

## べた書き文の仮名漢字変換システムとその同音語処理†

牧野 寛†† 木澤 誠††,†††

べた書き文の仮名漢字変換では、切れ目のない文字列から、いかに正しい語を推定するかが問題となるが、著者らはこの問題を次の二つ、すなわち、入力文の分かち書きと同音異義語の推定の問題に分けて考え、前者の問題に対して、二文節最長一致法、付属語分かち書きという二つの方法を用いることにより、べた書き文の文節単位への分かち書きが効果的に行えることをすでに示した。

本論では、同音語を識別するための情報として、用言に対しては、格文法の考えに基づいた格の情報とそれぞれの格の立ち得る語の意味的範ちゅうの情報、一方、体言に対しては、派生語、複合語および関連語の情報を検討整理して、機械辞書としての形式化を行う。さらに構文解析によって求められた係り受け関係に基づいて同音語を推定する具体的手法を示し、べた書き文の自動仮名漢字変換システムの効率の向上を図る。

実験例によれば、96.2%の文節が正しく変換され、単語の頻度による同音語選択の方法に比して、約7%の漢字変換率(文節単位)の向上が図られた。

### 1. ま え が き

電子計算機の普及発展に伴い、漢字仮名混り文を柔軟に入出力できるシステムが望まれているが、漢字の入力を簡単に行い得る入力方式として、仮名文を入力し、計算機内で自動的に漢字表記文に変換するいわゆる仮名漢字変換による方式が注目されるようになった。

従来提案された仮名漢字変換システムでは、変換が容易に実行できるように、入力文に分かち書きを施したり、字種を示す制御信号を挿入したりする、いわゆる分かち書き入力規則を採用しているものがほとんどであった<sup>1)~6)</sup>。しかしながら、通常の漢字仮名混り表記の日本語では、分かち書きの習慣がないため、入力に際して、分かち書きなどの知的作業の導入は、入力操作員の負担となり、仮名入力方式の本来の利点(タッチメソッドによる高速打鍵)を十分に生かせないなどの欠点があった。筆者らはこれらを打開するための方法として、べた書き文入力形式の採用を検討し、べた書き文の仮名漢字変換を実現するための第一段階として、切れ目情報のない文字列を文の構成単位である文節に分割する処理について報告した<sup>7)</sup>。

本稿では、その第二段階として、文節単位の文字列から同音語を選択する問題を考える。

同音語は表音表記が同一の語組として定義できるが日本語における語の定義そのものが機械処理という観点からは、比較的あいまいであり、また語の活用などによる語形変化が激しいことも手伝って、同音語に関する研究は必ずしも多くない<sup>8)</sup>。

一方仮名漢字変換においても、同音語の問題の解決が重要な課題とされている割には、必ずしも効果的な方法が見い出されていない。その原因は、大量の語を対象とした場合の語い情報(意味・用法までを含めた語の情報)をどのように記述し、利用するかについての研究が行われていなかったことにある。したがって以下では、語の用法上の差異から同音語を同定、認識するための語い情報の整理と辞書形式について論じるとともに、その辞書および一文の構文解析に基づいた同音語選択の方法について述べる。さらにすでに報告した分かち書きシステムとを統合することによって得られたべた書き文からの自動仮名漢字変換システムについて報告する<sup>9)</sup>。

### 2. 同音語処理の問題点

仮名漢字変換システムにおいて、同音語処理を考えると、まず問題となるのが自立語辞書の見出し単位と同音語の発生との関係である。辞書見出しの単位の設定には2通りの方法、たとえば、「情報処理の機器構成」という場合の「情報処理」、「機器構成」などの比較的長い単位までを含めて、一つの見出し単位と採る方法と「情報」、「処理」、「機器」、「構成」などのような単語単位を見出しと設定する方法が考えられる。国立国語研究所の語い調査<sup>10),11)</sup>では、前者の単位に対

† An Automatic Translation System of Non-segmented Kana Sentences into Kanji-Kana Sentences and its Homonym Analysis by HIROSHI MAKINO and MAKOTO KIZAWA (Faculty of Engineering Science, Osaka University).

