

外国人の書いた日本語の校正支援システムの研究*

5R-1

橋本利典 中挟知延子[†] 島田静雄[‡] 近藤邦雄 佐藤尚[‡]埼玉大学工学部情報工学科[§]

1 はじめに

日本では、海外からの留学生が増えている。彼らにとって、日本語を使うこと、特に、日本語の文章を書くことが必要になる。彼らが書いた日本語の文章には、誤字・脱字などの表記上の誤りや文法上の誤りのある文が混じると同時に、文法上は正しいが日本人から見て意味の通じない文がある[1]。そこで、彼らにとって、自らの文章を訂正してくれる文書校正支援システムは、有用である。それと同時に、日本人の書く文章の校正にも役立つものと思われる。また、外国人の書いた日本語の文章は、彼らの母国語での考えを日本語に翻訳したものであると考えられる。彼らの文章を分析することにより、逆に我々が、日本語の構造を理解することになり、将来的に日本語と外国語の自動翻訳に役立つと考えている。本稿では、外国人が書いた文の誤りの分類をし、その誤りの校正ルールの確立を行う。漢字使用語圏の外国人を対象にしたシステムを試作し、日本人による校正と比較、評価する。このシステムは、小規模な処理系で実現することを目標としている。

2 誤りの調査と分類

情報工学科に所属する漢字使用語圏の留学生を主な対象とし、彼らが作成した日本語文章の中から、誤りの調査を行なった。誤りが含まれると日本人が判断した365文中に、384の誤りが見つかった。その分類の結果の概要を示す。

2.1 助詞の誤り

助詞の誤りが多く発見された。そのほとんどが、格助詞の選択における使用誤りである。我々、日本

人であるなら間違わないものばかりであるが、外国人には以下のような助詞の誤りが見られた。

- 格助詞「が」と「を」の使用誤り(15)
- 格助詞「を」と「に」の使用誤り(12)
- 格助詞「に」と「で」の使用誤り(9)
- 格助詞「に」と「の」の使用誤り(9)
- 格助詞「を」と「の」の使用誤り(8)
- 格助詞「で」と「が」の使用誤り(7)
- 格助詞「の」の欠如(10)
- 格助詞「に」の欠如(5)
- 接続助詞「ので」の使用誤り(5)
- 接続助詞「で」と「ので」(5)
- 格助詞「で」と「から」の使用誤り(6)
- 不必要な助詞(12)

ただし、カッコ内の数字は、サンプル数である。

2.2 漢字・熟語の選択誤り

漢字を使用する場合、日本語として通じるものと通じないものが見られた。また文脈を考えると使用すべきでない漢字や熟語を使っている場合も見られた。

例1: 「昔前の」→「以前の」

例2: 「便理」→「便利」

2.3 複合語の誤り

複合語の作成時の誤りがいくつか見られた。

● 複合名詞の使用誤り

例: 「歩き方や走り方や飛ぶ方」
→「～飛ぶ方」

● 複合動詞の使用誤り

例: 「～という点で御理解くれると思う。」
→「～で御理解してくれると思う。」

*Intelligent Proofreading System of Japanese Documents Written by Foreigners

[†]東京国際大学

[‡]Toshinori HASHIMOTO, Chieko NAKABASAMI, Shizuo SHIMADA, Kunio KONDO, Hisashi SATO

[§]Department of Information and Computer Sciences, Saitama University

2.4 用言の活用誤り

動詞の活用の仕方で見られた。

・動詞の中止法の使用誤り

例：「～に伴う、～が変わっていく。」
→「～に伴い、～が変わっていく。」

・自動詞と他動詞の使用誤り

例：「集まった。」→「集めた。」

・受身形の使用誤り

例：「ふたをあげて、E L 基盤をとられた。」
→「～、E L 基盤をとった。」

2.5 その他

以上の他に、気のついた誤りがあるが、分類が難しいので、単に数例を列挙する。

・意味不明な用語

日本人の間違いとは異質である。

・主語と述語の係り受けの不自然さ

・時制の不自然さ

・音声上の聞き間違いからくる表記上の誤り

「～では」を、「～てば」など。

・音便の使用誤り

などが目立った。

3 誤り文の校正

本研究では、誤りの中でも、助詞の使用誤りに着目した。それぞれの助詞に対して、日本語文として許容できる形を判定のためのルールとする。例えば、主格助詞「が」の場合は、以下のような形のルールを用いる。

例：・判断の主意概念を主格として導く「が」

- ・ A が B の C である。
- ・ A の B が C の D だ。
- ・ A が B です。
- ・ A まだが B の C だ。

これは、1例である。この他に「が」だけで、59の形を用意した。これを句点で区切られた一文を抜き出し、その文の漢字を省き、マッチングさせる。この形にマッチしないものを誤りとして抽出する。動詞の活用語尾も考え、否定、受身、復文や名詞句にも対応させる。そのために、動詞の活用語尾の辞書を用意する必要がある。さらに「を・に・の・

へ・で・から・より・や」の格助詞の日本語文として許容できる形についても151の形を用意した。同様にして誤りを抽出する。

校正は、抽出された助詞を書き換える。例えば、「が」の判定のルールで抽出された助詞「が」は、誤りの調査のデータを参照した結果、「を」に書き換える。このように、調査結果をもとにして、助詞の書き換えを行う。

4 考察

以上より、格助詞の使用誤りの抽出と校正することができると思われる。

また、日本語文の典型的な形をルールとして用いた理由は、例えば、「A が B の C である。」の文を見ると、A、B、Cにどんな名詞をいれても文が成立する。

「彼 が 私 の 兄 である。」

「今日 が 学会 の 日 である。」

上のような文が成立する。日本語文には、このような形の文が存在する。よって形に合わない文が誤り文として抽出されたと考えたからである。

しかし、格助詞の誤りを抽出して、自動的に校正した助詞が誤っている可能性がある。現時点でこれを判別するには、人間が行わなければならない。その自動化の精度を上げるためには、さらなるサンプル数を増やし、誤りの分類の信頼性を上げる必要がある。

5 まとめ

本研究では、外国人が書いた日本語文について、誤りの調査と分類を行なった。そして、その誤りの中で、一部の助詞の抽出と校正方法を考察した。しかし、コンピュータによる自動的な校正は、難しいので、今のところ誤りを抽出し、最後は人間が判断し、校正しなければならない。今後も、助詞の校正を主に、サンプル数を増やし、さらに分類を行い、誤りの抽出と校正方法を確立していく予定である。

参考文献

- [1] 坂倉守 他:「外国人の書いた日本語文の訂正に関する基礎的研究」情処研報 94-NL-99-8 (1994)
- [2] 菅沼明 他:「字面解析による助詞「が」の抽出」情処第 47 回大会, 1W-10 (1993)
- [3] 江副隆秀:「日本語を外国人に教える日本人の本」創拓社 (1985)