

## 日本語UNIX・UX-300シリーズ OS/UX

中原 康 (東芝・青梅工場)

## 1. はじめに

UNIXは、その柔軟性において他のOSとは、“一味違った”OSである。それは、UNIXでは、徹底した“build-block”思想により部品化されたソフト群と、それらを自由かつ有機的に結び付けるツールが用意されていて、利用者が自分の利用環境に合わせて、自分なりの使い方や工夫ができるからであろう。Shellに見られる優れたマン・マシンインタフェースなどはその代表である。

このUNIXも日本での普及となると、やはり日本語の取り扱いが大きな問題である。UNIXの設計思想を損うことなく自然な形で日本語処理を組み込み、日本語処理特有なコマンドは勿論のこと、既存の豊富なコマンドでも容易に日本語を取り扱えるような改良が望まれるところである。

東芝のUX-300シリーズでは、こうした観点から、UNIXにOSレベルでの日本語処理機能を追加し、またその親和性の高い日本語処理方式の利点を生かして、日本語入出力機能の強化拡充やShellなど上位ソフトなどの日本語処理を充実させ、“日本語UNIX”としてその形態を整えている。〔1〕～〔6〕

本稿では、その日本語処理実現について解説する。

## 2. 日本語組み込みの問題点

計算機に日本語処理を取り込むには、一般に次の三点を解決する必要がある。

- 〔I〕日本語入出力システム(特に、日本語入力)
- 〔II〕日本語(漢字)コード
- 〔III〕日本語処理プログラム

## 2.1 日本語入出力システムの問題

まず、日本語入出力の問題であるが、日本語入力が最も重要であろう。その入力方式としては、漢字変換方式とそのシステムでの位置付けが問題である。

漢字変換方式としては、さまざまなものがあるが、やはり使いやすさ、親しみやすさの点から“かな漢字変換”であろう。この“かな漢字変換”にもいろいろあって、端末にその機能を持つもの、本体のソフトで処理するもの、また、“かな漢字変換”といっても単なる“音訓変換”のものや高度な文法処理機能を持つものなどがある。

次の日本語入出力のシステムでの位置付けとも関連するが、UNIXには、どれが向いているであろうか？ UNIXでは、前述のようにさまざまなコマンドがあり、プログラム作りから電子メールや文章作りまで幅広い作業が行われる。やはり、文章の自然な入力が可能な、文法処理機能を有した“かな漢字自動変換”が必要であろう。端末がこの機能を実現していれば問題はないが、文法処理や辞書(利用者の単語・外

---

UNIXは、米国 AT&T BELL 研究所が開発し、所有権を有するオペレーティングシステムである。

字登録など)の問題などから“音訓変換”にとどまっているものが多く、UNIXを日本語で使いこなす作業環境には、端末依存はあまり向いているとは言えない。一方、専用ワードプロセッサでの“かな漢字変換”は、べた書き・複文節処理へと高機能化が行われているが、この変換方式には、辞書の持ち方や文節の認識問題を含め、変換率などで残された問題も多く、今のところ、専用日本語エディタ以外での日本語入力には向いていない面が多い。

日本語入出力のシステムでの位置付けとしては、

- ①専用システム
- ②サブシステム
- ③サブルーチン(ライブラリ)
- ④標準入出力

が考えられる。①は、ワードプロセッサ専用機としてOSを表に出さない方式で、②は、ワードプロセッサの機能を専用のサブシステムやエディタなどでのみ実現するものである。いずれも日本語入出力機能が専用のエディタなどに限定される。③は、一歩進んで、日本語入出力機能を特定のサブルーチンで供給しようとするもので、専用エディタ以外のプログラムでも日本語の入出力が可能になるが、やはり日本語入出力機能がそれを意識したプログラムでしか実現できない。④は、更に進んで、日本語入出力機能をOSの標準入出力機能の中に組み込もうとするもので、幅広いプログラムでの日本語入出力が可能となる。UNIXへの親和性を考えると、①②は、システム切り替えやファイル変換が必要となる点と、専用の日本語エディタなどを経由しない日本語入出力が不可能となる点で、また③は、UNIXに既存の豊富なプログラムは勿論、ライブラリを意識的に使い分けていないプログラムでの日本語入出力が不可能となる点で、不向きである。