†† 大阪大学基礎工学部情報工学科

††† 現在は図書館情報大学

して長単位、後者の単位に対しては、短単位という言葉がそれぞれ用いられている。

いま、前者の方法、すなわち見出し単位を長単位に採れば、「キキコウセイ」という語が出現しても、見出しの一致のみでも語が一意に定まり、変換という面から非常に有利となる。しかしながら辞書規模の拡大は避けられず、さらに長単位の語（複合語）を構成する単語が、単独で用いられた場合には、同音語の同定に長単位の語の性質を有効に生かすことができないなどの理由により、以下に述べるシステムでは、後者の単語単位の見出しを採用している。したがって上例の「キキ」、「コウセイ」などの見出しにおいて同音語が生ずることとなる。

また、活用語尾を持つ用言では、辞書見出しとして語幹（不変化部分）を採用していることから、実際の文中での語形さらには同音語を求める際に、語の品詞、活用までを考慮に入れる必要がある。逆にこの文法情報すなわち語の接続規則を用いて同音語の限定を行うことが可能となるために、文の構成要素である文節単位中に存在する同音語を対象とする方が、機械処理上も有利であり、さらに辞書中の同音語のみならず、実際に生じる異品詞間の同音語、たとえば「会いたい」と「相對」などをも対象とすることになり、われわれの同音語に対するイメージにも一致する。したがってここでは文献 7) で述べたシステムによって分かち書きされた文節単位の文字列に含まれる同音語（以後同音語文節と呼ぶ）を対象として処理を考えていく。ただし上述のシステムでは派生語に対し、自立語辞書、接辞（接頭・接尾語）リストの見出しの一致によって機械的に合成されたものは対象として含まれることになる。以上の前提のもとでの同音語（文節）の例を図 1 に示す。

図中の例でも分かるように、辞書探索での見かけの同音語（書(い), 貝) は語の持つ品詞、活用形などの情報を用いて、分かち書きの過程で処理されており、同音語文節は文法的に文節として十分な資格を持つために、同音語文節の処理には、文節間の関係すなわち文節を単位とした構文情報、意味用法を考慮していかなければならない。したがって、ここでは同音語の選択基準を与える情報を、

(i) 体言、用言間に存在する意味的關係に基づく語い情報

(ii) 主として複合語、派生語に関する単語固有の情報

- 
- (1) 自立語の文法的性質（品詞、活用形）が同一。  
ハナ（花、鼻、…）  
カイト（書いた、欠いた、…） cf. 貝
  - (2) 自立語の品詞属性は異なるが一部で同一。  
イッタ（行った、言った、…） cf. 行く、言う
  - (3) 接辞が同一。  
ゼンセカイ（全世界、前世界、…）  
ニホンゴ（日本語、日本後、…）
  - (4) 文節として同一。  
ナイヨウニ（内容に、無いように、…）
- 

図 1 同音語文節の種類

Fig. 1 Examples of homonyms.

に大別し、これら二つの情報の整理とその結果作成された用法辞書の形式について次に述べる。

### 3. 用法辞書

#### 3.1 用言の格

複数個の同音語の中から最適な単語を判定する手段の一つとして、各用言が特有の格の支配構造を持つと考えられることに着目して、体言、用言の意味的または用法上の関係を格関係によって捉える方法を試みた。すなわち用言の取り得る格とそれぞれの格に立ち得る体言の意味的範ちゅうを用法辞書に明記しておくことによって、同音語の選択、同定に利用しようとするものである。この実現のためには、個々の用言がどのような格を要求し、さらにその格に対応する体言の意味的範ちゅうが適切であるかを整理する必要がある。この作業では、単語の意味分類をいかに行うかが問題となるが、すでに国立国語研究所によって作成された分類語彙表<sup>12)</sup>の意味範ちゅうおよび意味分類がこの目的に有効であると考えられたので、それらをそのまま用いた。

用言の格は個々の用言によって異なり、その数も一定ではなく、その情報を記載する用法辞書も可変長形式が望ましいが、処理時間の短縮という意味で固定長形式の辞書構成をとることにした。そのためにすべての格の項目を固定的に割りつけることは無駄が多いので用言（特に動詞）に共通性の高い格と個々の用言特有の格という観点から次の二種類

