Science and technology, Okayama University, 3-1-1 Tsusima-naka, Kita-ku, Okayama, 700-8530 Japan

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

Keyword Web resources, article, error detection error correction

1. はじめに

近年、グローバル化が進み、企業や学校などで、英語を使用する機会が増加している。大学でも海外の学会での発表や英語で論文を執筆することも多い。しかし、英語を母語としない日本人は英語の用法をしばしば誤ることがあるので、英作文の校正を自動で行うシステムが望まれる。特に日本語にはない冠詞の誤りは、日本人の書く英作文に多く含まれている[2]。

そこで本稿では、英語の冠詞の誤りを自動で修正する手法を提案する。提案手法は、岡山大学工学部情報工学科専門英語の講義で使用しているテキスト[1]の中の冠詞の用法に基づいて修正を行う。例えば、“The result can be shown by () simple computation.”という英文の()に入る冠詞を考える。computation は数えられない名詞だが、この場合 simple という対象を個別化する形容詞が computation を修飾している。そのため、他のもの(例えば、new computation など)と対比され、そういう性質をもつものという意味で個別化されるので、不定冠詞の a が入る。この例では、冠詞の用法を検出する際、対象を個別化する形容詞が名詞を修飾してい

るか否かで判断する。このような特定の単語に依存する冠詞の用法をシステムに実装した。またそのためには、このような対象を個別化する形容詞の単語リストなどが必要になる。本研究では、Web 上にそのようなリストがあればそれを利用し、ない場合は検索エンジンを利用して作成する。このようなリストの作成方法は 4 節で説明する。また、修正に利用する冠詞の用法については 3 節で説明する。本研究では[1]に書かれている冠詞の用法を修正ルールとして冠詞修正システムに実装するだけでなく、検索エンジンを用いて Web 検索し、その検索結果数も修正の判断に利用できるようにする。

本稿では、2 節で関連研究、3 節では冠詞修正システムに用いる冠詞の用法、4 節で提案する冠詞の修正方法を説明する。5 節で評価実験とその考察を述べ、6 節でまとめる。

2. 関連研究

冠詞の誤り修正に関する研究は、主にコーパスを用いるものと検索エンジンを用いるものがある。例えば

コーパスベースで冠詞の誤りを修正するものとして [5-8] が挙げられる．これらは，コーパスから統計量を抽出し，それに基づいて単語の出現状況をルールとして獲得することにより，そのルールを用いて冠詞の誤りを検出して修正する．問題点としては，コーパスの規模を大きくすることが困難である点と冠詞の例外的用法への対応が難しい点が挙げられる．

一方，検索エンジンを利用した冠詞の修正では，[3][4] などが挙げられる．これらは検索エンジンを用いて検索結果数を比較して冠詞を修正する手法を提案している．[3] では，コーパスベースのものよりも検索エンジンを用いたものの方が良い結果が出ることが報告されている．また[3]では，検索クエリのパターンを名詞の単数形と複数形の場合に分けて増やすことで，精度の向上が確認されているので，本研究でもそれを採用した．

他にも，[9]のような既出の名詞に着目した，最大エントロピー分類器を用いたものもある．

3. 冠詞の用法

本節では[1]を参考に修正システムに実装した冠詞の用法について説明する．まず名詞が可算か不可算か，単数形か複数形かに分けて冠詞の用法を述べ，次に不定冠詞，定冠詞，無冠詞の用法について説明する．

3.1 可算/不可算名詞と冠詞の用法

名詞が可算か不可算かによる冠詞の用法を表 1, 2 に示す．これらの表の○は可能な組み合わせで，×は通常不可能な組み合わせになる．名詞が可算か不可算かの情報は冠詞の修正において有用なので，冠詞修正の判断に利用する．また逆に，可算名詞か不可算名詞かを判断する上で表 1, 2 の情報が利用できる．

3.2 不定冠詞の用法

不定冠詞は基本的に複数あるものの一つを取り上げたことを表すのに使われる．用法としては，「一つ」という意味で one と同じ意味で使われたり，多数存在するものの中から任意に選んだ一つという意味を表したりする用法などがある．具体的な用法を以下に示す．

- (1) これまでに存在していなかった何かが新たに得られたり発生したりしたものを示す

例: From this, we can obtain **an** important theorem.

