

## 漢字カタカナ混じり変換システムの試作\*

3Q-2

田村 俊哉<sup>†</sup> 古市 佳男<sup>†</sup> 鈴木 謙二<sup>‡</sup>(株)東芝 マルチメディア技術研究所<sup>†</sup> (株)東芝 情報サービス・機器事業開発室<sup>‡</sup>

## 1 はじめに

漢字混じりかな漢字変換は、「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可能である[1,2,3,4]。これによって、比較的簡単な漢字は直接入力し、難しい漢字は平仮名で入力した文字列を変換することができ、ペン入力の利点を生かした入力が可能になる。さらに、ペン入力の自由度を高めるために、メモ書き等の場面で「会ギ」のように漢字がカタカナ表記されることに着目し、カタカナ入力を漢字への変換指示と定義した漢字カタカナ混じり変換方式を検討してきた[5]。

この変換方式を使う立場から調査してみると、実際のメモ書きなどでカタカナ混じりの表記が用いられる反面、平仮名の方が書き慣れている、あるいは崩すことができるので書き易いという声があることや、カタカナ/平仮名の厳密な使い分けは難しいことが判った。そこで、従来の漢字混じりかな漢字変換を維持しつつ、変換単位にカタカナが含まれる場合に限り、カタカナ入力を変換指示と解釈して漢字カタカナ混じり変換を行うという仕様でシステムを試作した。

## 2 漢字カタカナ混じり変換

カタカナで漢字を代用するという事は、漢字混じりかな漢字変換において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。他方、平仮名表記は無変換を指示していると考えることができる。そこで、カタカナ表記を入力文字列中の変換部分の指定、平仮名表記を無変換部分の指定として利用するという観点から、次の2つの変換方式を考案した。

## 2.1 変換部分指定方式

変換部分指定方式は、上述した考え方をそのまま取り入れ、入力の指定通りの表記に変換される語を変換候補として出力するものである。

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

## 2.2 変換単位指定方式

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、「オダヤか(穏やか)」か「オダヤか(穏か)」かのように送り仮名などで迷うことがある。こうしたことを考えると、カタカナ入力を上述のような狭義の“変換したい部分の指定”とした場合必ずしも使い易いものとはならない。そこで、通常のかな漢字変換と同様にカタカナ入力を“変換したい単位(自立語)の指定”にまで拡張した変換単位指定方式を考案した。このように拡張した場合、「オダヤカ」と入力しても「穏やか」が変換候補にあ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)との一致度によってはユーザが意図しないものを出力してしまうことがある。そこで、入力指示との一致度に応じて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。

1. カタカナの変換後の表記が漢字、且つひらがな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

## 3 ニーズ調査

社内でアンケート調査を実施し、漢字カタカナ混じり変換に対する利用者のニーズを調査した。

1. 電話、会議のメモといった日常の場面でカタカナ混じり書きを使用するという回答は、全体の約62%であった。使用する場面は、「会議・電話でのメモ」、「スケジュールなど自分用のメモ」が大半であった。また一方で、メモ書きする場合に

\*An Experimental System of Kanji-mixed Katakana to Kanji Conversion, by Toshiya TAMURA Yoshio FURUICHI Kenji SUZUKI

<sup>†</sup>TOSHIBA Corp. Multimedia Engineering Lab.

<sup>‡</sup>TOSHIBA Corp. Information Services & Portable Equipment Development Office

はひらがなの方が書きやすいという回答が約61%を占めた。この理由として、「ひらがなの方が書き慣れている」、「ひらがなの方が崩し易い」という意見があげられていた。

2. 漢字カタカナ混じり変換に対する要望としては、「シンプルな仕様」、「ひらがなを無変換、カタカナを変換指示と厳密に使い分けることは難しい」、「カタカナ語を誤って変換しないように」といったことが多かった。
3. 変換指示としてのカタカナ表記を使用する程度に応じ、「全く用いない」もしくは「慣用的なもの(会ギ etc)だけ用いる」という利用者と、「ある程度漢字の代用として用いる」という利用者に大別することができる。

#### 4 漢字カタカナ混じり変換の実現

ニーズ調査の結果を踏まえて、従来の変換を維持しつつ、慣用的なものについてカタカナ混じり変換機能を追加したものであることを前提とし、さらにカタカナと平仮名を使い分けることにより変換後の表記が指定できる応用的な使い方も可能な変換システムを提供するという方針で漢字カタカナ混じり変換を実現した。

##### 4.1 変換部分指定

カタカナと平仮名を使い分けることは難しいので変換単位指定方式を採用すべきと考えられるが、カタカナ混じり書きが利用されるのは名詞、サ変動詞であり、送り仮名を伴う動詞などではあまり利用されないこと、シンプルな仕様が適切であると考えられることから、漢字カタカナ混じり変換を実現する上で変換部分指定を採用している。

入力1: 「**後期**」  
 出力1: **1.後期 2.好奇 3.好機** ...  
 入力2: 「**好奇**」  
 出力2: **1.好奇 2.好機**  
 入力3: 「**好き**」  
 出力3: **1.好き 2.透き 3.空き** ...

図1: 単語の変換例

##### 4.2 漢字カタカナ混じり変換の起動条件

ユーザが入力する際に文節などを意識することは少ないこと、カタカナ混じり書きが単語単位の入力で行われやすいことを考慮し、漢字カタカナ混じり変換は該当する入力にカタカナが含まれる単語について動作させている。

入力1: 「ていあんしたいけん」  
 出力1: **1.提案したい件 2.提案した意見** ...  
 入力2: 「ていあんした**意見**」  
 出力2: **1.提案した意見**  
 入力3: 「**提案**したいけん」  
 出力3: **1.提案したい件**

図2: 連文節の変換例

#### 4.3 カタカナ未知語

漢字カタカナ混じり変換では、カタカナは変換指示と解釈されるので、辞書に登録されていないカタカナ語を入力した場合には、「ハリス→針巣」ように過度に変換してしまうことがある。そこで、入力されたカタカナ文字列をカタカナ語としても扱うようにし、他の変換候補と同様に処理している。

入力: 「ハリスがきた」  
 ハ [葉、歯、…]  
 ハリ [針、梁、…]  
 リス [リス] (..が)  
 ス [巣、洲、…] (..が)  
 き [来、着、…] (..た)  
 きた [北、喜多、…]

… 追加されるカタカナ文字  
 [ ] 内 … 辞書検索成功語  
 ( ) 内 … 接続可能付属語

図3: カタカナ未知語の処理

#### 5 おわりに

カタカナ入力を漢字への変換指示と定義した漢字カタカナ混じり変換の実現について述べた。試作したシステムはこれまで検討してきた方式を使う立場から評価し、ユーザのニーズを反映した形で実現しており、慣用的なカタカナ混じり書きの変換に加え、変換後の表記指定も可能である。今後、さらに評価を進め、より使い勝手の良いシステムとしていきたい。

#### 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*™, 東芝レビュー, Vol.48, No.2, pp.94-97(1993)
- [5] 田村他: 漢字カタカナ混じり変換方式の検討, 情報処理学会第49回全国大会, 1S-2(1994)