#### 1 基本的な格

#### 2 固有の格

に大きく分類し、前者をそれぞれ固定の項目とし、固有の格に対して、一つの項目を割り当てることとした。

基本的な格は、下の例に見受けられるように、格助詞「ガ」、「ヲ」、「ニ」によって代表される格で、通常主格、対格、与格などと呼ばれる。ただし、格助詞

「ニ」は Fillmore が与格としたもの、すなわち動作の表す状態や動作の影響を受ける生物を表す格<sup>13),14)</sup>以外に例のように無生物に対しても用いられる。

(例)

- [私]が 歩く
- [本]ヲ 読む
- [物]ニ さわる

ここで [X] は単語 X の意味範ちゅうを示す。

格助詞「ニ」はまた場所を表す名詞とともに用いられ(場所格と呼ぶ)、この意味的に異なる二つの用法に対して、述語の動詞はいずれかの格しか要求しない。

すなわち、

「一つの動詞は、格助詞「ニ」により与格、場所格のうち一方だけを要求する。」

ことが経験的に言える。たとえば

- [人]ニ [家]ニ { 言った  
行った

はいずれの動詞の使い方も奇妙に感じられる。したがって、辞書項目では、二つの格に対する明確な区別をしないこととした。一方固有の格は動詞によって異なっているため、格を示す格助詞とともに辞書に記載することとした。たとえば、

書く： [人]が [紙]ニ [字]ヲ [道具]デ

行く： [人]が [場所]カラ [場所]ニ

に見られるように、「デ」、「カラ」などとともにそれぞれの対応する意味範ちゅうを記した。固有の格として基本的な格(ガ, ヲ, ニ)以外の格助詞を採用しているが、格助詞「の」、「や」などは連体修飾語句を作る性質から、格としては用いていない。

用言の中で、形容詞、形容動詞は述語、連体修飾語となる場合、係り受け関係にある体言との関係は、一種の主格関係と見なすことができることから、形容詞形容動詞に関する辞書項目としては主格の項にそれに対応する意味範ちゅうを記載した。

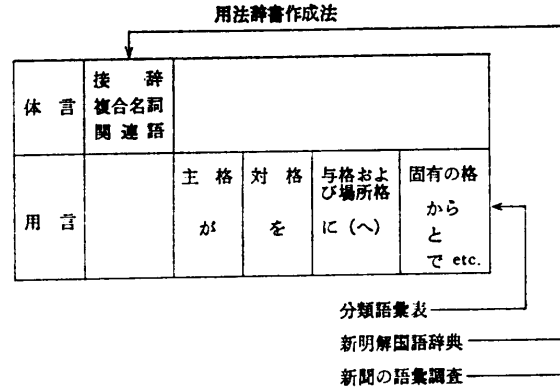
(例)

- [花]が 美しい。
- [人]が 綺麗だ。

以上述べてきた方針に従い、現在の自立語辞書中の用言 1,849 語(動詞 1,260 語、形容詞 156 語、形容動詞 433 語) およびサ変名詞 1,572 語について、格情報の整理を行った。なお用例の採集にあたっては、文献 15)~19) を参照した。

### 3.2 複合語、派生語

すでに述べたように辞書見出しの単位として単語単



項目 見出し	接頭語	接尾語	複合名詞、 関連語	主格	対格	与格 および 場所格	固有の格
労働	重	祭 者 省	委員 運動 基準 金庫 組合	[人]			テ[場所]*
気			利く する 付く 付ける				
読む				[人]	[図書]		
呼ぶ				[人]	[人]		

\* [場所] は分類語彙表の複数の意味範ちゅうに対応する。

図 2 用法辞書作成とその形式

Fig. 2 The format of the usage dictionary.

位を採用していることから、(名詞+名詞)、(接頭語+名詞)、(名詞+接尾語)によって表現される語、すなわち複合名詞、派生語に関する情報を辞書に記載する必要がある。ここでは用法辞書の接頭語、接尾語、関連語の項目名のもとに、直接該当する語を代入する方法を採った。各語の記載にあたっては、新明解国語辞典<sup>20)</sup>、新聞の語彙調査を資料<sup>10)</sup>とした。なお関連語項目には、体言、用言で示される慣用的表現をも合わせて記載した。たとえば、

- 気が 付く。
- 気を 付ける。

などである。

図 2 に用法辞書の作成法と辞書形式の例を示す。

## 4. 構文解析

同音語の用法の差異によって文中の同音語の中から正しい語を推定するために、まず文中の各単語または文節間の構文的な係り受け関係、すなわち文節の修飾、被修飾に基づいた解析を行う。

