

3P-9

ラップトップUNIXのかな漢字変換

福中 儀邦 山田 邦夫 中原 康

(株)東芝 青梅工場

1.はじめに

現在、UNIX*はEA・OAワークステーションなどの広い分野において、標準OSとして定着しつつある。しかし、その日本語処理機能は十分ではなく、UNIXが日本に導入されて以来様々な議論や拡張がなされて来た。

この間、AT&T日本語UNIXシステム諮問委員会や通産省指導プロジェクトのΣOS日本語拡張機能仕様の制定などの標準化活動を経て、ようやくAT&TのUNIX System V Release 3以降の標準リリースや日本語処理パッケージのJAE/JIOという形でUNIXの8ビット化、16ビットコードのサポート、日本語入力機能など、日本語処理には不可欠な問題も、UNIXの国際化の観点から議論され、統一的にサポートされるようになった。

日本語処理の観点からすると、このサポートは特定の日本語処理機能、特に日本語入力を固定的に実現したのではなく、多様な日本語入力のカスタマイズを可能にする柔軟な機構を実現しているところに特長がある。

当社では、UNIX System V Release 3及びJAEをベースとしたラップトップUNIXワークステーションを開発した。本稿では、その日本語入力、特にな漢字変換について報告する。

2.実現方法

日本語入力機能、特にな漢字変換は、OSのインタフェースの中に自然に溶け込ませながら、OSとは別のプロセスにより実現するのが一般的である。その特別なプロセスはユーザプロセスとデバイスドライバの間に存在することになるが、その実現方法は様々であり、統一されるには至っていない。

前述のJAE/JIOは、UNIX System V Release 3で導入されたストリーム機構を用いたもので、かな漢字変換のインプリメントに1つの方向を示したものとして注目される。

この方式はユーザプロセスとデバイスドライバの間に特別なモジュールを挿入し、そのモジュールを介してかな漢字変換プロセスを結合させるというものである。

この方式の特長は、ストリームモジュールの部品として

の独立性やかな漢字変換プロセスとのインタフェースの単純性にあり、その結果、任意のかな漢字変換を自由に組込むことができ、また切り替えて使用することができるというところにある。ただし、その特長を生かすにはストリームモジュールとかな漢字変換プロセスの役割分担を明確化しておくことが重要である。

今回開発したシステムでは、ストリームモジュールには単に通信を行なう為の機能を持たせるだけとし、かな漢字変換固有の処理、例えばマンマシーン・インタフェース(MMI)等、はすべてかな漢字変換プロセス側に持たせることとした。今後のかな漢字変換の発展性やストリームモジュールの独立性・汎用性を重視したからである。

3.日本語入力機能3.1 かな漢字変換

日本語の入力方式は、漢字の読みをそのままひらがなで入力し、対応する漢字に変換する“かな漢字変換方式”を採用した。さらに、基本的に変換指示の操作を必要としない「ベタ入力かな漢字変換」とした。この変換方式は、逐次入力される文字列をかな漢字変換が自動的に文法解析し変換処理を行い、入力と平行してその変換結果の表示・確定を行っていくので、連続的な文章の入力が可能な方法である。この方式を採用したのは以下の理由からである。

- ・漢字の入力方法としては、かなから漢字へ変換する方法がユーザにとってなじみ易く、自然な方法である。
- ・日本語の入力を主体に考えるプログラムへの応用(日本語エディタ等)まで考えると、現在日本語ワードプロセッサで主流になっている変換レベルまで実現すべきである。
- ・特に入力過程において、日本語も英語と同じように変換を意識せずに入力できる方式を追及すべきである。

3.2 辞書

- ・標準辞書
一般的に使用される単語の読み、見出し、品詞情報が格納されており、人名地名等の固有名詞もこの辞書に格納されている。
- ・単漢字辞書
音訓変換、部首変換に必要な読み、見出しが格納されている。
- ・新語辞書

ユーザが単語を登録するために使用する、また学習情報の格納に使用される。

ベタ入力かな漢字変換では、一般的に辞書へのアクセス回数が多くなり、変換スピードの低下をまねがちであるが、変換スピードの低下は、入力過程にくみこまれるかな漢字変換において、致命的であり、高速化の工夫が必要となる。