この場合生成や発見などを表す動詞と組み合わせられることが多い．

- (2) 不定冠詞と組み合わせられる構文や表現で用いられる

例: **There is a** situation in which this approach is necessary.

- (3) 対象を個別化する形容詞に修飾されている

例: The result can be shown by **a simple** computation.

3.3 定冠詞の用法

定冠詞は基本的に対象のものが一つしか存在しな

表 1 可算名詞と冠詞の組み合わせ

	a/an	the	φ (無冠詞)
単数形	○	○	○
複数形	×	○	○

表 2 不可算名詞と冠詞の組み合わせ

	a/an	the	φ (無冠詞)
単数形	×	○	○
複数形	×	×	×

いことを表す．3.2 節と同様に，具体的な用法を以下に示す．

- (4) 直前に述べたものを指す
- (5) 読者がみな知っていると思われるものの場合，了解を意味する形容詞で済ませるとき
例: the standard ... , the famous など
- (6) the ~ of ... の形で対象~の性質や特徴を示す
例: **The existence of** such function is obvious.
- (7) その名詞節の内容が that 以下に述べられていることを示す

例: This property is due to **the** assumption **that** the noise is Gaussian.

- (8) その個数だけ存在する基数を伴う名詞

例: When we view the line segment, **the two** endpoints may not be visible.

- (9) 普通名詞から転じた固有名詞

例: the United Nations, the White House など

- (10) 人名を冠した対象

例: the Fourier transform など

- (11) 最上級の前

例: It is **the easiest** way to prove this theorem.

- (12) 定冠詞と組み合わせられると決まったもの

例: the same as..., the past など

- (13) 特定の形容詞+of+名詞の複数形の形

例: some of the ... , most of the ... など

3.4 無冠詞の用法

名詞が無冠詞になるのは，上記の不定冠詞や定冠詞の用法の範疇に入らない場合といえる．用法を以下に示す．

- (14) 単数形なら不定冠詞のつく場合で名詞が複数形の場合

- (15) 不特定の数えられない名詞

例: rain, snow など

- (16) アポストロフィ s を伴う人名の所有格

例: Green's theorem など

- (17) 名詞が列記されていて，一まとまりの場合は冒頭の単語にのみ冠詞をつけて以降の名詞の冠詞を省略する

例: The symbols and **notations** used in this paper are as follows.

(18)動詞から転じた動作を示す名詞

例:Application of this identity proves the theorem.

(19)特定の形容詞がついた場合

例:In **every case**, **each variable** assumes a different value.

(20)記号を固有の対象として扱う時

例:Consider a line segment connecting **point P** and **point Q**.

(21)特定の数値を扱う時

例:Consider a circle of **radius 2**.

(22)学問や専門分野の名前

例:mathematics, vector analysis など

(23)慣用句的な表現

例:in general, for example など

(24)冠詞相当語に修飾されている

例:his, those などに修飾されている

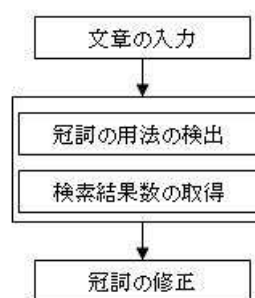


図 1 冠詞修正システムの概要

4. 冠詞誤り自動修正システム

本節では、提案する冠詞の自動修正システムについて説明する。

4.1 システム概要

ここでは、冠詞修正システムの処理の流れを説明する(図 1)。

まず、英文を入力する。既出の名詞には定冠詞が付きやすいので、入力する英語の文章は一文だけでなく、修正対象の文章全てを入力する。次に、入力文の名詞句に対して、名詞句ごとに 3 節で挙げた冠詞の用法に当てはまるかどうかを検出する。また同時に、検索エンジンを用いてその名詞句における 3 つの冠詞パターンの検索結果数を取得する。最後に、これら二つの結果を基に冠詞の適切性を判断し、不適切な場合は修正案を提示する。以下で冠詞の用法の検出、検索結果数の取得、冠詞の修正について詳しく説明する。

