

ワープロ用英字切り出し処理付き 入力モード自動変更機能の開発

1W-5

隈井 裕之 富田 ひろみ 中島 晃 板部 征典 柏 博文

(株)日立製作所 マイクロエレクトロニクス機器開発研究所 (株)日立製作所 情報映像メディア事業部

1. はじめに

ワープロのキーボード入力では、かな/ローマ字入力方式の切り替えや多数の字種入力のために、入力モードの変更操作が必須である。しかし、これを誤った場合、ユーザは入力モードを設定し直して再入力することを余儀なくされていた。これに対処するため、我々は、入力文字列全体の入力モードを判定して自動変更する“入力モード自動変更機能”を先に開発した[1]。これは、単一字種の入力文字列に対して有効であったが、今回、さらに、複数字種の入力の中で最も入力モードの戻し忘れの頻度の高い、英数からひらがなモードへの戻し忘れに対処する“英字切り出し処理を有する入力モード自動変更機能”の試作・評価を行い、実用性を確認した。

2. システムの概要

日本語文には漢字、ひらがな、カタカナだけでなく、多くの英字が使用されている。ワープロで作成されるこれらの英字を入力するために、ユーザは、キーボードに用意されているモード切り替えキーを用いて、かなと英字を打ち分けている。

図1(a)に英字が混在した日本語文の入力例を示す。図中 **英数** **ひらがな** は、モード切り替えキーを表している。今回、我々は図1(a)に示すような文例に対して、ユーザが「ひらがな」キーを押し忘れた場合を対象に、モード設定誤りの救済手段を開発した。すなわち、ユーザが図1(b)のように「ひらがな」キーを押し忘れて、そのまま英字を入力し続けても、自動的に英字を正しく切り出すとともに、残りの文字と入力モードをユーザの意図通りに切り替えてくれるシステムである。

表1は、日本語文中で使用される英字について、我々が文例や文献[2][3]を参考に調査・分類した結果である。一般に日本語文中で使用されている英字は殆どが、略語や単位の類である。そこで、今回は、英文や英単語ではなく、略語や単位など(表1 No1-4, 以降英略語と呼ぶ)を英字切り出しの対象とした。切り出し処理には、英略語を集めた辞書を用い、更に単位などの数字を伴う英略語に対処するために数字の切り出し処理を加えている。また、英字に後続する入力モードは、ローマ字モードとかなモードを対象とし、どちらのモードで入力されても正しく判定できるようにしている。

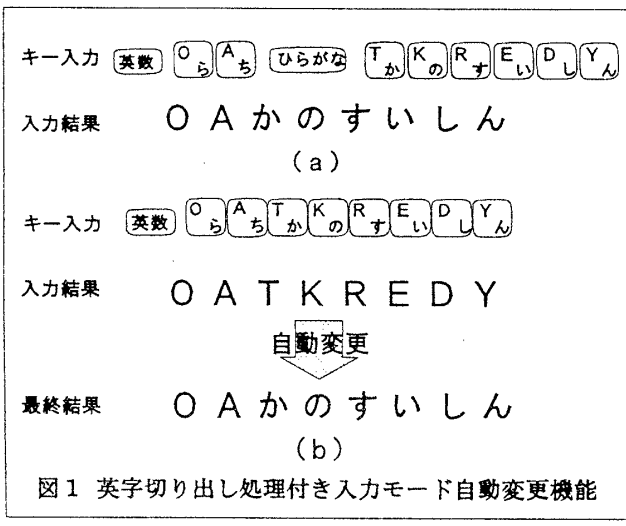


図1 英字切り出し処理付き入力モード自動変更機能

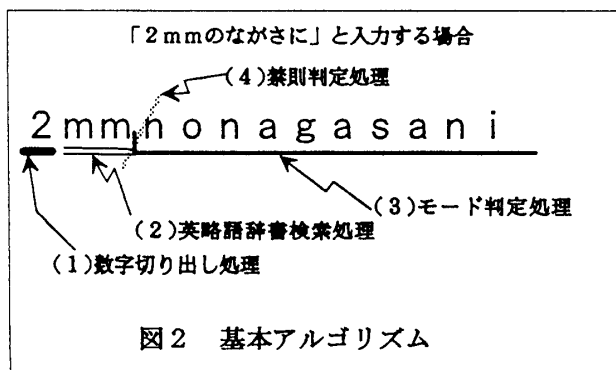
Method of Finding English Letters from Key Input in Undefined Mode for Japanese Wordprocessor.
Hiroyuki Kumai, Hiromi Tomita, Akira Nakajima, Masanori Itabe, Hiromi Kashiwa.
Hitachi, Ltd.

