

外国人の初級日本語文の誤り検出方式 Error Detection in Foreigner's Japanese Language Learning

杉野 勝也[†]
Katsuya Sugino

絹川 博之[†]
Hiroschi 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。多者択一の問題を行うシステム、映像、音声を利用したシステム、リーディングチュウ太[1]などの読解支援システムが開発され活用されているが、学習者が作成した誤りを含む日本語文を分析するシステムはほとんど見られない。現在の外国人向け日本語教育において、学習者の文を添削するのは教師で、機械化はされていない。したがって、学習者が独学で文章作成の学習は困難である。

日本語構文解析の分野では茶釜[2]や JUMAN[3]が多く利用されているが、これらは正しい日本語を対象にしており、日本語学習者の間違った日本語には対応していない。

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システムの開発を行っている。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表記ミスを検出する方法を研究している。表記ミスとは日本語の読み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. 研究対象

2.1 研究対象の日本語

本研究では初級日本語を研究対象にしているが、ここの初級日本語とは、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 研究対象者およびデータ

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。対象者は大学または大学院、専門学校への進学を目的として日本へ留学し、日本語学校で勉強している学習者である。

データとなる学習者作成の日本語文は、日本語学校で行われた試験の解答より採取した。主に文章での解答を採取したが、単語テスト、ディクテーションの解答からも採取した。採取した数を表1に示す。

表1 採取したデータ数

文章での解答	179 文
単語テストの解答	125 語
ディクテーションの解答	29 文

[†] 東京電機大学 大学院 工学研究科 情報メディア学専攻
Tokyo Denki University, Graduate School of Advanced
Science and Technology

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になっている。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、その分、誤りも多くなっている。

3. 誤りの分類

3.1 品詞による分類

採取したデータ(誤りがある日本語)を表2に示すように品詞ごとに分類した。「勉強する」というのは、「勉強をする」と表現することもあるが、「勉強する」の場合、一つの動詞として分類し、「勉強をする」は、「勉強」を名詞、「を」を助詞、「する」を動詞に分類した。また、文末でよく使われる表現で「～じゃない」、「～なければならぬ」等は複数の品詞で構成されているが、日本語教育において一つの文末として扱うので本研究においては「文末詞」という品詞として分類した。

表2 分類で用いた品詞とその例

動詞	食べる, 見る, 勉強する
名詞	学校, テレビ, 勉強, 買い物
形容詞 (い形容詞)	大きい, 多い, 暑い, 古い
形容動詞 (な形容詞)	有名, きれい, にぎやか, 親切
助詞	は, も, が, を, へ, に, の
副詞	とても, あまり, ぜひ, もし
数詞	1本, 2人, 3時, 4つ, 5個
文末詞	なければならない, じゃない
接続詞	そして, から, でも, とき

誤りを品詞別に分類した後、品詞毎にさらに細かく分類した。分類結果の一部を表3に示す。表3からわかるように品詞においても「表記ミス」が多い。そこで我々は初めに表記ミスの検出方法を研究することにした。

表3 品詞による分類結果の一部

動詞	表記ミス	19 件
	活用の間違い	7 件
名詞	表記ミス	87 件
	指示詞の間違い	7 件
形容詞	表記ミス	15 件
	「～て」の間違い	3 件

3.2 表記ミスによる分類

表記ミスは48種類に分類した。分類結果の一部を表4に示す。分類した結果、表4に示した誤りが多いことが判明したので、これらの検出方法を重点的に研究することにした。

表4 表記ミスによる分類結果の一部

非清音を清音にしている	31件
非濁音を濁音にしている	25件
長音(う)の間違い	21件
拗音を別の拗音にしている	15件

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を示す。

- (1) 対象となる文を形態素解析プログラム JUMAN にて形態素解析する。
- (2) 形態素解析の結果を文型辞書と比較し、初級日本語文型と一致するかを調べる。
※「文型辞書」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) 形態素解析結果と文型辞書が一致しない場合は、文に誤りがあると判断する。
- (4) 形態素解析結果と文型辞書が一致する時、対象文が正しい場合と誤りを含む場合があるので、各単語を正誤判断する。(4.2)

(4)の文型辞書と一致したにも関わらず、誤りがある例を以下に示す。

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2文は、正文も非文も同じ形態素解析結果となり、形態素解析だけでは非文に誤りがあることは判断できない。

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。

なお、「初級単語辞書」とは初級日本語で扱う単語のみを登録した辞書である。

4.3 誤りの解析

誤りの種類が多いため解析方法が多くなっている。本論文では、表4で示した誤りの解析方法を述べる。

4.3.1 「非清音を清音にしている」の解析

非清音を含む単語「ゆうびんきょく（郵便局）」が「ゆうひんきょく」に、「パン」が「ハン」にすることがある。この場合、濁音、半濁音になる可能性がある文字を濁音、半濁音にして、変更後の単語が初級単語辞書に登録されているか否かを調べる。登録されている場合は、それが正しい単語と判断する。

4.3.2 「非濁音を濁音にしている」の解析

「おとこ（男）」を「おどこ」に、「デパート」を「デバート」にする等、非濁音を濁音にすることがある。この

ような誤りに対しては、濁音になっている文字を清音または半濁音にして、初級単語辞書に登録されているかを調べる。

4.3.3 長音(う)の間違い

「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」にすることがある。この場合、単語の「う段」または「お段」の文字の後ろに「う」を加えて、初級単語辞書と比較する。

4.3.4 「拗音を別の拗音にしている」の解析

「しゃちょう（社長）」を「しゃちゅう」にする等、拗音を別の拗音にすることがある。この場合、単語の中の拗音を別の拗音にして初級単語辞書と比べる。拗音の誤りは「ゃ、ゅ、ょ」の間違いだけでなく、「ちよ」を「しよ」になる等、拗音全体を別の拗音に変えて正誤判断する必要がある。

5. 今後

外国人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、この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。また、外国人学習者が作成した文章の収集も続け、多くの誤りを検出できるようにしていく。

今回の研究では表記ミスに限定して解析を行ったが、これでは、「パンで食べた」「ごはんを飲んだ」等の誤りは検出できない。これらの誤りを検出するには動詞と助詞の組み合わせ、単語の意味等を配慮する必要があるが、今後対処していきたい。

また、今回の検出方式を利用し、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] <http://language.tiu.ac.jp/>
- [2] 松本裕治他, “形態素解析システム『茶釜』version 2.3.3 使用説明書”, 奈良先端科学技術大学院大学 情報科学研究科 (2003).
- [3] 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).
- [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).