

5 Q-2

OS/omicon 仮名漢字変換システム第2版の設計と実現

本宮志江, 酒井貴子, 下村秀樹, 並木美太郎, 高橋延匡
(東京農工大学 工学部 電子情報工学科)

1. はじめに

当研究室では、日本語の使用を前提として、全2バイトコードを採用したオペレーティングシステム、OS/omicon 第2版が実現されている。現在、OS/omicon 第2版上には、日本語アプリケーションとして、言語CコンパイラCAT、プログラムエディタAltheaなどが稼働しており、さらに、日本語文書の構造を考慮した、論文・研究資料作成用ワープロを開発中である。ここで問題として残るのが、日本語入力環境である。OS/omiconの特徴を十分に活かすために、日本語の標準的な入力手段として、仮名漢字変換は必要不可欠である。しかし、現在用いている仮名漢字変換システム [1] は、学習機能がない、連文節変換がないなど機能的にも不十分であり、また変換精度も悪い。

以上のようなことが背景にあり、我々は、OS/omicon 仮名漢字変換システム第2版を開発することになった。

本稿では、OS/omicon 仮名漢字変換システム第2版の設計と実現について述べる。

2. 仮名漢字変換システム第2版の開発目標

仮名漢字変換システム第2版(以降、本システムとする)では、次に述べることを目標としている。

- (1) 文字種の違いを意識する必要のない、モードレス入力を可能にする。
- (2) ユーザや作業に応じた仮名漢字変換機能を提供する。
- (3) 日本語解析などのアプリケーション(以降、APとする)に、仮名漢字変換システムの持つ辞書を開放する。
- (4) 変換操作に関するキー割当や変換結果の表示などのインターフェース部分を構築するAPに、設計の自由度を与える。

これらの目標を持つに至った背景の詳細については、[2] で述べているので参照されたい。

3. 仮名漢字変換システム第2版の全体構成と特徴

本システムは、SVC解釈部、資源管理部、環境部、文字列変換部、辞書部の五つのモジュールからなる(図1)。

次に、本システムの特徴である、文字列変換部、環境部および辞書部について、個々に説明する。

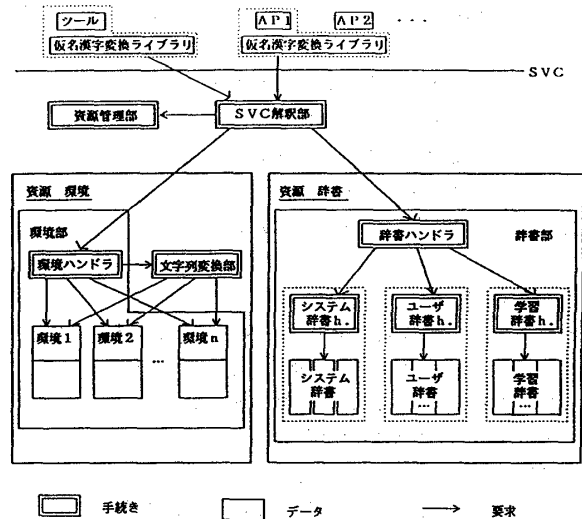


図1 仮名漢字変換システム第2版の全体構成

3.1 環境部

『環境』は、

- (1) 変換辞書情報領域

変換に使用する辞書名(複数可)および辞書の優先順位などの、変換に必要な情報を管理する領域

- (2) 変換作業領域

入力文字列の解析作業を行い、解析結果を保持しておく領域

からなる。

上記の2つの領域は、各環境ごとに用意される。この環境により、それぞれの作業内容や目的に応じた辞書を用いて、独自の変換を行うことができる。環境は、複数のAPで共有することができ、さらに、一つのAPが、複数の環境を用いて変換を行うことも可能である。各APは、変換に先立って、環境を設定する必要がある。

