

2S-8

かな漢字変換における  
構文意味解析速度の向上

限井 裕之<sup>†</sup>                      中島 晃<sup>†</sup>                      柏 博文<sup>††</sup>                      谷口 茂樹<sup>††</sup>  
<sup>†</sup>(株)日立製作所                      <sup>††</sup>(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                      多賀工場

1. はじめに

パーソナルワープロの日本語入力は、ベタ書き入力のかな漢字変換方式が主流となってきた。また、同音異義語の誤り、文節区切り位置の誤りなどの誤変換(多義)の問題に関しても、用例辞書等を用いて解消し、変換精度を向上する試みが行われるようになった。しかし、従来の方法の多くは、単純に語句の組合せを記憶することで行われていたが、対象語句の増加につれて辞書容量が級数的に増える問題がある。

我々は、この問題に対して、格文法による構文意味解析手法に着目し、パーソナルワープロへの適用を考えた。構文意味解析手法では、語順の変化や、新たな語句の登録への対応を辞書容量を増加させることなく行うことができる。しかし、従来の構文意味解析を用いたシステムは処理量が多く、殆どが大型機上で稼働するものであり、パーソナルワープロに使用されているマイクロプロセッサには負担が重く、そのままでは適用することはできない。処理量を軽減しなおかつ多義解消に効果のある方法の開発が望まれている。

本報告では、構文意味解析手法のパーソナルワープロへの適用について我々が検討、試作したシステムとその評価結果について述べる。

表1 構文意味解析かな漢字変換システムの概要

項目	スペック
CPU	16ビット8MHz
自立語辞書数	約60,000語
格フレーム辞書	約700用言
意味分類辞書	約5,000名詞
構文意味解析方式	表層格および深層格を用いた格引き当て方式
処理可能範囲	かな読み20文字まで

2. システムの概要

本システムのかな漢字変換方式は文献1)、2)に示される形態素解析法、構文意味解析法、最適文選択法を基本として、実用範囲内の制限を設けパーソナルワープロ向けに処理量を軽減したものである。本システムは、入力された20文字までのかな読みに対して形態素解析木を作成した後、順次候補列(パス)を抽出し、単文に分割処理した後、これを格フレーム辞書と意味辞書を用いて構文意味解析処理を行う。各パスには、構文意味解析の結果によって評価が与えられ、一定以上の評価を得たパスを変換結果として出力するものである。表1に本システムの概略スペックを示す。

以下今回の開発の要点について述べる。

(1) 構文意味解析用辞書

従来の構文意味解析方式は、入力された文章が、文法的に完全であるとの仮定のもとに開発されてきた。例えば、図1に示す文章では、用言に対し主格や目的格がより重要な格(必須格)であり、共起関係は主にこれらの格に依存するとされてきた。しかし、パーソナルワープロユーザの入力する文章の中には、主格の省略表現や、訂正追加等による部分的な入力も多く、これらに対応する必要がある。図1の例における手段格などは、文法的には存否を問われない格(自由格)であるが、用言と共起関係を持つ。

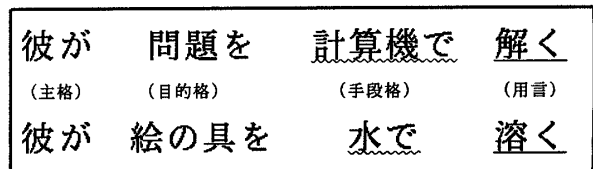


図1 構文意味解析における共起関係の一例

本システムでは、この観点に立ち、このような用例を収集、分析して作成したパーソナルワープロに適したコンパクトで、精度のよい構文意味解析用辞書を搭載している。

(2) 応答性の確保

パーソナルワープロでは、かな漢字変換の応答性は重要な問題となる。大型機上のシステムのように、すべてのパスを探索すれば、正しい変換を行う可能性は大きくなるが、応答性の低下が顕著となる。