以上の観点から、UNIXには、標準入出力方式の文法処理機能を持つ“かな漢字変換”が最も親和性があると言って良い。また、ユーザ独自の辞書(新語登録)や外字への対応も、日本語入出力の問題として重要である。

## 2.2 日本語(漢字)コードの問題

次に、UNIXへの親和性を考える上で重要な問題として、日本語のコード表現の問題がある。UNIXの世界は、基本的には7ビットASCIIの世界である。UNIXの核(特に、端末ドライバ)、Shell(B-Shell/C-Shellいずれも)、その他のUNIXの主要なコマンド(特に、エディタや清書プログラム)では、暗黙にその制約(1文字が7ビットASCIIであること)を前提として、作成されている。一方、日本でのコード表現としては、ASCIIに対応して、半角の英数文字とカタカナを制定したJIS英数コード(JISC6220)と、全角の英数文字やカタカナ/ひらがな/漢字を網羅したJIS漢字コード(JISC6226)があり(以後、前者を英数コード、後者を漢字コードという)、いずれも7ビット/8ビット系での規格が定められている。

UNIXに日本語を馴染ませるには、このコード混在の問題を解決する必要がある。それには、1)従来のUNIXに出来るだけ手を加えないで、2)JISコードとの整合性を保ちながら、3)一般のプログラムでも容易に対応でき、4)今後のUNIXの改訂や日本語文書交換/日本語テレテックスなどの標準化に柔軟に対応できることなどを考慮する必要があるだろう。

解決案としては、

- ① J I Sコード(6 2 2 0と6 2 2 6混在)で表現
- ② シフトJ I Sコード(6 2 2 0とシフト漢字コード混在)で表現
- ③ すべてを2バイトコードで表現

などが考えられる。①は、英数コードと漢字コードとの切り替えに特殊なエスケープシーケンス(J I S C 6 2 2 8に規定されている)を必要とするが、②や③ではその必要はない。プログラムを作る立場からすれば、③が適しているかもしれないがメモリ/ファイル容量や1),4)の点で難点がある。②は、3)や一部の1)(ライブラリや言語での文字列処理)の観点から有力ではあるが、U N I X核/Shell/各コマンドの改造と4)の観点での問題が残る。①は、同じくU N I Xのある程度の改造は必要だが、②ほど大作業にはならない。

O S /U Xでは、1)~4)の観点を総合して、①のJ I Sコード方式を採用した。なお、J I Sコード下での英数コードと漢字コードとの切り替えは、3)を考慮して、J I S C 6 2 2 8規格の3バイトのエスケープシーケンスを2バイトで表現し、実際の端末などとの入出力の際にO S /U X核で変換する方式とした。

### 2.3 日本語処理プログラムの問題

さらに、U N I Xを日本語化する上で重要なのは、日本語特有の処理ができるプログラムをいかに充実させるかということであろう。日本語入出力や漢字コードの問題を解決しても、U N I Xオリジナルなコマンドだけでは、もともと英語専用のものだけに、日本語特有の処理ができないからである。ただ、日本語入出力システムの機能や漢字コードの選択によってだけでも、U N I Xオリジナルなコマンドが、ほとんどそのまま日本語版として通用することがあることも注意して置く必要がある。U X - 3 0 0の電子メール(mail(1), write(1))などは、その例である。

U N I Xオリジナルなコマンドだけでは、不十分なものの代表として、1) エディタ、2) 清書プログラム、3) その他日本語入出力や日本語処理環境を整備するコマンド、などが挙げられる。1)、2)は特に重要であり、また3)以外でも日本特有のプログラムの開発が、今後ますます重要になって来るものと思われる。

