

K-027

外国人の初級日本語文における振り仮名の誤り訂正 Error Correction of Giving Kana in Foreigner's Japanese Language Learning

佐藤 俊也[†]
Toshinari Sato

杉野 勝也[†]
Katsuya Sugino

絹川 博之[†]
Hiroshi Kinukawa

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でも多く利用されるようになったが、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学習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振り仮名の誤りを検出、訂正する方法を研究している。なお、ここでの振り仮名誤りとは、日本語の読みの誤りをいい、例えば「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][2]。学習者には非漢字圏の学習者もあり、また、初級レベルにおいては正しい振り仮名を身に付けるために漢字を使わずに振り仮名だけで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振り仮名が多くなり、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。

- ・振り仮名の誤り
- ・活用間違い
- ・指示詞の間違い

これらの中で、振り仮名の誤りが71.0%を占め一番多かったため、振り仮名の誤りの検出、訂正方式を研究する

ことにした。この内、振り仮名の検出方式については別途報告するので、本稿では振り仮名の訂正方式を述べる。

4. 誤りの訂正方法

4.1 誤りの種類とその訂正

4.1.1 「非清音を清音にしている」の訂正

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」にしている誤りで、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, 半濁音は、Hである。よって単語中のK, S, T, HをG, Z, D, B, Pに変え単語候補を生成する。

4.1.2 「非濁音を濁音にしている」の訂正

「おとこ（男）」を「おどこ」にしている誤りで、この場合は、4.1.1と反対の処理をする。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に変換して単語候補を作成する。

4.1.3 「拗音を別の拗音にしている」の訂正

「しゃちよう（社長）」を「しゃちゅう」にしている誤りで、拗音の母音を変化させ単語候補を生成させる。

4.1.4 「濁音を半濁音にしている」の訂正

「しんぶん（新聞）」を「しんぷん」にしている誤りで、PをBに変え単語候補を生成する。

4.1.5 「濁音を別の濁音にしている」の訂正

「かぜ（風邪）」を「かぞ」、「ざっし（雑誌）」を「ざっし」にしている誤りで、母音のだけの誤りと母音と子音の誤りの2通りであり、濁音を他の19の濁音に変化させて単語候補を生成する。

4.1.6 「長音（う）の間違い」の訂正

(1) 「う」の欠如の訂正

「う、お段+う」の単語「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」にしている誤りで、「う段」は、*U, 「お段」は、*Oであるので、U, Oを検索し、その後ろに「う」を挿入する。

なお、*U, *Oの後に「う」が、すでに入っている場合には、「う」は、挿入しない。

(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。

この場合には、余計な「う」を*Uと*Oの後ろに発見して削除する。

4.1.7 「促音が抜けている」の訂正

「ざっし（雑誌）」を「ざし」にしている誤りで、この場合は、先頭の文字の後から文字と文字の間に「っ」を挿入して単語の候補を作成する。

[†] 東京電機大学 大学院 情報メディア学専攻
Tokyo Denki University, Graduate School

4.2 誤り訂正の実現方法

4.2.1 ローマ字表記への変換

プログラムにおいて誤り訂正方式実現のため振り仮名の誤り単語をローマ字へ変換し処理することにした。

また、その変換時には、訓令式を用いた。その他の工夫として、あ行は、通常一文字だが F を付けて処理しやすくし、小さいや行は、VA, VU, VO と表し、「ん」は、LL とし、全ての文字を2文字表記として扱うこととした。

4.2.2 文字を誤って書いた場合の訂正方法

4.1.1~4.1.5 の誤りに関する訂正をする際の方法としては、あらかじめ変換を実現するためのテーブル(表1参照)を用意しておく、このテーブルは、行列に作られた表形式の配列に変換に関する文字データを配置する。その後、探索文字は、第1列に入力し、その文字に対応した変換文字をその後に配置して作成されたものである。この第1列に格納された文字が、単語中に出現するかを調べてある場合には、文字を変換テーブルにしたがって変換し候補とし単語データを保存していく。

行数:68					
対象文字	変換数	変換後の文字			種類
KA	1	GA			4.1.1
HA	2	BA	PA		4.1.1
...	...	...			...
ZA	1	SA			4.1.2
BA	2	HP	PA		4.1.2
...	...	...			...
VA	2	VU	VO		4.1.3
VU	2	VA	VO		4.1.3
...	...	...			...
PA	...	BA			4.1.4
PI	...	BI			4.1.4
...	...	...	...	...	...
ZO	19	ZE	GA	...	4.1.5
GI	19	ZA	DE	...	4.1.5

図1 変換テーブル形式

4.2.3 「長音(う)の間違い」の訂正方法

(1) 「う」の欠如の訂正方法

単語中の「う段」*U と「お段」*O を検索し、その後ろに FU を追加し、単語候補を作成する。ただし、正しく表記している箇所余計に FU を追加しないために*UFU、*OFU となっているところには FU を追加しない。

(2) 余計な「う」の訂正方法

単語中の*UFU、*OFU を検索しあった場合には、FU を削除し単語候補を作成する。

4.2.4 「促音が抜けている」の訂正方法

単語中に出てくる X を検索し、その後続く母音の部分をも他の母音に変換して単語候補を作成する。

4.3 正しい単語候補の絞り込み

複数できた単語候補を絞り込む必要がある。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だけをすべて収集登録したものである。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。このような絞り込みを行っても複数の単語候補が残る場合があり、さらなる絞り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

本実験成果として、入力データとして与えられた誤り単語をひらがなからローマ字に変換するプログラムを作成した。また、文字を変換と長音の「う」の欠如の訂正と長音の余計な「う」の削除と促音が抜けている場合の訂正を実現するプログラムを作成した。

今回の研究で振り仮名誤りにおける7つの訂正方法は、非常に少ない数であるが、理論的に振り仮名誤りの65%を訂正できる見通しとなっており、非常に有意義な研究が、出来たと考えられる。

今後の課題としては、振り仮名表記誤りを100%訂正できるようにする。また、それだけでな、活用や指示詞の誤りについても訂正できるように研究を進める。また、正しい単語候補の絞り込みの所で少し触れたが、今回の研究においては、訂正の方法として単語候補を作成してその中から正しい単語候補を絞り込む形式を取っているが、この単語候補の絞り込みにおいて最終的にひとつに絞り込むことは、非常に困難であると考えられるので絞り込みに関しては、今後の大きな課題である。

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] 杉野勝也, 絹川博之, “外国人の初級日本語文における振り仮名の誤り検出”, 情報処理学会 第71回全国大会分冊 4 pp.615-616(2009).
- [3] 杉野勝也, 佐藤俊也, 絹川博之, 外国人の初級日本語文における振り仮名の誤り検出”, 第8回情報科学技術フォーラム(FIT2009)第3分冊(2009).
- [4] 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).
- [5] 益岡隆志, “24週日本語文法ツアー”, くろしお出版(1993).
- [6] 吉川 武時, “日本語文法入門”, アルク(1989).
- [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 1, 千駄ヶ谷日本語研究所(1999).
- [8] 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).
- [9] 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).
- [10] 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).
- [11] 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).