### 4.1 係り受けの基本規則

日本語文の係り受け構造はよく知られるように、次のような特徴を挙げることができる。

(1) 文節の修飾の型は連体修飾か連用修飾であり、文節末尾の語（活用語の場合はその活用形）の修飾の型によって決定される。

(2) 受けの文節（被修飾文節）はそれより文頭側の文節を0個以上受けることができる。

(3) 係り受けの関係は交差しない。

さらに、構文解析の簡単化のために、次の二つの規則を導入している：

(4) 係りの文節はそれより文末側の文節のいずれか一つに係り得る。

(5) 係り受けの関係は規則的に許されるかぎり、最も近い文節間で成立する。

上記(1)～(3)の規則のみでは、構文的なあいまいさを生じ、このあいまいさを解消するには、単語の持つ意味概念の詳細な情報を必要とし、意味的な解析の負担が重くなりすぎることから、通常の文でも比較的良好に成立している上記の規則(4)、(5)を導入した。以上の規則を用いて文節間の係り受け関係を求める。なお付属語の修飾の型は連体、連用のいずれかであるが、格助詞「の」、「と」に関しては、いずれの修飾の型も許している。

(例)

- 私の本(連体修飾), 私の書いた本(連用修飾)
- 彼と私(連体修飾), 彼と行く(連用修飾)

#### 4.2 補助規則

同音語選択の過程で行われる用法辞書項目の参照を容易にするために、いくつかの補助規則を設けている。

(1) 係り（修飾）の性質は文節末の語によって、受け（被修飾）の性質は自立語の品詞によってほぼ決められるが、一部の付属語（だ、であるなど）は述語としての性格を文節に与えることから、体言節から用言節への変換を行い、係り受け関係を求める。

(例)

それは 机だ

cf. 彼の机は

(2) 文節を形成している付属語が複数個存在する、いわゆる複合付属語の場合は文節末尾が係助詞（は、もなど）、副助詞（ほど、なりなど）などの場合が多く、格関係を示す格助詞はその直前に位置することが多い。したがって付属語の変換を行い、用法辞書

への参照を容易にする。

(例)

東京へも 行った → 東京へ 行った

(3) 受身、使役の文では格助詞を介した体言と用言の構文的関係が同一で、意味的には能動文のそれとは異なる。用法辞書中の格関係は能動文に対するものであるので、受身および使役の文では以下のように体言節の格助詞の変換を行う。

(i) 受身の場合（用言の付属語が「れる、られる」の場合）

N1ガ N2ニ Vサレル（ラレル）

→N2ガ N1ニ Vスル

(ii) 使役の場合（用言の付属語が「せる、させる」の場合）

N1ガ N2ニ N3ヲ Vサセル（セル）

→N2ガ N3ヲ Vスル

ただし、N1、N2、N3はそれぞれ体言、Vは用言（動詞）を表す。

(4) 接尾語および助数詞の付属した自立部を持つ文節では、その自立部の表す意味がそれぞれ接尾語、助数詞の意味によって代表されると考えられることから、以下の規則を用いて変換を行う。

[自立語+接尾語]→[接尾語]

[数字列+助数詞]→[助数詞]

(例)

[日本人]→[人]

[100人]→[人]

以上述べた規則の中で(1)は係り受け解析の段階で、(2)～(4)は同音語選択の直前に適用される。

### 5. 同音語の選択

同音語の選択は単語の意味分類に基づいた語い情報と単語固有の情報を用いて行われる。語い情報は文中での体言用言の格関係による選択および並列関係にある体言間の選択基準として用いられ、単語情報は派生語、複合語、慣用的表現の同音語の選択基準となる。ここでの同音語処理は、用言句（体言、用言の係り受け）および名詞句（体言、体言の係り受け）の処理に分けられる。