4.2 冠詞の用法の検出

入力文の名詞節に当てはまる冠詞の用法の検出方法について説明する。用法は 3.1 節で述べた可算/不可算による用法と 3.2~3.4 節の冠詞の用法(1)~(24)を調べる。

4.2.1 検出手法

3.1 節の表 1,2 に示したように名詞の可算/不可算の判別は冠詞を検討する上で有効である。よって、オンライン辞書を利用して可算/不可算を判断する。しかし、可算名詞と不可算名詞両方の意味がある場合は、冠詞の修正に用いない。

3 節の冠詞の用法の検出方法をそれぞれ説明する(表 4)。(4)の用法については、対象の名詞よりも前の文を見て判断する。詳しくは 4.2.2 節で述べる。(6)~(9), (11), (14)~(17), (20), (21), (24)は形態素解析の結果と文の形から検出する。(1)~(3), (5), (10), (12), (13), (18), (19), (22), (23)は特定の単語に依存するため、

あらかじめその単語を収集しリストを作成するか、あるいは、Web 上にある情報を利用する。(2), (13), (19)は[1]にある例文の構文や表現を用いた。(10)で用いる人名と(12), (18), (22), (23)については Web 上にあるリストまたはオンライン辞書から抽出した。(1)の生成、発生、発見、取得、導入、定義を表す動詞と(5)の了解を意味する形容詞は[1]に挙げられている例の類義語をオンライン辞書を用いて収集しリストを作成した。また、(1)で用いられる動詞の対象になる名詞も冠詞の修正に有用であると考え、リストを作成した。(1)の名詞と(3)の対象を個別化する形容詞は Web 上にリストがなかったので、検索エンジンを用いて単語を収集し、リストを作成した。単語の収集方法は 4.2.3 節で述べる。

4.2.2 既出の名詞

3 節の(4)の用法ではその名詞が既出かどうかをその名詞より前の文を調べて判断する。単に同じ単語が出現しているだけでなく、以下のような方法で調べる。

- 検討する名詞句に名詞が一つの場合は単にその名詞が出現しているか
- 検討する名詞句に名詞が熟語として存在している場合、その熟語または熟語の最後の名詞が出現しているか

4.2.3 単語の収集

3 節の(1)の動詞の対象となる名詞と(3)の個別化する形容詞は Web 上にリストがないので検索エンジンを用いて単語を収集し、リストを作成する。検索に用いるクエリは、[1]の例文にある収集対象の単語を含む名詞句と対象を個別化する形容詞はその前の単語を、(1)の動詞の対象となる名詞はその後の単語を用いる。集める単語の部分を実カードに置き換え、これを検索クエリとする。例えば“The result can be shown by a simple computation.”という例文から個別化する形容詞を集める場合、“by a * computation”となる。

(1)の場合は[1]の例文とオンライン辞書で集められた動詞を基に名詞のリストを作成する。手順を以下に

示す．

- (a) [1]の例文から収集対象の単語をワイルドカードに置き換えた検索クエリを生成する
- (b) (a)のクエリを用いて検索し，検索結果のサマリからワイルドカードにマッチした名詞を抽出する．
- (c) 抽出した名詞が(1)の用法として使われるかを確認するために，ワイルドカードを抽出した名詞に置き換えた検索クエリで検索し，冠詞の部分のを他の冠詞に置き換えた場合の検索結果数と比較する．
- (d) (c)で得られた検索結果数の一番多い冠詞が元の冠詞と同じであれば，その名詞をリストに加える．
- (e) 検索クエリの動詞をあらかじめ集めておいた動詞と置き換え，これを新しい検索クエリとする．そして，(b)～(d)の処理を行う．これをオンライン辞書で集めた(1)で用いる動詞全てで行う．

