外国人の初級日本語単語の訂正方式

杉野 勝也† 絹川 博之†
†東京電機大学 大学院 工学研究科 情報メディア学専攻

1. はじめに

近年，コンピュータが教育分野で利用されるようになり外国人を対象とした日本語教育においても多く利用されるようになってきた．しかし，外国人日本語学習者が作成した文章を添削するシステムはほとんど見られず，日本語教師等の人手によって添削されているのが現状である．そのため，学習者が独学で文章作成を学習することは困難である．

そこで我々は外国人学習者が独学で文章作成を学習できることを目標として日本語教育支援システムを開発している．現段階では，対象を初級日本語にしぼり，学習者の作成した文の誤りのうちひらがな表記誤りを検出，訂正する方法を研究している．なお，ひらがな表記誤りとは日本語の読みのひらがな表記の誤りをいい，例えば「がっこう（学校）」を「がこう」，「たべる（食べる）」を「たへる」などである．

2. 外国人の学習する初級日本語

2.1 初級日本語

本研究では外国人のための初級日本語を研究対象にしているが，ここでの初級日本語とは，財団法人日本国際教育支援協会と独立行政法人国際交流基金が行っている日本語能力試験の 3 級レベルに相当しており，日常生活に役立つ会話ができ、簡単な文章が読み書きできる程度の日本語である．漢字は 300 字程度，語彙は 1,500 語程度が必要とされている．

2.2 外国人の初級日本語

本研究では，実際に日本語を学習している外国人が作成した日本語を収集、解析して研究している[1]．学習者には非漢字圏の学習者もおり，また，初級レベルにおいては正しいひらがな表記を身に付けるために漢字を使わずにひらがな表記することがある．そのため，初級日本語学習者の日本語ではひらがな表記が多くなり，その中には誤りも多く見られる．

3. 外国人の初級日本語の誤り

3.1 誤りの分類

収集した外国人の日本語を解析し，どのような誤りがあるかを調べた．以下に代表的な誤りを示す．

- ひらがな表記誤り
- 活用の間違い
- 指示詞の間違い

これらの中で，ひらがな表記誤りが 71.0%を占め一番多かったので，我々はまず，ひらがな表記誤りの検出，訂正方式を研究することにした．ひらがな表記誤り検出方式に関しては別途報告している[1]ので，本稿では訂正方式について述べる．

3.2 ひらがな表記誤りによる分類

ひらがな表記誤りは 48 種類に分類した．ひらがな表記誤りの件数は 179 件あったが，それらの中で発音関係に分類される誤りが 79.9%あった．そこで，発音関係誤りの訂正方法を重点的に研究することにした．以下に発音関係誤りの一部を示す．

- 非清音を清音にしている
- 非濁音を濁音にしている
- 長音（う）の間違い
- 拗音を別の拗音にしている
- 促音が抜けている
- 濁音を半濁音にしている
- 濁音を別の濁音にしている

ここに示した 7 種類の誤りで発音関係誤り全体の 81.8%を占め，ひらがな表記誤り全体では 65.3%を占める．

4. ひらがな表記誤りの訂正方法

4.1 ローマ字表記への変換

発音に関係するひらがな表記誤りを扱う場合，母音，子音がはっきり分かれているローマ字表記が都合がよい．そこで，外国人がひらがな入力した単語をローマ字に変換し，誤りを訂正することにした．その後，再びひらがなに直し，正しい単語を提示する．ローマ字表記の場合，「し」，「ち」，「つ」，「ふ」を SHI，CHI，TSU，FU とするヘボン式ローマ字を使うことが多いが，処理を簡単にするため，SI，TI，TU，HU で表記する訓令式ローマ字を用いることにした．

4.2 正しい単語候補の作成

前記の通り，ひらがな表記誤りには 48 種類あるが，訂正方式はそれぞれ異なる．しがたがって訂正方式も 48 通りある．単語のひらがな表記誤りがどの分類に属し，どの訂正方式を用いればいいかをコンピュータに判断させるのは困難である．本研究では，一つの単語に対して，48 通りの訂正を行い，正しい単語の候補を複数作成する．その後，複数できた単語候補を絞り込むことにした．以下に 7 種のひらがな表記誤りの訂正方式を示す．

4.2.1 「非清音を清音にしている」の訂正

「非清音を清音にしている」というのは，濁音および半濁音を含む単語「ゆうびんきょく（郵便局）」を「ゆうひんきょく」に，「パン」を「ハン」にしている誤りのことである．

ローマ字表記での濁音は G，Z，D，B，半濁音は P である．濁音に対応する清音はそれぞれ K，S，T，H．半濁音に対応する清音は H である．単語中の K，S，T，H を G，Z，D，B，P に変えれば，正しい単語の候補を生成するこ

Error Correction in Foreigner's Japanese Words
†Katsuya Sugino, Hiroshi Kinukawa
Tokyo Denki University, Graduate School of Engineering

とができる．なお，HはBとPに変えて2つの単語候補を生成する．

4.2.2 「非濁音を濁音にしている」の訂正

「非清音を濁音にしている」というのは，「おとこ（男）」を「おどこ」に，「デパート」を「デバート」にしていることである．

この場合，4.2.1と反対の処理を行う．単語の中のG，Z，D，Bをそれぞれ K，S，T，H，Pに変えて単語候補を作る．なお，BはHとPに変換し2つの単語候補を作る．