#### 5.1 格関係による同音語選択

##### 5.1.1 用言を修飾する体言の処理

すでに求められた係り受け関係をもとにして、用言を修飾する体言の中で、主格、対格、与格および場所格あるいは固有の格のいずれかに該当するかを体言に

付属する最後尾の付属語によって決定する。この判定の前に前節で述べた変換が可能ならば適用されている。格関係の判定には、それぞれの格表示に用いられる係助詞などにも用いられる。たとえば主格関係には、「は、も」などが参照される。そして体言の意味コード、すなわち分類語彙表<sup>9)</sup>の意味分類コードと用法辞書中のそれに対応する格項目に記された意味コードとの照合を行い、ある閾値以上で最も良く一致する体言を選択する。このとき修飾される用言が複数の場合(同音語が存在する)があるので、それぞれ用言の格の一致度が高い用言が選択され、それに対応する体言が選択される。なお格助詞「の」を介して用言を修飾する体言に対しては、用言の主格項目との参照が行われる。さらに、{だ、であるなど}の付属部を持つ体言では、係り受け関係にある体言との関係が同義または意味的に近い場合が多いという事実から、意味コードの近い体言の対を選択する。なお、単語間の意味の近さの尺度としては、意味コードの上位桁からのコード(文字)の一致数を用いている。

### 5.1.2 連体修飾される体言の処理

この場合は、文中に格を表示する助詞を持たず、受けの体言が係りの用言のいずれの格に相当するかの判定が困難なために、係りの用言の格の中でまだ見出されていない格項目の参照が試みられる。たとえば、

彼に もらった 本

では、

{本が もらう  
本ヲ もらう

として、すなわちこの場合は与格がすでに存在しているので、主格、対格項目への参照が試みられる。

## 5.2 名詞句の処理

前節では主として用言を中心とした同音語の処理について述べたが、本節では主として体言間の関係に基づいた処理について述べる。

### 5.2.1 接辞および複合名詞の処理

接辞(接頭語、接尾語)を持つ文節に対してその自立語の接辞項目を参照して、接頭語、接尾語を決定する。また、連続する体言からなる二文節すなわち複合名詞において、関連語項目を参照して、その単語が登録されている場合に決定を行う。さらに体言節が連用修飾している場合にも、受けの用言が関連語項目に記されているかを調べる。これは慣用的表現に対処するためである。

### 5.2.2 サ変名詞を含む複合名詞の処理

体言からなる文節が連続するという形態的な面からは上述の複合名詞と同じであるが、いずれかの名詞がサ変名詞である場合、この種の複合名詞を構成する単語間の関係は格関係と見なせる場合が多い。したがって、サ変名詞が動詞として用いられる場合の格項目との参照が試みられる。たとえば、

「ジョウホウ ショリ」

では、{上方、乗法、情報、…}の中から

情報ヲ 処理スル

として語の選択が行われる。

また、「ジョウホウノ ショリ」のように(名詞+サ変名詞)の場合にも同様の処理が行われる。

### 5.2.3 並列句の処理

並列助詞{と、や、だの、なりなど}を介して係り受け関係にある二つの体言は意味的に近い範ちゅうに含まれることが多いことから、意味コードがある閾値以上で最も近い体言の対が選択される。

(例)

専制と 隷従 cf. 先生、宣誓、先制

## 5.3 頻度処理

5.1, 5.2 で述べた同音語選択の処理で決定されなかった単語または文節に対しては頻度最大のものが選択される。

以上の同音語選択の処理の概略図を図3に示す。

## 6. べた書き文の仮名漢字変換システム

### 6.1 辞書構成

べた書き文から漢字仮名混り文への変換に用いる辞書を表1に示し、各辞書について以下に説明を加える。

#### (a) 自立語辞書

単語情報の内容は、通し番号、見出し、漢字コード、漢字数、品詞、活用、頻度、意味分類コード、継続情報の各項目からなり、現在の見出し語数は7,956語で多品詞語を含めると8,422語となる。

#### (b) 用法辞書

表1 変換用辞書

Table 1 Dictionaries for segmentation and translation from Kana to Kanji.