上で説明した手順を例文 “We want to find a solution of this equation.” を用いて説明する．(a)では，“find a * of”を検索クエリとする．(b)では検索結果のサマリから*が名詞のものを集める．例えば，discovery という名詞を収集できたとする．(c)で，*に discovery を挿入し，元の不定冠詞の場合と冠詞を定冠詞，無冠詞に変更して検索結果数と比較する．そして，(d)の通り不定冠詞の検索結果数が最も多ければ discovery を名詞のリストに追加する．次に(e)で，find を locate などの別の動詞と入れ替え(b)～(d)の手順でこのような名詞をさらに収集する．

一方，(3)の個別化する形容詞の収集では，[1]の例文のみを用いて検索を行う．以下に手順を示す．

- (a) 検索クエリを例文から作成する
- (b) (a)のクエリで検索し，検索結果のサマリからワイルドカードにマッチした形容詞を抽出する．
- (c) これらの形容詞がその用法として使われるかを確認するために，ワイルドカードを収集した形容詞に置き換えた検索クエリで検索し，さらに冠詞の部分のを他の冠詞に置き換えた場合の検索結果数と比較する．
- (d) (c)で得られた検索結果数の一番多い冠詞が元の例文の冠詞と同じであれば，その形容詞をリストに加える．
- (e) 一つの例文からなるべく多くの形容詞を収集するために検索クエリの形容詞につづく名詞一つを変更する．まず，ワイルドカードを(d)でリストに加えた形容詞と置き換え，置き換えた形容詞の後の単語一つをワイルドカードに置き換える．次にそれを検索クエリとして検索し，ワイルドカードにマッチした名詞を抽出する．抽出した名詞をワイルドカードと置き換え，再び形容詞をワイルドカードに置き換えて新しい検索クエリとする．そし

て，再度(b)～(d)を行う．これを最初の検索で収集した形容詞全てで行う．

3節の(3)の対象を個別化する形容詞の例文“The result can be shown by a simple computation.”を用いてこの処理を説明する．(a)では，“by a * computation”を検索クエリとする．(b)では検索結果のサマリから*が形容詞のものを集める．ここで例えば new という形容詞を抽出できたとする．(c) “by a * computation”の*に new を挿入し，不定冠詞の a の場合とそれを定冠詞，無冠詞に変更した場合を比較する．そして，a の検索結果数が多いとき，new を対象を個別化する形容詞としてリストに追加し，冠詞の用法の検出に利用する．次に(e)で，“by a simple computation”において simple を集めた形容詞に置き換えた検索クエリを用いて，computation 以外の名詞を収集する．例えば，先ほど得られた new を simple と置き換え，“by a new * ”という検索フレーズを生成し，今度は computation に相当する名詞を収集する．例えば，method が収集できたとすると，今度は “by a * method” というフレーズを生成して，(b)～(d)の手順で形容詞をさらに収集し，対象を個別化する形容詞のリストに追加していく．

4.3 検索結果数の取得

検索エンジンを用いて不定冠詞と定冠詞，無冠詞の場合の検索結果数を取得する方法について説明する．

4.3.1 検索クエリの生成

検索クエリは名詞句とそれに名詞句の一つ前の単語を加えたものを用いる．それに名詞句の最後の名詞を単数形と複数形に分けた 6 パターンを検索クエリとする．“I read a book.”を例に book について検索クエリを生成した例を表 3 に示す．検索クエリを単数形と複数形に拡張した 6 パターンで検索をした方が良い結果が得られるという報告[3]から，これを採用した．

4.3.2 検索手順

表 3 に示した検索クエリを用いて検索を実行する．以下に検索の手順を示す．

- (a)検索クエリの生成
- (b)検索エンジンを使い，6 パターンの検索クエリ検索結果数を取得する
- (c)検索結果数が閾値を下回った場合，検索クエリの単語を一つ減らし，(b)に戻り再検索をする
- (d)検索結果数が閾値を超えた場合，それを結果として利用する．

