

漢字カタカナ混じり変換方式の検討*

1S-2

田村 俊哉[†] 福永 幸弘[‡] 井上 健[†] 鈴木 謙二[†]
(株)東芝 マルチメディア技術研究所[†] 東芝 AVE(株)[‡]

1 はじめに

漢字混じりかな漢字変換[1,2,3,4]は、「会ぎ→会議」のように漢字かな混じりの文字列をかな漢字変換することができる。この漢字混じりかな漢字変換によって、比較的簡単な漢字は直接入力し、難しい漢字はひらがなで入力した文字列を変換することができる。ペン入力の利点を生かした入力が可能になる。

一方、メモ書き等の場面では「会ギ」のように漢字がカタカナで表記されることがある。カタカナはひらがなに比べて字体が単純であり、カタカナ表記は筆記の手間を少なくするために有効な手段である。カタカナ表記をペン入力に取り入れることよって、ペン入力の自由度が高まり、ユーザに対してより自然な入力環境を提供できると考えられる。そこで、漢字ひらがな混じりの入力に加えて、カタカナ混じりでの入力も可能とする漢字カタカナ混じり変換方式について検討を行なった。

2 カタカナ混じり入力

メモ書き等の場面では複雑な漢字をカタカナ表記で代用することができる。

会ギ(会議) 事ム所(事務所)
庶ム(庶務) 休カ(休暇)
キ材(機材) 直キ(直帰)

図 1: カタカナ混ぜ書きの例

これは、複雑な漢字を画数が少なく字体の単純なカタカナで表記することで幾らかでも筆記の手間を省こうとする行為といえる。さらに、この書き方をメモ書きに限らず文章を記述する際に利用することを想定してみると、ひらがな部分をひらがなで筆記し、漢字部分を漢字またはカタカナで筆記することになる。ペン入力においてこのよう

*A Study of Kanji-mixed Katakana to Kanji Conversion, by Toshiya TAMURA Yukihiro FUKUNAGA Takeshi INOUE Kenji SUZUKI

(e-mail: tamtam@pms.mmlab.toshiba.co.jp)

[†]TOSHIBA Corp. Multimedia Engineering Lab.

[‡]TOSHIBA AVE co.,Ltd.

な文章入力が可能になれば、ユーザフレンドリーなインタフェースを提供するという観点からも有効である。

漢字混じりかな漢字変換においては、カタカナで漢字を代用するという事はカタカナ表記によって漢字への変換を指示していると解釈できる。これを逆の面からみると、ひらがなで表記することは無変換を指示していると考えられる。そこで、カタカナ表記を入力文字列中の変換部分の指定、ひらがな表記を変換しない部分の指定として利用するという考え方に基づいて、次章で述べる2つの漢字カタカナ混じり変換方式について検討した。

3 漢字カタカナ混じり変換

3.1 変換部分指定方式

変換部分指定方式は、2章で述べた考え方をそのまま取り入れ、入力の指定通りの表記となる語を変換候補として出力するものである。この方式ではカタカナ入力によって変換後の表記が指定できるため、同音語候補の絞り込みや文節区切り間違いの抑制等の効果が期待できる。

テイ案したイケン → 提案した意見
テイ案したいケン → 提案したい件

図 2: 変換部分指定方式の変換例

3.2 自立語指定方式

変換部分指定方式では漢字表記する部分を意識しながら入力しなければならないため、例えば、「ハシる(走る)」か「ハしる(走しる)」かのように入り仮名などで迷うことがある。また、入力の指定通りの表記が辞書に登録されていない場合には変換候補が出力されないという不具合もある。こうしたことを考慮すると、カタカナ入力を上述したような狭義の“変換したい部分の指定”とした場合必ずしも使いやすいものとはならない。そこで、カタカナ入力を“変換したい自立語の指定”にまで拡張した自立語指定方式を考案した。このように拡張した場合、「ハシル」と入力しても「走