辞書名		辞書名	
(a)	自立語辞書	(e)	付属語表
(b)	用法辞書	(f)	付属語分かち書き用辞書
(c)	接続行列	(g)	接辞、助数詞表
(d)	用言の活用語尾表		

図2に示した辞書内容を持ち、自立語辞書に対応した参照番号を持つ。

(c) 接続行列

自立語および付属語の接続規則を表形式に表現したもので、その大きさは154×108である。

(d) 用言の活用語尾表

用言およびサ変名詞の活用語尾とそれらの接続行列の行番号との対応を示した表である。

(e) 付属語表

助詞、助動詞、形式名詞、補助用言とそれらの接続行列との対応を示した表である。付属語の中でも同形(同音)の助詞などに対処するために、最大3通りの対応を許している。

(f) 付属語分かち書き用辞書

未登録語などによって文字列中に自立語が探索されない場合に参照される高頻度の付属語の表である(文献7)、表2参照)。

(g) 接辞、助数詞表

接頭語辞書は見出しと漢字コードからなり、接尾語辞書、助数詞表はさらに意味分類コードが付加されている。助数詞表には漢字の字数の情報が記されている。各辞書の収録語数は、それぞれ47語、311語、141語で、資料<sup>11)</sup>の接辞表および国語辞典<sup>20)</sup>の用例中より採録した。

6.2 変換システム

べた書き文の仮名漢字変換システムは図4に示すハードウェア構成上で実現されており、プログラム言語はPL/Iである。

変換システムが起動されると、まず変換に必要な各種辞書群が読み込まれ、各種の初期設定が行われる。入力されたべた書き文は図5に示す二つの処理、すなわち分かち書き処理、同音語処理の過程をへて漢字仮名混り文に変換される。

分かち書き部は基本的にはすでに報告した分かち書きシステムと同一であるが、その後の実験によって得られた結果をもとに若干の改良を行っている。その一つは二文節最長一致法に関するもので、二文節形\*の長さが同一の場合、二文節形の構文的関係をより重視した分かち書きを行っている。

いま図6に示す入力文字列では、(a-1), ..., (a-3)

\* 文節形...文節の性質を有する文字列。

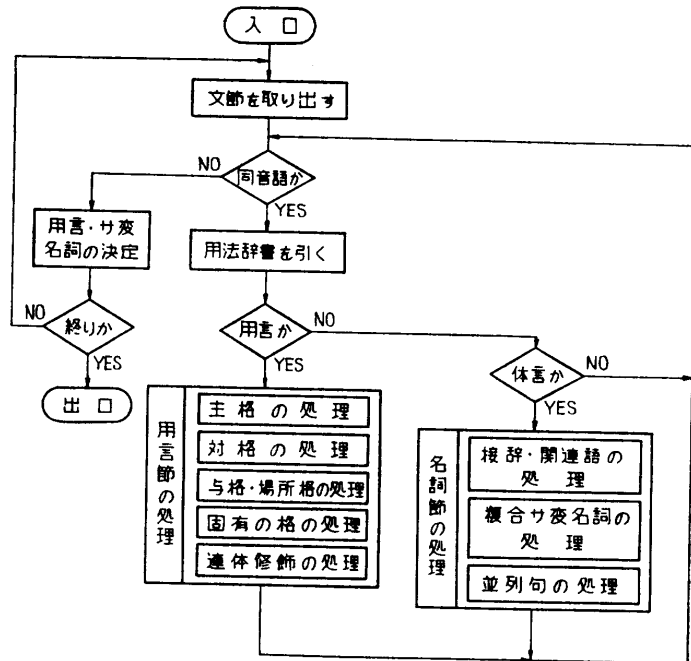


図3 同音語処理の概略図

Fig. 3 The flow-chart of homonym analysis.

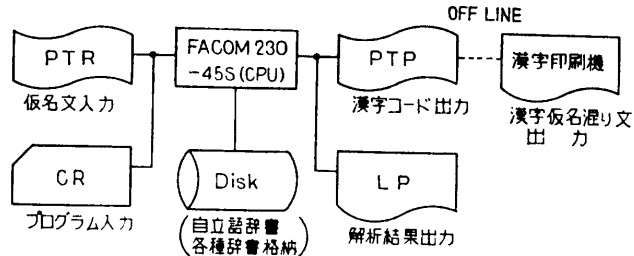


図4 ハードウェア構成

Fig. 4 Hardware configuration.