本研究では閾値は 0 に設定した．(c)において 6 パターンの検索クエリで閾値を下回った場合，複合名詞の場合は前の名詞を減らし，単体名詞が場合は，修飾している形容詞や副詞などを削除して再検索を行う．

4.4 冠詞の修正

4.2，4.3 節の結果を基に修正を行う．まず，4.2 節の

表 3 検索クエリの作成例

	a/an	the	φ (無冠詞)
単数形	read a book	read the book	read book
複数形	read a books	read the books	read books

冠詞の用法の検出の結果によって以下の 3 パターンに分類できる。

- (1) 検出された用法によって使用すべき冠詞が一意に決まる。
- (2) 検出された用法から複数の冠詞の修正案が提示される。
- (3) 用法が検出されない。

結果が(a)の場合はそのままその冠詞を修正案として提示する。結果が(b)の場合は考えられる冠詞を修正候補とし、その修正候補を 4.3 節の検索結果数を用いて比較し、修正案を決定する。結果が(c)の場合は、表 1, 2 に従って不定冠詞、定冠詞、無冠詞を検索結果数で比較し、修正案を提示する。

例えば(c)で、名詞が複数形とだけ分かった場合は、定冠詞と無冠詞の場合が考えられる。この場合、定冠詞と無冠詞の場合の検索結果数を比較し、修正案を決定する。また、(b)の場合では用法の組み合わせによっては例外的に冠詞が一意的に決まるものもある。その場合はあらかじめ実装しておいた例外パターンで修正案を決定する。例えば “The result can be shown by * simple computation.” を考える。この場合 simple は「対象を個別化する形容詞」で、computation は「数えられない名詞」となっているが、[1]には、数えられない名詞を個別化する形容詞が修飾した場合、不定冠詞とするという規則がある。このように、例外的な規則がある場合はそれを優先し、そうでない場合は、検出された用法の候補を検索結果数で比較し、それを修正案とする。また、優先度の高い用法についてはそれを優先し、修正案を決定する。もしそれで決められない場合は優先度のない冠詞の用法を検出した場合と同じ扱いで、修正案を決定する。優先する冠詞の用法を以下に示す。

- (16) アポストロフィ s を伴う人名の所有格
- (21) 特定の数値を扱う時
- (23) 慣用句的な表現
- (24) 冠詞相当語

5. 実験

実験では、4.2.3 節で述べた用法の検出に必要な名詞および形容詞のリストの作成と、冠詞誤りの修正精度の評価実験について述べる。

5.1 リストの作成

3.2 節の(1)で用いられる動詞の目的語になる名詞と

(3)の個別化する形容詞のリストを作成した。検索エンジンは Yahoo!検索 WebAPI[10]を用いた。その結果、表 5 で示したように、(1)では 1,036 の名詞、(3)では 3,458 の形容詞を収集した。

収集した名詞および形容詞の例を挙げる。(1)の名詞では、solution, method などが収集できた。これは発見や定義などを表す動詞の目的語になっていたため収集されたと考えられる。(3)の個別化する形容詞はさらに多く収集できた。また、収集の際同じ形容詞が何度も出てくることが多かったので信頼性は高いといえる。ここでは、vivid, difficult などの形容詞が収集された。

また、収集した名詞および形容詞の中にはいくつか品詞が違うものなども含まれるが、用法の検出は形態素解析によって入力文中の品詞を判断してからリストと照合するので問題ない。

5.2 冠詞誤りの修正実験

評価実験では、冠詞誤りを含む英文に対して冠詞の自動修正を試みてその修正精度を評価した。本実験では、[1]の冠詞を扱った節に挙げられている英文と New York Times[14]の記事、論文の抄録を修正対象として用いた。実験では(i)これら実験対象の英文の誤りを含まない正解データと同じ冠詞を修正案として提示できるか、(ii)冠詞のうち半数をランダムに誤った冠詞(無冠詞も含む)に置き換え、それを修正できるかを評価した。本手法は単語に依存しているので、スペルミス、冠詞以外の単語の誤りはないと仮定する。また、実験は提案手法と検索結果数のみを用いて修正を行った場合を比較した。本実験でもリストの作成と同様に検索エンジンは Yahoo!検索 WebAPI[10]を用いた。オンライン辞書としては Weblio 英和英辞書[11]を用いた。また、人名については欧羅巴人名録[12]、米国国勢調査局 1990 Census Name Files[13]を用いた。