3.2 文字列変換部

本システムでは、次の4種類の変換を用意する。

- (1) ローマ字片仮名変換・ローマ字平仮名変換

- (2) 平仮名片仮名変換・片仮名平仮名変換
 (3) 単漢字変換
 (4) 日本語変換

日本語変換は、入力文字列を、普通の日本語の文に現れる、“漢字仮名混じり文字列”に変換する機能である。一般に言う、連文節変換とほぼ同じである。

仮名漢字変換システムとしては、この機能が最も重要である。内部処理は、大まかには次のようになる。

APからの変換要求を受けて、文字列変換部が、

- a) 環境部に書かれた辞書の情報を参照し、
- b) 辞書部にその辞書を引きにいき、
- c) 環境部に変換結果文字列を作成し、

APに、変換結果文字列を返す。

次に、日本語変換について説明する。

日本語変換の特徴

本システムの日本語変換では、現在の市販の仮名漢字変換の機能より、さらに高機能な変換を用意することを目標としている。次に、その特徴を挙げる。

- (1) 入力文字列を仮名に限定しない。

この日本語変換では、

sisutemuno, systemの, しすてむの, システムノ
 など、どのような入力でも、

システムの

と変換できるようにする。

内部処理としては、辞書の検索を、a) 入力のままの文字列、b) ローマ字仮名変換した文字列、の2通りで行うことによって実現する。この方法の利点は、英数字と平仮名、漢字が混在するものを頻繁に入力するような場合に、モード切り替えの手間がいらぬことである。

- (2) 常に変換結果全体を考えて、最もよいと思われる候補を返す。

本システムでは、

仮名漢字変換は、入力文字列を解析して変換結果の木を作り、それを探索する問題に帰着する。
 と捉えている。

したがって、我々の方法では、変換結果の木を用いて、常に全体を考えて、最もよい変換結果を返すことに重点をおく(図2)。エンドユーザからみた操作を例にとり、説明すると次のようになる。

a) 次候補キーで先頭の単語を変えた場合、それにもなって、後ろの単語列も、最もよいと思われるものになる。

b) 先頭単語に対応する入力文字列の長さも、次候補キーを押すだけで変わる。つまり、次候補を押していれば、必ずほしい単語が得られる。

c) カーソルキーで変換範囲を変えるような指定は必要ない。

この方法の利点は、文節の切り直しの操作が必要ない

入力: このことは

変換:

```

<先頭>--- 子 --- の ---こと --- は
      |           |--- 子 --- と --- は
      |           |--- 琴 --- は
      |
      |---この---こと--- は
           |--- 子 --- と --- は
           |--- 琴 --- は
  
```

図2 変換結果の木

ことである。変換処理が正確に行われることを仮定した場合、変換に必要なキー操作数が少なくてすむ[3]。

3. 3 辞書部

辞書部の役割は、次の2点である。

- (1) 仮名漢字変換に利用する辞書の管理
 - (2) 辞書レコードアクセス(読出し、書込み)関数の提供
- つまり、辞書部は、これらの機能を、仮名漢字変換内部の文字列変換部にだけでなく、辞書メンテナンスツールや辞書情報を必要とするAPに提供する。

変換に利用する辞書には、

- ・システム辞書(自立語、付属語、単漢字)
- ・ユーザ辞書
- ・学習辞書

がある。この辞書部は、これらの辞書をすべて同一のインタフェースでアクセスできる機能を提供するとともに、辞書の管理を行う。

4. おわりに

本稿では、OS/omicon 仮名漢字変換システム第2版の設計と実現について述べた。仮名漢字変換システム第2版の実現後は、すべてのAPに標準的な入力環境を提供できるように、日本語入力フロントエンドプロセッサを構築する予定である。

参考文献

- [1] 下村秀樹, 並木美太郎, 高橋延匡: “OS/omicon 仮名漢字変換システムとユーザ辞書”, 情報処理学会第38回全国大会講演論文集4N-7, 1989
- [2] 下村秀樹, 本宮志江, 酒井貴子, 並木美太郎, 高橋延匡: “OS/omicon 仮名漢字変換システム第2版の設計思想”, 情報処理学会第42回全国大会講演論文集(予定), 1991
- [3] 本宮志江, 酒井貴子, 下村秀樹, 高橋延匡: “日本語ワードプロセッサの仮名漢字変換の解析と評価”, 情報処理学会ヒューマンインタフェース研究会報告33-2, 1990