

## ピン音表記を用いた文法解析型複数 文節中国語文章入力システム

3 Q-4

河野 勝也 小滝 房枝 隈井 裕之

(株)日立製作所 中央研究所

### 1. はじめに

#### 1. はじめに

従来、中国語入力方式では、四角号碼法、五筆法など漢字を分解してコード化して入力する方法が多く提案されている[1]。

このような入力方法は、相当の学習を必要とするため、専門オペレータが原稿を見ながら入力するには適しているが、文章を考えながら入力する場合には、思考の妨げとなる。

そこで、ユーザの思考を妨げない入力には読みからの入力が最適と考え、中国語の英字表記であるピン音を用いた複数文節中国語文章入力システムを開発した。本システムはパーソナルコンピュータ上のフロントエンドプロセッサとして稼動する。

### 2. システムの概要

ピン音は発音表記に近く、従来のコードや字画、部首を用いた入力方法に比べ、初心者でも容易に効率的に入力できる。

本システムは、図1に示すように中国語を文単位に読みで入力し、中国語漢字文字列に変換するものである。中国語用語辞書は約5万5千語の用語を持ち、文法解析に必要な品詞[2]情報を持つ。

読みを入力して漢字変換するときの技術課題は日本語のかな漢字変換と同様に同音語の中から、いかに適切な漢字を選択するかにある。

この問題を解決するためには、①文法知識の応用による漢字の特定化、②漢字の使用頻度情報の

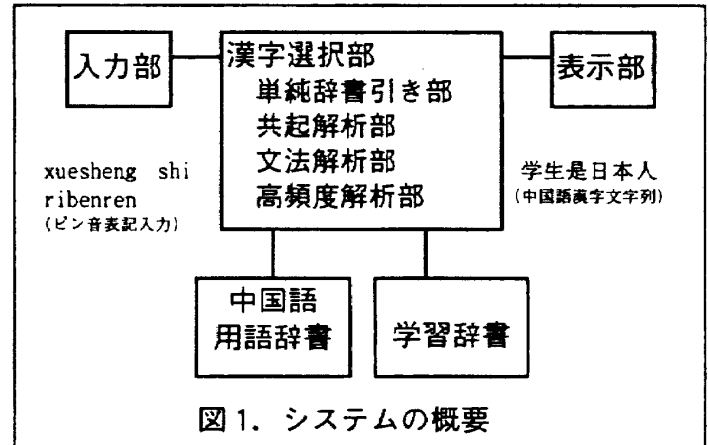


図1. システムの概要

利用、が有効である。

### 3. 漢字選択の方法

以下、同音語を高精度に選択する漢字選択部の動きについて説明する。

キーボードから中国語の読みが標準ピン音表記で文単位に入力されると、以下の手順で処理する。

- ①単純辞書引き部：単純に辞書検索を行って、読みに同音語がなければ漢字を決定。
- ②共起解析部：共に現れる可能性の高い読みが現れれば漢字を決定。
- ③文法解析部：単語の位置および両隣の単語の品詞の並びの特性を、文法ルールとして定義し、文法ルールに一致したとき、品詞または漢字を決定。
- ④高頻度解析部：上記で決定できない場合、同音語中最も使用頻度の高い漢字で決定。

変換経過の例を図2に示す。

- ① XUESHENG には同音語はなく辞書引きだけで「学生」に決定する。
- ② BEN には9個、SHU には46個、同音語があるが BEN SHU が同時に現れたときには、共起解析部において量詞と名詞の組合せである「本」「書」の可能性が高いと判定し漢字を決定する。

Chinese Text Input System.

Katsuya KOONO, Fusae ODAKI,

Hiroyuki KUMAI

Hitachi, Ltd., Central Research Laboratory

中国語：学生买一本书。（学生は一冊の本を買う。）

- ピン音入力： XUESHENG MAI YI BEN SHU.  
 ①単純辞書引き： 学生 MAI YI BEN SHU。  
 ②共起解析： 学生 MAI YI 本 书。  
 ③文法解析： 学生 MAI 一 本 书。  
 ④高頻度解析： 学生 买 一 本 书。

図2 変換経過の例

③ YI には 110 個同音語があるが、「本」が量詞で決定されているため、量詞の前には数詞が来るという文法知識（ルール）を用いて、「一」に決定する。

④上記で決定できなかった MAI には 9 個の同音語があるが、辞書に格納されている頻度情報を参照し、最も使用頻度の高い漢字で決定する。

#### 4. 文法ルール

文法解析で用いられる文法ルールには、中国語の文法知識に基づいて、以下のようなルールを設けている。

##### (1) 文頭・文末解析ルール

文頭、文末の位置情報を用いて解析する。ルールの例を図3 R1 から R3 に示す。

- R1: 文頭で「,」が続くとき、文頭は感嘆詞あるいは助詞とする。  
 R2: 文頭で「,」が続かないとき、文頭の品詞に代詞があれば代詞とする。代詞がなく介詞があれば介詞とする。  
 R3: 文末の読みの品詞に助詞があれば助詞とする。

- R4: 読みが DE で前が名詞または代詞のとき「的」で決定  
 R5: 読みが DE で後が名詞または「。」のとき「的」で決定  
 R6: 読みが DE で後が動詞で前が形容詞のとき「地」で決定  
 R7: 読みが DE で後が動詞で前が形容詞でないとき「得」で決定

- R8: 前が量詞なら後は名詞  
 R9: 後が量詞なら前は数詞  
 R10: 後が動詞なら前は副詞、形容詞  
 R11: 前が助動詞なら後は動詞、形容詞  
 R12: 前が介詞なら後は名詞 代詞 数詞

- R13: 同じ読みが重なっているとき 動詞か 形容詞か 名詞  
 R14: 「一」または「了」をはさんで同じ読みが重なっているとき 動詞か 形容詞

図3 文法ルール

##### (2) 漢字決定両となり解析ルール

前後に隣接する単語の品詞や特定の読みなどに注目し、漢字を決定する。一例として、助動詞「DE」の同音漢字「的」「得」「地」の選択し分けのルールを図3の R4 から R7 に示す。

##### (3) 品詞決定両となり解析ルール

前後に隣接する単語の品詞や特定の読み、あるいは既に決定した漢字などに注目し、品詞を決定する。もし、決定した品詞に漢字が複数ある場合には、最終的に高頻度解析で漢字を決める。図3の R8 から R12 に例示する。

##### (4) 読み特性による品詞決定両となり解析ルール

中国語では、同じ漢字を繰り返すことによって意味を添えるものがあり、よく使用される。前後の漢字が決定されていれば、繰り返される漢字の品詞を決定することができる。図3の R13 から R14 に例示する。

#### 5. 評価

以上の解析処理を組込んだ中国語入力システムを開発した。40文の文例による評価では、文法ルールを用いた場合には27文(68%)で正しい結果が得られ、用いなかった場合の23文(58%)に比べ10%の向上が見られた。単語単位では、文法ルール有り92%に対し、文法ルール無しが87%であった。以上の結果から、本システムが、実用に供しうる変換率を実現できる見通しを得た。

#### 6. 今後の課題

より大規模な評価を行い、ルールの適用条件の妥当性を検証し、変換精度を更に向上させることが今後の課題である。

##### [参考文献]

- [1] 陳他：中国語の漢字入力の一方法；情報処理学会第35回全国大会  
 [2] 香坂：現代中国語辞典；光生館