なお、ここで、日本語エディタや日本語清書プログラム作りには、日本語入出力システムや漢字コードの選択の問題が大きく影響することを改めて注意して置きたい。

### 3. U X - 3 0 0のローマ字・かな漢字変換

U X - 3 0 0の日本語入出力は、O S標準入出力方式を採用した。また、かな漢字変換としては、文法処理機能を持った漢字指定方式のかな漢字変換を採用している。漢字指定方式としたのは、入力を受けるプログラムが日本語専用エディタのような同音異義語の選択や文節識別能力を持つことを前提とできないことや、バタ書き下での余計な文法処理が、かえって自然な入力の妨げになり得ることを考慮したためである。次に述べるように、同音異義語の選択操作もO S側で行っている。ユーザプログラムに余計な選択操作を強いないようにするためである。

ところでUNIXの端末入出力の特徴には、次の3つがあり、

- 1) 全二重 I/O,
- 2) data-driven I/O,
- 3) stream I/O

一方日本語、特に漢字の入力の特徴として、入力過程に、

- ア) 辞書の参照,
- イ) 同音異義語などの選択操作

が介在することが挙げられる。このため、UX-300ではOS核と一体になって動作する漢字デモンプロセッサが存在する。一般の非漢字端末からの入力は、そのままユーザプログラムに渡されるが、漢字端末からの入力は、一度この漢字デモンプロセッサで受け、漢字への変換とその同音異義語などの選択処理が終了したものがユーザプログラムに渡される。(UNIX流に言うところ、この漢字デモンはキーボード入力の画面表示、かな漢字変換、候補漢字の選択を行う“フィルタ”群から成り、各々独立したモジュール構造を持っていて、プログラム部品として独立している。端末特性に依存するのは、表示部と選択操作部に局所化されている。)

このように、UX-300の日本語入力方式は、全体として、OS標準入力での“逐次変換・逐次選択方式”の“文法処理機能付き漢字指定方式”のかな漢字変換となっている。この方式により、UX-300上のソフトは、統一されたインターフェースで容易に、文章やデータとしての幅広い日本語入出力を行うことができる。以下、[1]に詳しく報告したのでここでは簡単にその機能を説明する。

### 3.1 かな漢字変換

#### (1) かな漢字自動変換

漢字の読みをそのままひらがなで入力し、送りがななどから文法チェックを行って、対応する漢字に変換する“漢字部分指定のかな漢字変換”である。単一名詞変換の他に、送りがな・助詞・助動詞の認識機能、接辞・数詞・漢数字変換機能を持っている。

#### (2) 固有名詞変換

日本語入力では、地名・人名・会社名などの入力も重要である。これらの固有名詞をその読みから漢字に変換する。

#### (3) 一字変換

上記の変換で変換できないような漢字一字一字について、その音読みや訓読みから漢字に変換する。(音訓変換)

#### (4) 区点変換

更に上記の変換では変換できない特殊記号などを、JISコードの区点番号や16進コードで直接指定する。部分的に分からない桁を「?」で指定し、表示を見ながら入力する“選択方式”も可能である。

#### (5) 学習機能

同音異義語の表示は、辞書に設定された「頻度」順に行われるが、使用者ごとの違いに対して、使用者(端末)ごとに暫定辞書を設け、一度選択された漢字を最優先に表示する学習機能がある。

### 3.2 ローマ字変換

英数配列のキーだけで漢字入力を可能とするローマ字→かな漢字変換がある。その特徴には、次のものがある。

- (1) 綴り方は、訓令式+日本式+ヘボン式+ $\alpha$ 。
- (2) 音節単位のローマ字入力・ひらがな/カタカナ表示。
- (3) 小文字/大文字のローマ字をひらがな/カタカナに変換。
- (4) 英数配列キーによる漢字変換制御キーの代行。
- (5) 前記のかな漢字変換のすべてに対して、ローマ字変換が可能。
- (6) 特殊記号の拡張ローマ字変換(非ローマ字の子音綴りを漢字に変換)。
- (7) 端末ごとに動的に、ユーザ定義可能な拡張ローマ字変換。