このシステムでは高速化のために、1)逆引き構造の標準辞書、2)結合検定の重複処理を避けるために同音同品詞語のグループ化、を採用している。

3.3 操作環境の設定

かな漢字変換用の環境ファイルを用意し、ユーザが操作環境を設定できるようにしている。この環境ファイルには、変換操作キーを定義するマッピング情報、ブリンク、反転など、入力や候補表示で使用する表示属性、入力モードの初期値、使用する辞書ファイル名、などを設定することができる。この様な環境ファイルを用意することにより、ユーザごとに操作環境を設定できると共にいろいろな漢字端末での使用が可能になり、MMIの柔軟性を高めている。

4.入力時の操作・編集機能

入力時の操作・編集に係わるMMIは、ベタ入力かな漢字変換の特長を生かすものでなければならない。

一般に変換キーを必要とするかな漢字変換の入力操作の流れは、

- (1) 読み入力
- (2) 変換
- (3) 候補選択
- (4) 決定

であり、(1)から(4)までを繰り返して日本語の文章を入力していくのが基本である。

しかし、ベタ入力では自動的に(2)の操作が行われるが、自動的に文節切りを行ない変換結果を表示するために、漢字の確定に至るまでの入力領域として従来以上の長さが必要になる。また、思わしくない候補表示が行なわれた時は、容易にその修正が行なえるような編集機能もより一層充実させなければならない。一方、いかなるかな変換方式においても共通な、同音意義語の選択操作や、ひらがな/カタカナ/英字/記号、全角/半角など多字種の取扱い、読みのわからない漢字の入力方法など様々な問題がある。

これらを解決するために本システムでは、以下の機能を用いたかな漢字変換プロセスのMMIとして持たせ操作性の向上を計っている。

・入力文字種の制御

全角/半角、ひらがな/カタカナ、かな/ローマ字などの入力する文字種の設定。

・入力文字種の変換

入力した後にその文字列を他の文字種に変換する機能。例えば、ひらがなで入力したものをカタカナに変換する。(あめりか → アメリカ)

・文節の区切り位置指定、再変換機能

文節の区切りを間違えて変換された場合や、入力ミスにより間違った読みを入力して変換してしまった時の修正機能。

・音訓/部首/区点変換

通常の変換では変換できない単漢字の音読み/訓読みや部首名による漢字入力、記号の区点コードによる入力。

・ユーザによる単語の登録/削除機能

標準辞書に登録されていない単語の登録/削除機能。

・入力領域超過時の自動確定

連続入力を妨げないよう、入力領域を越えそうになった時の未確定漢字を順次自動的に確定させる機能。

この様な方法をとることにより、UNIXのシステムでサポートするかな漢字変換といえども、不必要な変換操作を意識することなく日本語の連続入力が可能となっている。

5.おわりに

ラップトップUNIXは、個人の情報処理環境として位置づけられ、多様な個人環境にも十分耐えるものでなければならない。今回は、そのような観点から、かな漢字変換のMMIを含めた環境設定の柔軟性・多様性を考慮した開発を行なった。しかし、変換率やMMIにはまだまだ残された問題も多い。今後とも、辞書チューニングやAI技術の導入等による変換率の向上、MMIの向上を計り、日本語入力環境の充実に努めていきたい。

【参考文献】

- (1)『ΣOS-V0機能仕様書(第2版)』,情報処理振興事業協会 Σシステム開発本部,1987.
- (2)H.Kogure,R.McGowan: "A UNIX System V STREAMS TTY Implementation for Multiple Language Processing," AT&T Secho Asia/Pacific Edition,1987.
- (3)中原,他:「UNIX環境下での日本語処理」,情報処理学会コンピュータシステムシンポジウム論文集,1985.
- (4)山田,他:「UNIX System Vにおける日本語環境」,情報処理学会第36回全国大会論文集5D-9,1988.
- (5)武田,他:「ベタ入力かな漢字変換方式のワープロへの実現」,情報処理学会第30回全国大会論文集1G-8,1985.
- (6)楠元,他:「ベタ入力かな漢字変換方式における自動分割」,情報処理学会第32回全国大会論文集2T-6,1986.

*UNIXはAT&Tが開発しライセンスしているOSです。