の文節形が抽出され、それぞれ後続の文字列に対し、文節形抽出が試みられ、(b-1)、(b-2)の連続する二文節形が求められる。この場合二つの二文節形の長さが同一なので、前システムでは自立部の最長性から、(b-2)の分かち書きを採っていたが、本システムでは構文的により望ましいと考えられる(b-1)の解釈、すなわち(連体形+名詞)が採用され、前半部の文節「勝れた」が決定される。もう一つの改良点は未登録語の出現に関するもので、(図6(II)参照)いま「キハン」が未登録語とすると、前システムでは図中(c-1)の分かち書きが行われた。これはすでに決定された区切りからの文節形が複数の区切りの可能性を持つときに、無制限に二文節最長一致を適用して、付属語分かち書きを適用していたからであるが、本システムでは



表 3 誤変換の例

Table 3 Examples of translation errors.

	誤	正
分かち書き	(1) 確か似合った (2) ポツポツミエテイテ (3) 泣き香の	確かに合った ポツポツ見えていて* 無きかの**
同音語の語誤選り	(イ) 時刻の主権 (ロ) これを犯してはならない (ハ) 左官に使われ	自国の主権 これを優してはならない さかんに使われ

\* 「ポツポツ」が未登録語

\*\* 形容詞活用は文法的に未処理

分かち書きの誤りは二文節最長一致法による誤り(表3(1)参照)、未登録語による誤り(表3(2))および未処理の文法に起因する誤り(表3(3))とに分けられ二文節最長一致法による誤りは、データ文中7例(14文節)で、総文節数の0.5%程度であった。本システムでは、文節形抽出過程での形態素分析の強化、二文節最長一致法の若干の改良などによって前システムより、誤りの少ない分かち書き結果を得ることができた。

同音語選択の誤りは、ほぼ頻度処理による誤りと言えるが、その原因は本システムの同音語処理の不備によるものであり、第一に名詞句(名詞+の+名詞)を形成する名詞間の関係が同音語処理に反映されていないこと(表3(イ))、第二に代名詞の意味的な処理が行われていないために、格関係による処理が有効に働かなかったこと(表3(ロ))、副詞に対する配慮がなされていないことによって、格関係が優先されたこと(表3(ハ))などが挙げられる。なお図8に漢字仮名混り文出力の例を示す。

## 7. あとがき

仮名漢字変換の主要な問題である同音語選択の方法とそのための辞書構成を中心に述べ、既報告の分かち書き法とを組み合わせることによって、べた書き文からの自動仮名漢字変換システムの実現を試みた。

実験例によれば、96.2%の漢字変換率(文節単位)が得られたが、これは頻度のみによる同音語選択に比べると約7%程度の変換率の向上となっている。この変換率の向上がすべて同音語選択の方法によるものではないが(分かち書き部での改善によるものが0.7%程度ある)、本論で探った同音語処理法は良好な結果を示したと言える。

今後の課題は、収録語数を増加し、より実用性を高めるとともに、現在もよく使用される古語的表現など

日本国民は、政党に選挙された国会における代表者を通じて行動し、我等と我等の子孫のために、諸国民との共和による成果と、我国全土に渡って自由のもたらす恵沢を確保し、政府の行為によって再び戦争の醜化が起こることの内用にすることを決意し、ここに主権が国民に存することを宣言し、この憲法を確定する。そもそも国政は、国民の厳粛な信託によるものであって、その権威は国民に由来し、その権力は国民の代表者がこれを行使し、その福利は国民がこれを享受する。これは人類普遍の原理であり、この憲法は、かかる原理に基づくものである。我等はこれに反する一切の憲法、法令及び詔勅を排除する。日本国民は、恒久の平和を念願し、人間相互の関係を支配する崇高な理想を深く自覚するのであって、平和を愛する諸国民の公正と信義に信頼して、我等の安全と生存を保持しようと決意した。我等は平和を維持し、専制と隷従、圧迫と偏狭を地上から永遠に除去しようと努めている国際社会において、名誉ある地位を占めたいと思う。我等は全世界の国民が等しく恐怖と欠乏から免れ、平和のうちに生存する権利を有することを確認する。我等は、いずれの国家も、時刻のことにのみ専念して他国を無視してはならないのであって、政治道徳の法則は、普遍的なものであり、この法則に従うことは、時刻の主権を維持し、他国と対等関係に立とうとする各国の責務であるとシズル。日本国民は国家の名誉に掛け、全力を上げてこの崇高な理想と目的を達成することを誓う。天皇は、日本国の象徴であり、日本国民統合の象徴であって、この地位は、主権の存する日本国民の総意に基づく。日本国民は、正義と秩序を基調とする国際平和を誠実に希求し、国権の発動たる戦争と、武力による威嚇または武力の行使は、国際紛争を解決する手段としては、永久にこれを放棄する。前項の目的を達するため、陸海空軍その他の戦力は、これを保持しない。国の交戦権はこれを認めない。国民は、総ての基本的人権の享有を妨げられない。この憲法が国民に保障する基本的人権は、犯すことの出来ない永久の権利として、現在及び将来の国民に与えられる。思想及び両親の自由は、これを犯してはならない。集会、結社及び言論、出版その他一切の表現の自由は、これを保障する。検閲は、これをしてはならない。通信の秘密は、これを犯してはならない。学問の自由は、これを保障する。総ての国民は、健康で文化的な最低限度の生活を営む権利を有する。国は、総ての生活面について、社会福祉、社会保障及び公衆衛生の工場及び増進に努めなければならない。人類は勝れた目と指先の感覚を持っているが、それは、今ごく一例を示したように、人類の思考法形成の一助を担っていると見える。

