

仮名漢字変換における誤入力自動補正機能の実現

1 M-6

小林勉 石塚靖 中里茂美

(株) 東芝 マルチメディア技術研究所

1 概論

キーボードを用いて日本語入力を行なう場合、一般的に使われているキーボードのキーの総数は、日本語で用いられる文字集合の要素数に比べて圧倒的に少ないため、キー打鍵のシーケンスを日本語文字集合へ変換する工夫が必要となってくる。

現在、その主流の位置にあるのが「仮名漢字変換方式」である。仮名文字列の入力方式には、キーボードから仮名を直接入力する「仮名入力方式」と、キーボードからはローマ字を入力し、これを仮名文字列に変換した後に仮名漢字混じり文字列へと変換する「ローマ字入力方式」とが存在する。

仮名文字列から仮名漢字混じり文字列への変換は、100%自動的に行なわれるのではなく、正解候補の選択など、入力者の補助が必要となる。このため、頭の中にある思考素片を文字列として表現するとき、日本語入力方式においては、欧米の入力方式に比べて余分なプロセスが必要となってくる。この余分なプロセスにかかるユーザの負担を極力減らすことが、より良い入力システムであるための重要な要件となる。

2 誤入力自動補正の位置づけ

我々は、円滑な日本語入力処理系を実現すべく、仮名漢字変換の変換率を高める研究を進めてきた。この結果、変換精度は向上し、日本語入力操作における、変換キー押下数に対する次候補選択キー押下数の割合は、年々減少してきている。しかし、依然として仮名文字列から仮名漢字混じり文字列への変換というプロセスが介在していることには変わりがなく、より効率の良い日本語入力のためには、もう一步踏み込んだ工夫が必要である。仮名漢字変換処理系に、単なる「仮名文字列」から「仮名漢字文字列」への変換を任せるだけではなく、よ

り積極的にユーザの入力作業を補助する機能を持たせれば、入力作業の負担を軽減することができる。

従来の仮名漢字変換処理系は、ユーザから入力された仮名文字列には間違いがないものとして変換処理を行なっていた。だが、実際の入力作業には、入力ミスが伴う。ユーザが入力ミスに気づかずに、そのまま変換を行なってしまうと、得られる変換候補はユーザが予期している候補とかなり違うものになることが多い、このためユーザは思考の中断を迫られる。この場合、ユーザは変換結果が予期したものと違っている原因が入力ミスであることを認識し、その入力ミスを補正した後、再度変換を行なう必要がある。

こういったユーザの入力ミスを検知し自動的に補正するため、誤入力自動補正機能を実現した。

3 設計方針

誤入力自動補正機能の設計に当たっては、次の点に留意した：

1. 誤訂正を極力避ける
2. 誤訂正をしてしまったときの復旧手段を設ける
3. 誤入力自動補正が機能したことを視覚的に伝える
4. 速度低下を抑える

誤入力自動補正の目的は、ユーザの入力作業を積極的に支援することであるから、誤訂正をしてしまっては、かえってユーザに負担をかけてしまうことになる。これは避けなければならない。また、万が一誤訂正をしてしまったときに、それを復旧する手段を設けておくことが望ましい。さらに、自動補正が行なわれたことにユーザが容易に気づくようにしておく必要がある。そうでないと、万が一誤訂正をしてしまったときに、ユーザがそれと知らずに入力を��けてしまう可能性がある。これらの

機能は、仮名漢字変換処理系の変換速度の低下を極力抑える形で実現しなければならない。

4 システム構成

誤入力自動補正機能を持つ仮名漢字変換系の概要構成図を、図1に示す。

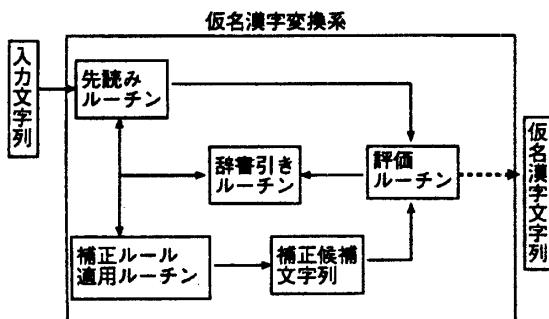


図1: システム構成

補正ルールは、補正パターンと補正アクションとから構成されている。補正パターンにマッチする文字列が入力されたときに、補正アクションが実行される。補正アクションでは、ルール適用のための追加条件のチェック、候補文字列の生成、ルール適用ルーチンの制御などの処理を行なう。これらの処理は、ユーザが文字列を入力している時に並行して実行される。ユーザが入力作業に専念しているときのCPUのアイドル時間を利用することで、変換キー押下後の処理を軽減させている。

また、仮名漢字変換系の処理では、ディスクへのアクセスがオーバーヘッドとなるため、各種キャッシュデータを有効活用することによって、ディスクへのアクセスを最小限に抑えている。

ユーザによって変換キーが押下されると、ユーザの入力文字列および補正候補文字列を評価関数に通し、誤入力があったかどうかの最終判断を行なう。評価関数は、各ルールごとの評価点数、ユーザ入力文字列から生じる変換文字列の評価点、候補文字列の評価点などが絡み合った経験則を反映した関数となっている。

誤入力自動補正の例を以下に挙げる：

1. 語彙ルール (コミュニケーション, いづれ)
2. 長音とハイフンの使用誤り (キーボード)
3. ローマ字入力の誤り (ほにやく, にゅりょく)

4. 仮名入力の誤り (ふあーぶるこんちゅうき)

これらの誤入力補正ルールがかかると、ルールがかかる箇所の表示色を変えて表示する。この機能により、システムが万が一、誤って補正してしまった場合に、ユーザがそれにすぐに気づき、補正を取り止めることができる。また、「いづれ」などの単語が本来は誤りであることをユーザに認知させるのにも役立つ。

5 評価および今後の改良点

誤入力自動補正機能を実現していく上で、正しい文章を誤って補正してしまうことを防ぐためのトライ＆エラーが最もコストがかかる作業であった。たとえば、ユーザが文章の一部分を入力して変換した場合、誤って訂正してしまう可能性が高くなる。「しゅまつにあう」を「しゅうまつにあう(週末に会う)」と補正するためのルールが「1しゅ」を「1しゅう(1種)」にしてしまうなどの問題が生じた。この場合、文章としての完成度(文章の断片ではなくて、主語・動詞・目的語があるなど)を評価関数のパラメータとして盛り込むことで誤った補正を回避できる。

こういった改良作業は、主に次の2点に集約される：

- 補正ルールの詳細化。補正アクションにおけるルール適用の見直し処理の追加など
- 評価ルーチンの詳細化。例えば、単語間をまたがる補正に対しては点数を低くするなどの処理

この一連の作業により、評価ルーチンが複雑化してしまったことは反省すべき点である。これについては今後の改良項目としたい。他の改良項目として、ユーザの入力ミスの学習、および評価ルーチン内で使われている評価値の動的変更などを考えている。

6まとめ

我々は、日本語入力作業の軽減のために、仮名漢字変換処理系に単なる「仮名文字列」から「仮名漢字混じり文字列」への変換だけをやらせるのではなく、より積極的にユーザの入力作業を支援することが大切であると考えている。そのアプローチの一環として、誤入力自動補正機能を搭載した仮名漢字変換処理系を実現した。今後、その他の入力支援機能を統合することで、より効率のよい日本語入力の枠組を実用化していく。