### 3.3 漢字ヒストリ

文章なりデータなり、日本語入力を行う場合、何度も同じ漢字を入力することがある。漢字の入力は、(ローマ字→)かな→漢字変換→選択操作が必要なので、厄介である。同音異義語の表示に頻度処理や学習機能があり、一方“新語機能”により、短い読みで辞書に登録して置くことはできるが、毎回漢字変換と選択操作を行っていたのでは効率が悪い。

このため、使用者ごとに交換・選択された漢字を複合語単位で記憶して置き、その再入力は、ファンクションキーを用いて一気に漢字そのものを入力する“漢字ヒストリ機能”がある。これは、ローマ字/ひらがな入力や漢字変換、選択操作を必要としないので、非常に効率的な入力ができる。

### 3.4 ユーザ漢字キー

使用頻度の高い漢字(文字列)や定型句、会社名を効率良く入力するには、前述の漢字ヒストリ機能や新語機能(辞書登録)があるが、更に効率良く特定の漢字列を入力できるよう、ユーザ定義可能な漢字キー機能がある。

これは、使用者(端末)ごとにユーザ指定の漢字列を特定のファンクションキーに任意に割り付けるもので、そのファンクションキーの押下により設定された漢字列を直接入力することができる。その設定や切り替えも自由である。

### 3.5 辞書管理と新語機能

かな漢字変換で使用する辞書には、一般の国語辞典に相当する一般語辞書と、地名人名などの固有名詞を収録した固有名詞辞書、更に漢和辞典の音訓表に相当する一字変換辞書がある。このうち、一般語辞書と固有名詞辞書は1つのファイルにまとめられており、日本語辞書と総称する。システムには、1つの標準日本語辞書(/etc/j.dict)がある。標準日本語辞書の登録されている語(約3万語)には限りがあり、使い方や分野によっては、すべてが目的の語に変換されるとはかぎらない。

このため、使用者の使用目的に合わせて、単語を追加・削除して個人用の辞書を作ることができるよう、辞書エディタ(shingo)をサポートしている。新語(一般名詞扱い)の他、システム登録の一般名詞や活用語の追加・削除も必要に応じてできるようになっている。

作成した日本語辞書は、漢字端末ごとに自由に切り替えることができる。

### 3.6 外字機能

JIS C 6226にないユーザ独自の特殊な漢字やマークを、外字として作成し、自由に登録して使用することができる。全システム共通な外字パターンファイル以外にも、端末ごとの外字パターンファイルを自由に設定したり、また全くローカルに個人用外字パターンファイルを持つことができる。

## 4. おわりに

以上、日本語UNIX:UX-300シリーズOS/UXについて、日本語化実現方式やその機能から報告した。日本語入力の方式や操作性に様々な工夫を行い、OS標準機能として日本語処理を考えているのが大きな特徴である。この方式により、UNIXに非常に親和性の良い日本語処理が実現されたと思っている。

今後共、この方式や漢字コードなどの日本語UNIX標準化について検討を加え、さらに日本語UNIXをベースとして日本特有な機能を実現して行きたい。

### 【参考文献】

- [1] 中原：「日本語UNIXでの日本語入力方式と新機能」, 情報処理学会日本文入力方式研究会14-2, 1984
- [2] 中原：「UX-300での日本語処理」, 情報処理学会第27回全国大会, 1983
- [3] 中原, 谷内：「UNIXの漢字テキスト」, 情報処理学会第26回全国大会, 1983
- [4] 三浦, 鈴木：「UNIXの日本語文書清書プログラム」, 情報処理学会第26回全国大会, 1983
- [5] 中原, 山田, 小倉：「日本語UNIX-日本語sh, csh」, 情報処理学会第28回全国大会, 1984
- [6] 山田, 中原, 谷内：「日本語UNIX-日本語スクリーン」, 情報処理学会第28回全国大会, 1984