5.2.1 評価尺度

評価尺度として以下の 3 つを用いる。F 値の P は適合率、 R は再現率のことである。

$$\text{再現率} = \frac{\text{正しく修正された冠詞誤りの数}}{\text{実際の冠詞誤りの数}}$$

$$\text{適合率} = \frac{\text{正しく修正された冠詞誤りの数}}{\text{修正された誤りの数}}$$

$$F\text{値} = \frac{2PR}{P+R}$$

5.2.2 実験結果

本実験では検索結果数のみを用いて冠詞を修正した場合と提案手法との比較を行った。検索結果数のみを用いる修正は、表 3 のような 6 パターンの検索結果数を比較し、検索結果数の最も多い冠詞を修正案とす

表 4 検出方法一覧

検出方法	冠詞の用法
前方参照	(4) 直前に述べたものを指す
形態素解析, 文の形	(6) the ~ of …の形で対象~の性質や特徴を示す (7)その名詞句の内容が that 以下に述べられていることを示す (8)その個数だけ存在する基数を伴う名詞 (9)普通名詞から転じた固有名詞 (11)最上級の前 (14)単数形なら不定冠詞のつく場合で名詞が複数形の場合 (15)不特定の数えられない名詞 (16)アポストロフィ s を伴う人名の所有格 (17)名詞が列記されてるとき (20)記号を固有の対象として扱う時 (21)特定の数値を扱う時 (24)冠詞相当語
[1]の例文から作成したリスト	(2) 不定冠詞と組み合わせられる構文や表現で用いられる (13)特定の形容詞+of+名詞の複数形の形 (19)特定の形容詞がついた場合
Web 上にあるリスト, オンライン辞書	(1) 新たに得られたり発生したりしたものを示す(動詞) (10)人名を冠した対象 (12)定冠詞と組み合わせられると決まったもの (18)動詞から転じた動作を示す名詞 (22)学問や専門分野の名前 (23)慣用句的な表現
検索エンジンを利用して作成したリスト	(1) 新たに得られたり発生したりしたものを示す(名詞) (3) 対象を個別化する形容詞に修飾されている

る。また、検索結果数の最も多い冠詞が二つ以上ある場合には、修正案を一つに絞れないので修正を行わない。様々な文章で評価を行うため、修正実験の対象として 3 種類の文章を用意した。それぞれの実験結果を以下に示す。

5.2.2.1 専門英語教科書の英文を用いた実験

[1]にある英文を用いて実験を行った。この例文は全 30 文で構成され、145 の検討すべき名詞句があった。また、この英文はそれぞれ単文でお互いに関連はないので、3.3 節の(4)の用法は同一文内の前後関係をみて判断した。実験結果を表 6, 7 に示す。表 6 は得られた修正案が正解データと一致しているか、表 7 は冠詞誤りの修正実験の結果で、無作為に正解データの冠詞の半数(73 箇所)を誤った冠詞に入れ替えて実験を行った結果である。

[1]の英文は、提案手法で用いた冠詞の用法を説明する章の練習問題になっているので、英文中に検出できる冠詞の用法が多くあった。表 6 や表 7 の評価では、検索結果数を用いた手法に比べ高くなったが、その点は留意する必要がある。また、[1]は学術論文を英語で書くためのテキストでもあるので、その例文では使用頻度の低い単語が多く使用されていた。そのため、検索結果数を用いる手法の修正精度は低くなったと考え