る」が変換候補に挙がるようになる反面、入力のカタカナと変換後の表記(漢字・カタカナ)の一致度によってはユーザが意図しないものが出力されることがある。そこで、変換候補を下記のような4段階に分類し、1>2>3>4の順に優先して出力することにより対処した。これにより、本方式ではユーザの意図に近い変換候補から出力されるようになる。

1. カタカナ部の変換後の表記が漢字、且つひらがな部の変換後の表記がひらがな。(変換部分指定方式)

アカサビ→赤錆、サク→咲く…

2. カタカナの一部の変換後の表記がひらがな。

アカサビ→赤 さび

3. ひらがなの一部の変換後の表記が漢字。

サク→佐 久、まなデシ→愛 弟子

4. 上記を満たさず、カタカナ/ひらがな入力に関わらず読みが一致する。

サク→策、ミズから→自ら

4 検証

当社で開発した漢字混じりかな漢字変換システム[4]をベースに2つの方式による変換システムを試作した。上述した方式の有効性を検証するために1つの定量的な指標である変換率に着目し、本方式により変換率がどのように変化したかを調査した。図3は7文章¹について測定した変換率²を示したものであり、漢字混じりかな漢字変換には原文中の6画以下の漢字は漢字、7画以上の漢字はひらがなで表記したデータ、2つの漢字カタカナ混じり変換には6画以下の漢字は漢字、7画以上の漢字はカタカナで表記したデータをそれぞれ用いた。

どちらの方式とも平均で5%弱変換率が向上している。これは、カタカナ混じりの入力を適切に変換でき、さらに同音語候補の絞り込みや文節区切り間違いの抑制等の効果があることを示していると考えられる。この測定結果から、変換率の面でも従来の漢字混じりかな漢字変換システムを上回ることが判り、入力の自由度を高めたことと共に本方式の有効性が確かめられた。

¹新聞記事(150文節程度)が4文章と、論文(700文節程度)が3文章。

²ここで示した変換率は1次変換率である。

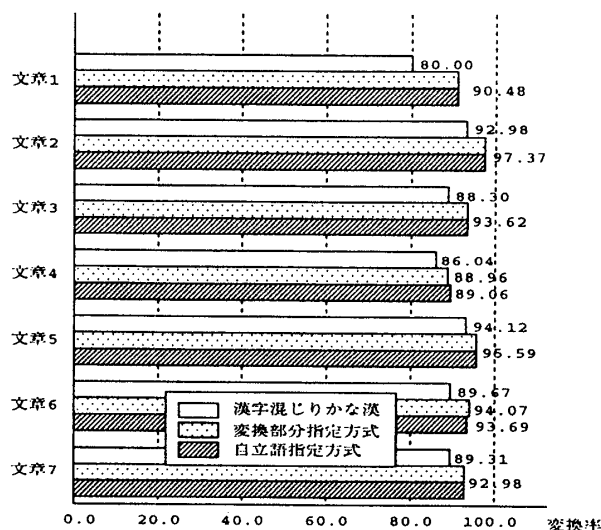


図3: 変換率の測定結果

5 おわりに

メモ書き等で複雑な漢字をカタカナ書きすることに着目して、カタカナ混じりの入力も可能とする漢字カタカナ混じり変換に対する1つの考え方を示した。また、システムを試作し、カタカナ混じりの入力が適切に変換できること、及び本方式によって得られる効果が確かめられた。今後は、入力の簡便さ等利用者の立場からの評価実験も行なって最適な漢字カタカナ混じり変換の方式を確立するとともに、実機上での漢字混じりかな漢字変換システムへの組み込みについても検討していきたい。

参考文献

- [1] 塩見他：2ストローク入力のための仮名漢字変換，情報処理学会論文誌，Vol.33，No.7，pp.920-928(1992)
- [2] 山田：漢字混じりかな漢字変換用単語辞書の作成，情報処理学会第46回全国大会，5L-3(1993)
- [3] 金子他：代表読み辞書を用いた交ぜ書き漢字変換，情報処理学会論文誌，Vol.35，No.6，pp.1111-1118(1994)
- [4] 山本他：ペン入力パソコン *Dynanote*TM，東芝レビュー，Vol.48，No.2，pp.94-97(1993)