

外国人の初級日本語学習時における単語の仮名表記誤りの訂正方式 Error Correction in Kana Expressed Word for Foreigner's Basic Japanese Language Learning

谷之口 優人†
Yuto Taninokuchi

杉野 勝也†
Katsuya Sugino

絹川 博之†
Hiroshi Kinukawa

1. はじめに

初級日本語を学習している外国人は、仮名表記において初歩的な誤りをする事が多い。特に、濁音・半濁音の付け間違い、促音の抜け、長音などの誤りが多く見られる。例えば、「がっこう(学校)」を「がごう」、「くうき(空気)」を「くき」のような誤りである。このような誤りを含む日本語文を解析する方式はあまり発表されておらず、これらの検出、訂正は日本語教師などの人手に頼っ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。

初級日本語学習者は、単語、漢字の読み書き、文法、発音、文章作成、聞き取りなど、様々な分野の学習をする必要がある。その中でも文章作成は、添削が必要である問題などから、コンピュータで支援をおこなう必要があると考えられる。

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日本語学習支援システムの開発を目指している。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振り仮名の誤りを検出、訂正する方法を研究している。

本稿では、訂正候補を生成する際の促音、長音などの追加・削除に関する方式や、無駄な訂正候補を減らす方式について報告する。

2. 外国人向けの初級日本語学習支援システム

2.1 対象とする日本語

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 N3 レベルに相当しており、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 文章作成支援

本システムは学習者が文章を入力すると、システムが誤り検出、訂正を行い、学習者に誤りの指摘と正解を提示する。なお入力される文章は、アラビア数字や漢数字などは使わず、全て平仮名表記したものを想定している。

3. 外国人の初級日本語の誤り

表 1 に外国人が誤りやすい仮名表記誤りの分類と、その例を示す。このような誤りの特徴を利用し、誤りの訂正候補を自動的に生成する。

表 1. 仮名表記誤りの分類

(1) 非清音を清音にしている
「ゆうびんきょく(郵便局)」を「ゆうひんきょく」に、「パン」を「ハン」
(2) 非濁音を濁音にしている
「がっこう(学校)」を「がごう」
(3) 拗音を別の拗音にしている
「しゃちょう(社長)」を「しゃちゅう」
(4) 濁音を半濁音にしている
「しんぶん(新聞)」を「しんぷん」
(5) 濁音を別の濁音にしている
「かぜ(風邪)」を「かぞ」、「ざっし(雑誌)」を「ぎっし」
(6) 長音の誤り
「くうき(空気)」や「ひこうき(飛行機)」の「う」が欠如して、「くき」「ひこき」にしているもの、「きょねん(去年)」に余計な「う」を追加して「きょうねん」
(7) 促音が抜けている
「ざっし(雑誌)」を「ざし」

4. 仮名表記誤りの訂正方式

4.1 全ての候補を生成する訂正方式と問題点

訂正方式として、濁音・半濁音、長音、促音などの誤りの特徴を利用し、誤った単語の文字列から訂正候補を生成する。濁音に関する訂正であれば、清音を濁音に、濁音を清音に変換する処理を各平仮名にし、全ての組み合わせを生成する。

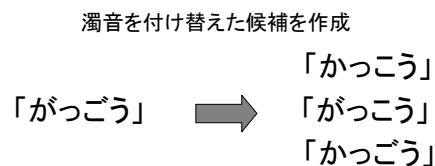


図 1. 濁音に関する候補生成

図 1 の例では、「がっこう(学校)」を「がごう」と誤った場合、濁音に関してだけでも「かっこう」「がっこう」「かごう」という候補を生成することになる。文字数が多いほど組み合わせが多くなるため、無駄な候補が多く生成されてしまう。更に濁音だけでなく、長音や促音、半濁音などの候補の組み合わせを考えると、大量の候補が出来てしまうことが問題であった。

†東京電機大学大学院 未来科学研究科

この問題に対処するため、候補を生成する際、清音のみを扱うことで、濁音・半濁音を考慮しない方式を提案する。更に長音や促音についても無駄な候補を生成しないように改良がされている。