られる。

本実験で、両手法で修正案を誤った箇所や提案手法のみ誤った箇所、提案手法のみ正解した箇所などについて述べる。検索結果数のみを用いた手法は、使用頻度の低い単語が使われている場合に間違いが多かった。また、検索結果数が十分でも修正案を間違えることも多少あったが、この場合、正解と修正案の検索結果数は近いことが多かった。両手法共通で間違っていた箇所は、提案手法で用法が検出されていない、または検出されているが複数の用法が検出されていて修正案を決定できずに検索結果数を用いて修正している場合が多かった。提案手法のみで誤っている箇所については、4.4 節で挙げた優先する冠詞の用法以外が検出されている場合で、特に 5.1 節で作成した名詞および形容詞のリストを用いる(1), (3)の用法が検出された場合と(6)の用法が検出された場合が多かった。しかし、提案手法のみ正解している箇所でも、冠詞の用法の(1), (3)が検出されている場合が多かった。よって、作成した名詞および形容詞リストは有効だが、品詞の収集方法等に改善の余地がある。

5.2.2.2 New York Times を用いた実験

New York Times の記事一つを用いて実験を行った。この記事は 51 文で構成され、398 の検討すべき名詞句

表 5 作成した形容詞および名詞のリスト

リスト	個数
(1)の名詞	1,036
(3)の形容詞	3,458

表 6 修正案の適切性評価([1]の英文)

	正解	誤り	正解率
検索結果数のみの手法	102	43	0.70
提案手法	117	28	0.81

表 7 冠詞誤りの修正性能([1]の英文)

	再現率	適合率	F 値
検索結果数のみの手法	0.71	0.60	0.65
提案手法	0.81	0.72	0.76

があった。表 8, 9 にそれぞれ表 6, 7 と同様の実験結果を示す。[1]の英文に対する実験と同様に表 8 は正しい英文を入力として与えて得られた修正案がその正解と一致しているか(正しいものを正しいと出せるか), 表 9 は冠詞誤り修正実験の結果で、無作為に冠詞の半数を誤った冠詞に置換して実験した結果を表す。

New York Times[14]では一般的に使われる単語が多いため、十分な検索結果数を得ることができ、[1]の英文の場合より検索結果数のみを用いた手法の精度が高くなった。提案手法も冠詞の用法の検出で修正案を決定できない場合は検索結果数を用いるが、表 9 で適合率が少し高くなっている。また、表 8 で二つの手法の正解率は同じになったが、検索結果数のみを用いた手法において検索結果数が同じで最多の冠詞が二つ以上ある場合が、[1]の英文では誤り 43 件のうち 2 件だったが、この場合 63 件のうち 27 件もあった。

具体的な修正案を調べると[1]の英文の場合と大体同じような例が多かった。

5.2.2.3 論文抄録を用いた実験

論文抄録を用いた実験では、英語圏の著者が英語で書いた論文の抄録を五つ用いて実験を行った。全 47 文の中に 334 箇所の検討すべき名詞句があった。表 10, 11 に実験結果を示す。5.2.2.1 節の実験と同じく表 10 は得られた修正案がその正解と一致しているか、表 11 は冠詞誤り修正実験の結果を示す。

冠詞修正案の適切性評価の実験では、使用頻度の低い単語が多く使われているため、修正精度は New York Times[14]を用いたときに比べ低くなった。しかし、提案手法は検索結果数のみを用いた場合に比べ単語の使用頻度の影響は受けにくいので、使用頻度の低い単語が増えても精度の減少は小さい。

具体的な誤った修正事例を調べると[1]の場合や New York Times[14]と大体同じような例が多かった。

表 8 修正案の適切性評価(New York Times)

	正解	誤り	正解率
検索結果数のみの手法	335	63	0.84
提案手法	336	62	0.84

表 9 冠詞誤りの修正性能(New York Times)

	再現率	適合率	F 値
検索結果数のみの手法	0.84	0.75	0.79
提案手法	0.81	0.79	0.81

表 10 冠詞誤りの修正性能(論文)

	正解	誤り	正解率
検索結果数のみの手法	259	75	0.78
提案手法	269	65	0.81

表 11 冠詞誤りの修正性能(論文)