4.2.3 「長音（う）の間違い」の訂正

「長音（う）の間違い」には「う」の欠如の訂正と，余計な「う」の訂正の2種類ある．

(1) 「う」の欠如の訂正

「う」の欠如とは，「う段＋う」の単語「くうき（空気）」や「お段＋う」の単語「ひこうき（飛行機）」の「う」が欠如して，「くき」「ひこき」にしている誤りである．

「う段」は*U，「お段」は*Oであるので，U，Oを検索し，その後ろにUを追加する．ただし，正しく表記している箇所に余計なUを追加しないために*UU，*OUとなっているところにはUを追加しない．

なお，今回のデータ収集では得られなかったが，他の長音の間違いをすることもある．日本語の長音は上記以外にも「あ段＋あ」，「い段＋い」，「え段＋い」，「え段＋え」，「お段＋お」があり，これらの誤りは同様な方式で訂正できる．

(2) 余計な「う」の訂正

余計な「う」とは，「きょねん（去年）」に余計な「う」を追加して「きょうねん」にする誤りである．

この場合，「きょねん（去年）」を「きょうねん」にしている等の余計な「う」に関しては，*UU，*OUを探してUを削除して単語候補を作る．

4.2.4 「拗音を別の拗音にしている」の訂正

「拗音を別の拗音にしている」とは「しゃちょう（社長）」を「しゃちゅう」にしていることである．

拗音は「子音＋y＋母音」であるので，「子音＋y＋母音」を検索し，母音の別の母音に変換する．ただし，拗音の母音はA，U，Oのみ．

4.2.5 「促音が抜けている」の訂正

「促音が抜けている」とは「ざっし（雑誌）」を「ざし」，「きって（切手）」を「きて」にしていることである．

促音は同じ子音が2つ続いているので，子音の前に同じ子音をつけて，単語候補を作成する．ただし，単語の最初に子音および，すでに同じ子音が続いている場合は，子音の追加は行わない．

4.2.6 「濁音を半濁音にしている」の訂正

「濁音を半濁音にしている」とは「しんぶん（新聞）」を「しんぷん」，「コンビニ」を「コンピニ」にしている誤りである．

半濁音はP，それに対応した濁音はBであるので，PをBにし単語候補を作成する．

4.2.7 「濁音を別の濁音にしている」の訂正

「濁音を別の濁音にしている」とは「かぜ(風邪)」を「かぞ」，「ざっし（雑誌）」を「ぎっし」にしている誤りである．

「かぜ」と「かぞ」の誤りは，「ぜ（ZE）」と「ぞ（ZO）」で，母音の間違いであるので，他の母音の変換をすれば単語候補を作成できる．

一方，「ざっし」と「ぎっし」の場合は，「ざ（ZA）」と「ぎ（GI）」の間違いで，子音，母音共に変える必要がある．濁音は全部で20文字あるので，一つの濁音を他の濁音19個に変化させて19個の単語候補を生成する．

なお，今回のデータ収集では見られなかったが，子音のみを間違っている場合も考えられる．この場合，子音のみを変更すれば単語候補を作ることができる．

4.3 正しい単語候補の絞込み

複数できた単語候補を絞込みする必要がある．ここでは「初級単語辞書」[1]を使用する．「初級単語辞書」とは初級日本語で扱う単語を収集したもので，現在作成中である．複数ある単語候補が「初級単語辞書」に登録されているかを調べ，登録されていない場合は，誤りと判断して候補から外し，登録されているもののみ候補として残す．このような絞込みを行っても複数の単語候補が残る場合があり，さらなる絞込みが必要となるがこれに関しては今後の課題とする．

5. おわりに

今回，提案した訂正方式で発音関係誤りの81.8%に対応することができたが，ひらがな表記誤りの全体からみると65.3%にすぎない．今後，発音に関係しないひらがな表記誤りに対しても訂正方式を検討していきたい．

また，今回の訂正方式を利用し，外国人向け初級日本語独習システムの開発する予定である．これはコンピュータが学習者に問題を出し，学習者は文で解答するシステムで，学習者の文を解析し，誤りを指摘し，正しい文を提示するものである．

謝辞

本研究を行うにあたり，学校法人 吉岡学園 千駄ヶ谷日本語学校に御協力を頂きました．この場を借りて御礼を申し上げます．

参考文献

[1]杉野勝也,絹川博之, “外国人の初級日本語文の誤り検出方式”, 第7回情報科学技術フォーラム(FIT2008)第3分冊 pp563-564 (2008).
[2]益岡隆志,田窪行則, “基礎日本語文法 –改訂版-”, くろしお出版 (1992).
[3]益岡隆志, “24週日本語文法ツアー”, くろしお出版 (1993).
[4]吉川 武時, “日本語文法入門”, アルク(1989).
[5]千駄ヶ谷日本語教育研究所著, “コミュニケーション日本語1, 千駄ヶ谷日本語研究所 (1999).
[6]千駄ヶ谷日本語教育研究所著, “コミュニケーション日本語2, 千駄ヶ谷日本語研究所 (1999).
[7]千駄ヶ谷日本語教育研究所著, “コミュニケーション日本語3, 千駄ヶ谷日本語研究所 (1999).
[8]スリーエーネットワーク編著, “みんなの日本語 初級Ⅰ 本冊”,スリーエーネットワーク(1998.)
[9]スリーエーネットワーク編著, “みんなの日本語 初級Ⅱ 本冊”,スリーエーネットワーク(1998.)