4.2 清音化による訂正方式

上記の問題を解決するために、濁音・半濁音に関しての候補を生成しない、清音化による訂正方式を提案する。

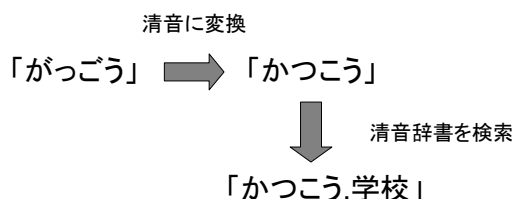


図 2. 清音化による訂正方式

「がっこう(学校)」の誤りの例では(図 2 参照)、濁音を付け替えた候補を生成せずに、清音に変換する。促音・拗音は通常の仮名に変換する(「っ」は「つ」に、「ゃゅよ」は「やゆよ」に)。そして清音に変換された文字列から、訂正候補を見つけ出すための辞書として、見出しの単語を清音にした“清音辞書”を使用する。「がっこう(学校)」という単語は、辞書の見出しが「かつこう」と登録されているので、濁音の候補を生成することなく、訂正後の単語である「がっこう(学校)」を導き出すことができる。更に辞書の見出しにおいて、促音・拗音を通常の仮名に変換したものを使用することで、促音・拗音を考慮せずに訂正することができる。

4.2.1 清音辞書

清音辞書には、初級日本語で使われる単語、約 2000 語が登録されており、検索用の見出しとして濁音・半濁音を清音に、促音・拗音を通常の仮名に変換したものをを用いる。その見出しに対して、漢字表記、仮名表記の情報を持っている。例えば、「学校」であれば、見出しが「かつこう」、漢字表記、仮名表記は「学校, がっこう」のように登録されている。

4.2.2 誤り訂正の流れ

訂正は、次のような流れでおこなわれる。

- (1) 誤った単語を清音に変換する。
- (2) 更に、促音・拗音を通常の仮名に変換する。
- (3) 長音、促音を追加した候補を生成する(4.3.2, 4.3.3 で後述)。
- (4) 清音辞書から候補と一致する見出しを探し、その見出しに登録されている漢字表記を訂正後の単語とする。

4.3 候補生成の方法

4.3.1 平仮名に対する付加情報

各平仮名に対して、表 2 の情報が付加されている。これらの情報を利用することにより、長音の追加、促音の追加、濁音変化などの処理をおこなう。

表 2. 平仮名に対する付加情報

(a) 五十音表の何行, 何段か
(b) その平仮名が, 清音・濁音・半濁音のどれか
(c) 濁音, 促音, 拗音の変化の有無
(d) 後ろに長音が付くかどうか
(e) 前に促音が付くかどうか

4.3.2 長音の追加

長音の追加には、表 2 の(d)の情報を用いる。例えば、平仮名「が」に対して、長音が付くことは無く、「があ」という文字の並びが現れる単語は存在しない。更に「そ」に対して長音が付き、「そお」となる場合はあるが、「ぞ」に対しては「お」が付くことは無い。これらの情報をまとめ、利用することで無駄な長音を付けた候補を減らすことができる。

4.3.3 促音の追加

促音の追加には、表 2 の(e)の情報を用いる。例えば母音の前に促音が付くことはなく、「〜っあ」という単語は存在しない。初級日本語で使われる単語、辞書で調べた結果、83 字中、22 字のみが、その平仮名の前に促音が付く可能性があることがわかった。これらの情報を用いて、促音を追加するかどうか判断することで、無駄な候補を減らすことができる。

5. 考察

本手法を用いることで、訂正する際の候補を減らすことが出来るが、清音に変換した辞書によって検索をおこなうため、いくつかの単語が最終的な訂正候補が出力される可能性がある。例えば、「しけん」という候補で清音辞書の検索をおこなった場合、「しけん(試験)」「じけん(事件)」という訂正候補を出力する。よって候補に対して優先度を付け、可能性が高いものから表示させる必要がある。そのためには、初級日本語においてよく使われる単語であるか、誤りやすい間違いであるか、などの情報を利用して優先度の重み付けをおこなう事が考えられる。

6. おわりに

本稿では、濁音・半濁音を清音化することや、長音や促音についての情報を利用することで、生成される候補を減らし、効率的に訂正する方式を提案した。現在、提案した方式に基づいてプログラムを開発している。開発したプログラムを用いて実験評価し、さらにプログラムを改良していく予定である。

参考文献

- [1] 杉野勝也, 佐藤俊也, 絹川博之: 外国人の初級日本語学習における仮名表記と文法の初歩的誤りの検出方式, 第9回情報科学技術フォーラム(FIT2010)第3分冊(2010).
- [2] 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).