	再現率	適合率	F 値
検索結果数のみの手法	0.74	0.67	0.70
提案手法	0.80	0.74	0.77

5.2.3 考察と今後の課題

[1]は理系の論文を書く人向けの著書であるので、冠詞の用法についても、理系の論文によくある用法が多い。そのため、New York Times のような一般的な単語が使われる文章を校正する場合は、検索結果数を用いる手法と同程度だったと考えられる。今後の課題としては、用法の追加や用法の検出方法と検出した用法の利用方法の改善が挙げられる。十分に検索結果数が得られない場合は検索クエリの生成方法を工夫する必要もある。

6. まとめ

本稿では、岡山大学工学部情報工学科の専門英語の講義で使用しているテキスト[1]にある冠詞の用法を用いて冠詞を自動修正する方法を提案した。提案手法では、Web 資源を用いて入力された英文から予め定めた冠詞の用法を検出し、検出された用法と、検索エンジンによって得られた検索結果数を用いて冠詞の修正を行う。また、冠詞の用法検出に必要な名詞や形容詞のリストを検索エンジンを用いて作成した。評価実験では学術論文の抄録や一般的な文章と考えられる新聞記事を用いて、検索結果数のみを用いた修正方法と提案手法を比較した。実験結果より、学術論文などの文章では本手法は検索結果数のみを用いる方法を上回った。しかし、一般的な文章の場合は検索結果数のみを用いた手法とあまり精度に差はなかった。今後の課題として、冠詞の用法の拡充や、冠詞の用法と検索結果数を用いた冠詞の修正方法の改善が挙げられる。また、

冠詞の用法の検出のため作成した名詞および形容詞のリストについてもさらに精査したい。

参 考 文 献

- [1] 金谷健一 “これなら書ける論文英語―理系のための英文練習帳―” 岡山大学大学院自然科学研究科 2011 年 4 月(第 1.0 版)
- [2] 河合敦夫, 杉原厚吉, 杉江昇, “英文の誤りを検出するシステム ASPEC-I” 情報処理論文誌 Nov.1984 Vol.25 No.6, pp.1072-1079, 2007
- [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” 日本データベース学会, Letters vol.6, No.3, pp1-4, 2007
- [4] 綱嶋祐一, 岡田壮史, 安藤一秋, “検索エンジンを利用した多言語作文支援” 電子情報通信学会, 信学技報 ET2007-97, pp.73-78, 2008
- [5] 永田亮, 井口達也, 脇寺健太, 榊井文人, 河合敦夫, 井須尚紀, “前置詞情報を利用した冠詞誤り検出” 電子情報通信学会論文誌 D-I, Vol.J88-D-I No.4, pp.873-881, 2005
- [6] 乙武北斗, 荒木建治, “単語出現状況の特徴を用いた英文誤りの検出および自動校正” 社団法人情報処理学会, 研究報告, NL-171, pp.25-30, 2006
- [7] 永田亮, 若菜崇宏, 森広浩一郎, 榊井文人, 河合敦夫, 井須尚紀, “可算/不可算の判定に基づいた英文誤りの検出” 電子情報通信学会論文誌 Vol. J89-D, No.8, pp.1777-1790, 2006
- [8] 乙武北斗, 荒木建治, “単語出現状況の帰納的学習による英文誤りの検出及び自動校正” 電子情報通信学会論文誌, D Vol.J90-D No.6 pp.1592-1601, 2007
- [9] 竹内裕巳, 河合敦夫, 永田亮, 乙武北斗, “英文自動冠詞付与における前方照応の考慮” 情報処理学会研究報告, Vol.2011-NL-204 No.10, pp.1-7, 2011
- [10] Yahoo!JAPAN が提供する検索 WebAPI
<http://developer.yahoo.co.jp/webapi/search/>
- [11] Weblio 辞書 英和和英辞書 <http://ejje.weblio.jp/>
- [12] 欧羅巴人名録 <http://www.worldsys.org/europe/>
- [13] 米国国勢調査局による米国の人名頻度順リスト 1990 Census Name Files
<http://www.census.gov/genealogy/names/>
- [14] New York Times <http://www.nytimes.com/>