

K-026

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

杉野 勝也[†] 佐藤 俊也[†] 絹川 博之[†]
Katsuya Sugino Toshinari Sato 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、ここでの振り仮名とは日本語の読みの誤りをいい、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
- ・活用の間違い
- ・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ので、我々はまず、振り仮名の誤りの検出、訂正方法を研究することにした。この内、振り仮名の訂正方式につ

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

いては別途報告する[3]ので、本稿では検出方式について述べる。

4. 誤りの検出方法

4.1 形態素解析による検出

以下に誤りの検出方法の概略を示す。

- (1) 対象となる文を形態素解析プログラム JUMAN[4]にて形態素解析する。
- (2) 形態素解析の結果を初級構文辞書と比較し、初級日本語文型と一致するかを調べる。
※「初級構文辞書」とは初級日本語で扱う文型の形態素解析結果を登録したものである。
- (3) 形態素解析結果と初級構文辞書が一致しない場合は、文に誤りがあると判断する。
- (4) 形態素解析結果と初級構文辞書が一致する時、対象文が正しい場合と誤りを含む場合があるので、各単語について初級単語辞書登録の有無により正誤判断する。(4.2参照)

※「初級単語辞書」とは初級日本語で扱う単語だけを全て登録したものである。

(4)の初級構文辞書と一致したにも関わらず、誤りがある例としては、振り仮名が初級単語辞書非登録単語を表す場合である。

正文： わたしは がくせい です

非文： わたしは かくせい です

形態素解析の結果

がくせい：名詞（「学生」と判断）

かくせい：名詞（「覚醒」と判断）

上記の2文は、形態素解析では正文も非文も同じ結果となり非文に誤りがあることは判断できないので、初級単語辞書により誤りを検出する。

4.2 単語の正誤確認

単語が初級単語辞書に登録されているか否かによって単語の正誤を判断する。登録されている場合は、単語は正しいと判断し、登録されていない場合は誤りと判断する。誤りと判断した場合は、誤りの解析を行う。

5. 検出方法の詳細

5.1 形態素解析の前処理

JUMANによる振り仮名文の形態素解析では、一文字でも誤りが存在すると、単語の境界を誤り、誤解析が多くなることがある。そのため、そのままでは、形態素解析結果を利用できない。そこで、形態素解析を行う前に可能なものは誤り検出、訂正を行う。この前処理を行うことにより形態素解析の精度を高めることができる。以下に前処理について述べる。

5.1.1 文末処理

初級日本語学習者でも文末表現を間違っていることは少ない。また、文末表現に誤りがあっても推測容易なものが多く、正しい文末表現に訂正できる。そこで、文末を調べて、場合によっては訂正して、文末に接続されている単語の正誤を判断する。判断方法は以下の通り。

- ・初級単語辞書に登録されているか、
- ・品詞の正誤(文末に接続可能な品詞であるか)
- ・活用の正誤(正しい活用をしているか)

単語の活用が誤りと判断した場合、正しい活用に訂正する。

お酒を飲むすぎました

↓

お酒を飲みすぎました

初級単語辞書非登録単語または誤った品詞の場合は、初級単語辞書の中の単語に置き換えるが、意味は考慮せずに、文型のみを考慮して、文末との接続が正しくなる単語を選択する。したがって、文型は正しいが、意味が正しくない文になる。

京都へいかんつもりです

↓

京都へ歌うつもりです

(文型は正しいが意味が正しくない)

ここでは JUMAN の形態素解析が正しくなることを目的としているので、意味は考慮しない。この後の形態素解析結果と初級構文辞書との比較で一致すれば、置換単語が誤りであると判断する。

5.1.2 助詞間単語処理

初級日本語では、各名詞文節に必ず助詞がついていることが多い。

弟の学校に英語の先生がいます(助詞間は1単語)

以下の例文のように自立語のみの文節を含む場合もある。

私はきのう買った本を読む(助詞間に3単語ある)

そこで助詞の前に自立語があると仮定し、それらが基本単語辞書に登録されているかを調べる。登録されていない場合は、訂正を試みる。(単語の訂正の仕方については別途報告する[3]) 訂正できた場合は、訂正した単語に置き換える。訂正できなかった場合は、置き換えは行わない。

ここでは助詞「か」、「が」、「で」を対象外とする。これらが助詞であるか、単語の一部であるかの判断が困難であるためである。

5.2 形態素解析の後処理

5.2.1 付加情報の削除

形態素解析後、形態素解析結果と初級構文辞書との比較を行うが、単純比較では、正常な結果が得られない場合が多い。以下にその例を示す。

文1: 私は本を読んだ。

文2: きのう、私は本を読んだ。

文3: 私はゆっくり本を読んだ。

文4: 私は本を読んだよ。

上記の文は、意味はほとんど同じであるが、構文は全て異なる。単純比較をする場合、これらの構文を全て初級構文辞書に登録する必要があり、登録数が非常に多くなり、実用的とは言えない。文2から文4の下線部は、一種の付加情報で、無くても文の意味が通る単語である。そこで、

形態素解析結果より、付加情報を削除した後、初級構文辞書との比較をすることにした。名詞、動詞、形容詞、形容動詞、助詞と文末以外の品詞を削除する。名詞のうち、時を表すもの(例「きのう」)も削除する。

また、削除する付加情報より取得可能な情報は利用する。例えば、「きのう」の場合、「過去を表す」という情報が得られるので、文が過去形であるか否かを確認する。「あまり」という副詞では文が否定文であるかを確認する。

6. おわりに

外国人学習者が作成した日本語の誤りの検出方式を検討したが、今後は、本検出方式と別途報告[3]の訂正方式を組み込んだシステムを開発し、評価、改良をしていく予定である。

今回の研究では、構文比較による誤り検出以外に、「きのう」など時を表す単語と文の時制が一致しているか否か、「あまり」など副詞が肯定文または否定文で正しく使われているかという確認も行った。

しかし、初級日本語に限定し、比較的簡単に短い文を想定しているため、すこし複雑な文になると対応できない。例えば、「今日、きのう買った本を読む」という文の場合、「過去」を示す「きのう」と「現在・未来」を表す「きょう」が存在するため、文の時制を判断するのが困難である。また、意味を考慮していないので、「本を飲んだ」という文を誤りと判断できない。これらの問題点に関しては今後対処していきたい。

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単語の訂正方式”, 情報処理学会 第71回全国大会分冊 4 pp615-616 (2009).
- [3] 佐藤俊也, 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訂正”, 第8回情報科学技術フォーラム(FIT2009)第3分冊 (2009)
- [4] 黒橋禎夫, 河原大輔, “日本語形態素解析システム JUMAN Version5.1”, 東京大学大学院情報理工学系研究科 (2005).
- [5] 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).
- [6] 益岡隆志, “24週日本語文法ツアー”, くろしお出版 (1993).
- [7] 吉川 武時, “日本語文法入門”, アルク (1989).
- [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1, 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2, 千駄ヶ谷日本語研究所 (1999).
- [10] 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).
- [11] 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).
- [12] 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).