3. 英字切り出しアルゴリズム

英字切り出しの精度が悪いと、かえってユーザの操作性を低下させることにもなりかねない。切り出しアルゴリズムは精度に主眼をおいて開発し、図2に示すように4種類の処理によって構成している。

表1 日本語文中の英字の性質

No.	英字の種類	英字の例	特徴・性質
1	略称(頭文字)	BGM, OK	大文字2~4文字で構成
2	英単語の省略	etc., cf.	小文字2~3文字で構成
3	英字+記号	QA	英単語の略字と記号
4	単位, 標識等	10cm, R246	数字と単位, 標識
5	英字+かな	Tシャツ, ビタミン	777777 ット+かな
6	日本語化英単語	OPENする	サ変名詞化, 形容動詞化
7	英文	A happy day	英単語の羅列



(1) 数字切り出し処理

先にも述べたように英略語の中には、単位など数字とともに使用されるものがある。これに対処するため英略語の前後に現れる数字を切り出す。ただし、英略語辞書検索の結果、得られた英略語が数字を伴わない単語である場合には、数字の切り出しは行わない。

(2) 英略語辞書検索処理

入力されたキーコードを検索子として、英略語辞書を検索する。辞書に格納された各英略語毎に、数字接続の可否を示すフラグを付加し、「45BGM」「OA7」などの誤った切り出しを防いでいる。また、「OA(オフィスオートメーション)」と「OAD(政府開発援助)」のように2以上の英略語が候補になる場合には、禁則判定処理や数字の接続において整合性がとれるもので最長の英略語を優先する。

(3) モード判定処理

先に報告した入力モード自動変更機能の累積頻度判定方式を用いて、英字を切り出した残りの部分について入力モードを判定する。この処理は、一見英字切り出し処理とは無関係に見えるが、ここで判定したモードの情報は次の禁則判定処理で使用する。

(4) 禁則判定処理

英字を切り出した残りの部分について、その区切り目が日本語の並びとして正しいか否かを判定する。例えば、区切り目が拗音(ゃ, ゅ, ょ)や促音(っ)、「ん」で始まる場合には、区切り目としては正しくない。このような禁則にあてはまる場合には、英字切り出しを無効とし、次の英略語候補の処理に移る。

このとき、ローマ字モードとかなモードでは、キーの解釈が異なるため、(3)の判定結果を用いて全体として整合性が取れるようにしている(例えば、「NN」はローマ字モードでは「ん」となるが、かなモードでは「みみ」である)。

以上4つの処理の組み合わせによって、英字とひらがなの区切り目を精度良く検出することができる。

表2 評価結果

項目	ローマ字	かな
正しく切り出された文例数	64	65
正しく切り出されなかった文例数	切り出し位置誤り	0
	モード判定誤り	1
	辞書に英略語なし	35
合計	100	100

4. 試作と評価

以上述べたシステムを試作し評価を行った。評価は英字で始まるかな・ローマ字の文例各100文に対し英字が正しく切り出される文例数を調べた。英略語辞書は語数約370語のものを用いた。表2に、評価の結果を示す。この結果から、辞書に対象となる英略語が存在すれば、ほとんど正しく切り出されることがわかる。辞書に存在しなかった英略語の殆どは、商品名や型番など固有の英略語であった。

また、別に全体がかな又はローマ字の文例各100文に対して、誤って英字切り出しが行われる文例の有無を調査した結果、かな、ローマ字共に1文例のみであった。すなわち、従来の入力モード自動変更機能に対する影響は最小限に押さえることができた。

以上のように、今回開発した英字切り出し処理のアルゴリズムは、区切り位置の誤りが少ないことから、十分実用的であると考えられる。

5. 結論

英数からひらがなへの入力モードの戻し忘れを対象とし、英略語辞書によって英字を正しく切り出し、後続するひらがなを正しいモードに変更する「英字切り出し処理付き入力モード自動変更機能」を開発した。評価の結果、対象の英略語が辞書に搭載されていれば、ほぼ完全に英字を切り出せることが判った。今後の課題としては、(1)英略語辞書語数の増加、(2)大量文例での評価、(3)ユーザ固有の英略語に対応できる方式の開発である。

参考文献

- [1]中島、他：「ワードプロセッサにおける入力モード自動変更方式」、情報処理学会第43回全国大会
- [2]「カタカナ語・略語辞典」、旺文社
- [3]小泉袈裟勝監修：「単位の辞典」、ラテイス