

漢字に関する異表記の調査

6F-5

伍井啓恭、鈴木克志、丸山冬樹、太細孝

三菱電機(株) 情報電子研究所

1. はじめに

前回の報告^[1]において日本語の片仮名表記のゆれに対応した処理方式として「カタカナ異表記処理」の報告をした。今回さらに漢字に関する表記のゆれについて、機械翻訳用辞書の一部、約8800語に対してどのような漢字異表記が存在するか調査したので報告する。

2. 漢字異表記の種類

漢字異表記は以下の3つに大別される。

2-1. 漢字自体の表記のゆれ

漢字自体の表記のゆれには以下のものがある。

2-1-1. 異体字

異体字は本来の親字に対して、同音、同義でありながら、その字体の違いにより異表記となってしまふものであり J I S 漢字コード中には次のような種類がある。

俗字	奇(奇)	劍(劍)
旧字	應(応)	劍(劍)
古字	气(氣)	劍(劍)
本字	苺(莓)	
別体字	壺(壺)	
誤字	達(達)	壺(壺) 劍(劍)

() 内は親字

2-1-2. 異漢字

異漢字も同音、同義(または、ほぼ同義)でありながら、その漢字表記の違いにより異表記となってしまふものであるが、異体字と決定的に異なる点は、必ずしも表記の対応が1文字毎でないことである。

例 竹の子 筍
日照り 旱

2-1-3. 繰り返し符号

繰り返し符号「々」が用いられることによって異表記となるものである。

例 人々 人人
日々 日日

2-2. 送り仮名の許容のゆれ

送り仮名の許容のゆれは、用言の送り仮名の表記の曖昧性から生じる異表記で例えば以下のようなものがある。複合動詞の88.7%がこれにあたるという報告もある。^[3]

例 受け付け 受付け 受付

2-3. 仮名書きによる表記のゆれ

仮名書きによる表記のゆれは漢字の読みが平仮名または片仮名で書かれることによって起こる表記のゆれである。この類の異表記には以下のものがある。

2-3-1 単純な漢字仮名の混ざり書き

漢字の読みを平仮名または、片仮名で書く事により生じる異表記。一般に、連体詞、感動詞、助詞、補助動詞、形式名詞、接辞、当て字、または誤読のおそれのあるものは、平仮名書きしやすく、動植物名、外来語、擬音、擬声語、擬態語、俗語、隠語などは片仮名書きしやすい。

例 出どころ 出所
非常ボタン 非常釘

2-3-2 読みが音便化して変化する

漢字の読みが仮名で書かれ、これがさらに音便化することにより生じる異表記。音便にはイ音便、促音便、拗音便、ウ音便がある。

例 取付く とりつく とっつく
香しい かぐわしい こうばしい

2-3-3 読みが濁音化して変化する

漢字の読みが仮名で書かれ、これがさらに濁音化(以下、連濁という)することにより生じる異表記。連濁するか否かは以下に示すような規則がある。

★前の音が促音のときは連濁しない
取っ手(トッテ) 取っ付き(トツキ)

★擬声語、擬態語は連濁しない

カンカン クンクン

★用言+用言は連濁しない

積み重ねる(ツミカケル) 振り掛ける(フリカケル)

★連用形名詞+用言は連濁する

行き詰まる(イキジマール) 行き止まる(イキトマール)

★一般に慣用されることの少ない語

抗する(コウスル) 吊する(チウスル)

しかし、必ずしもこの規則が適用できるとは限らない。

2-3-4 その他の音の変化を伴うもの

漢字の読みが仮名で書かれ、これが上記以外の音の変化をすることにより生じる異表記。

★拗音の直音化

新宿(シンジュク→シンジツク)

★一種の促音化

三角形(サンカクケイ→サンカクケイ) 菊花(キカケイ)

★接辞の音の変化

雨宿り(アメドリ) 稲穂(イネ) 酒屋(サカ)

3. 漢字異表記の調査結果

機械翻訳用の辞書約5万語より漢字仮名混じりのもの約8800語に対して送り仮名の許容、仮名書きによる表記、異漢字のバラエティーを全てピックアップした。異体字、繰り返し符号による異表記は機械処理が比較的簡易に実現できるため除外した。その結果、表記のバラエティーは約4万表記(正表記は除く)であることがわかった。

4. おわりに

機械翻訳用の辞書の一部について漢字に関する表記の曖昧性を調査した。その結果、1語につき平均して約4.5パターン of 異表記が存在することがわかった。これらの殆どは送り仮名異表記であり自動修正の報告^[8]もあるが音便、連濁などが複雑に絡んでいるため、音素だけの片仮名異表記のような機械的な処理は難しいと思われる。

参考文献

- [1] 伍井他: カタカナ異表記処理, 情報処理学会第38回全国大会 (1989).
- [2] 日本語発音アクセント辞典, 日本放送協会編.
- [3] 梅田他: 漢字カナ変換の一方式, 情報処理学会第32回全国大会 (1986).
- [4] 小林他: 文章作成支援システムにおける日本語処理(1), 情報処理学会第35回全国大会 (1987).
- [5] 亀田他: 未知語の分類とその処理規則, 情報処理学会第36回全国大会 (1988).
- [6] 清水他: 辞書検証システムの構想, 情報処理学会第36回全国大会 (1988).
- [7] 高尾他: 日本語文書リーダ後処理の実現と評価, 情報処理学会論文誌, Vol.30, No.11 (1989).
- [8] 中村他: 和文形態素解析における異なった送り仮名の自動検出と自動修正, 情報処理学会第34回全国大会 (1987).

Classification of notational variants in Kanji
Hiroyasu Itsui, Katsushi Suzuki, Fuyuki Maruyama, and
Takashi Dasai
MITSUBISHI Electric Corp.