本システムでは、従来の形態素解析と尤度情報によるかな漢字変換精度がある程度得られており(90%以上)、形態素登録順が日本語のヒューリスティックルールに則り正解率の高いものから登録されることを考慮に入れ、パス抽出回数を制限する方法で応答性を確保している。すなわち、まず従来の方法で得られたパスを構文意味解析処理によって評価し、以降制限回数に達するまで、形態素解析木を縦型探索することによってパスを抽出し構文意味解析処理をおこなう。

3. 評価

本方式を評価するために、変換精度及び速度について調査し、従来の形態素解析のみによるものと比較した。

変換精度は学会誌、論文集、ビジネス文書約2万9千文字による調査では、変換精度は91.4%で0.4%の上昇であったが、文節区切り位置の誤り件数は従来比で14%低減した(表2)。また、同音異義語の誤り件数が増加したが、これは、構文解析により文節区切り位置が正しくなったものの、意味解析が不十分なケースがあったためである。

変換速度は約130文例について構文意味解析により正解候補が得られるまでのパスサーチ回数及び処理時間を調査した(図2)。この結果パス抽出制限回数15前後で、処理

表2 要因別誤変換件数

誤変換分類区分	同音同義語	同音異義語	同音異品詞	文節区切り	合計
構文意味解析有(a)	417	241	106	177	941
意味解析無(b)	435	232	108	206	981
低減率 <sup>*)</sup>	4.1	▲3.9	1.9	14.1	4.1

▲は増加を示す

<sup>\*)</sup> 低減率 =  $\frac{(b) - (a)}{(b)} \times 100 (\%)$

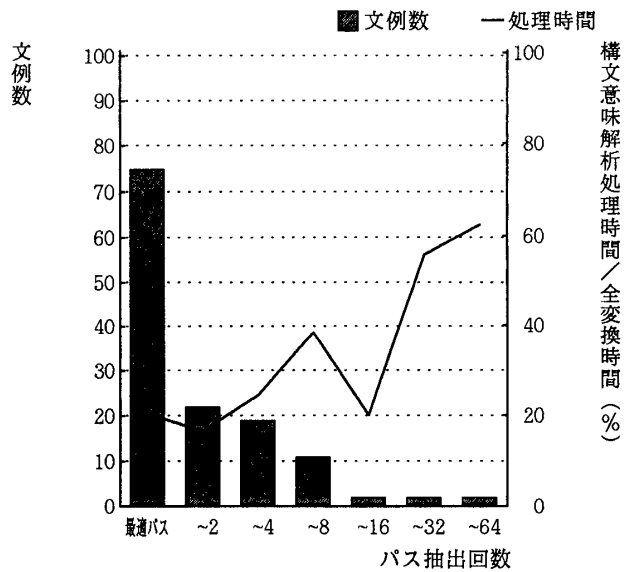


図2 パス抽出回数及び変換時間測定結果

時間を従来の2倍以下に抑えられる見通しを得た。これは、我々のシステムで1秒前後に相当する。また、このパスサーチ回数制限により構文意味解析が適用できなくなる文例の割合は3%であった。

4. おわりに

本方式により、処理量を許容範囲に抑えるための制限を設け速度を向上させることで、マイクロプロセッサを使用したシステムにおいても構文意味解析を用いたかな漢字変換を実現できること、また、変換誤り、特に文節区切り位置の誤りに対して効果の有ることを確認できた。

しかし、かな読み文字数、パスサーチ回数等適用範囲を狭めた結果、新たな誤変換を引き起こすこともあった。今後はハードウェアの進歩にあわせて、適用範囲を拡大し、さらに変換精度を向上していきたい。

5. 参考文献

- 1) 阿部他: 「べた書き文仮名漢字変換における最適文選択法」他、情報処理学会第28回全国大会講演論文集、4M-5~8 (1984)
- 2) 大島他: 「格文法による仮名漢字変換の多義解消」、情報処理学会論文誌、vol.27 No.7 pp 679 (1986)
- 3) 長尾他: 「日本語情報処理」、電子通信学会、コロナ社(1984)