(注) 片カナは付属語分かち書きによる。

図 8 出力例

Fig. 8 Output examples.

の形態素分析を強化すること、格助詞「の」によって関係づけられる名詞間の関係を整理し、同音語処理に反映させることが挙げられる。

最後に、日頃ご指導を頂く大阪大学基礎工学部豊田順一助教授に深謝する。さらに本システムの作成に援助を願った岡田真和君(現兵庫県立工業高校)に感謝する。

なお本研究の一部は文部省科学研究費一般研究Bによるものである。

## 参 考 文 献

- 1) 栗原, 黒崎: 仮名文の漢字混り文への変換について, 九州大工学集報, Vol. 39, No. 4, pp. 659-664 (1967).
- 2) 松下, 山崎, 佐藤: 漢字カナ混り文変換システム, 情報処理, Vol. 15, No. 1, pp. 2-9(1974).
- 3) 江澤, 江原: 計算機によるカナ漢字変換, NHK



- 技術研究, Vol. 25, No. 5, pp. 261-298(1973).
- 4) 木村, 遠藤, 小橋: 日本語入力用カナ漢字変換システムの試作, 情報処理, Vol. 17, No. 11, pp. 1009-1016 (1976).
  - 5) 河田, 天野, 武田, 森: ミニコンピュータを用いたカナ漢字変換システム, 電子通信学会技報, PRL 76-47 (1976).
  - 6) 牧野, 勝部, 木澤: カナ漢字変換の一方法, 情報処理, Vol. 18, No. 7, pp. 656-663 (1977).
  - 7) 牧野, 木澤: べた書き文の分かち書きと仮名漢字変換—二文節最長一致法による分かち書き, 情報処理学会論文誌, Vol. 20, No. 4, pp. 337-345 (1979).
  - 8) 国立国語研究所: 同音語の研究, 秀英出版, 東京 (1961).
  - 9) 牧野, 木澤, 岡田: べた書き文の仮名漢字変換—同音語処理について, 電子通信学会技報, AL 79-67 (1979).
  - 10) 国立国語研究所: 計算機による新聞の語彙調査 (IV), 秀英出版, 東京 (1971).
  - 11) 国立国語研究所: 計算機による新聞の語彙調査 (II), 秀英出版, 東京 (1971).
  - 12) 国立国語研究所: 分類語彙表, 秀英出版, 東京 (1973).
  - 13) Fillmore, C. J.: THE CASE FOR CASE, In *Universals in Linguistic Theory*, Hort, Reinhart and Winston, New York (1968).
  - 14) Fillmore, C. J.: 格文法の原理, 田中, 船城訳, 三省堂, 東京 (1975).
  - 15) 宮島達夫: 動詞の意味・用法の記述的研究, 秀英出版, 東京 (1973).
  - 16) 石綿敏雄: 人間の精神活動を意味する動詞の用法—一言語情報処理のための動詞句の分析・その1, 国研報告 49, pp. 185-235, 秀英出版, 東京 (1973).
  - 17) 石綿敏雄: 自然現象を意味する動詞の用法, 国研論集 4, pp. 51-63, 秀英出版, 東京 (1974).
  - 18) 石綿敏雄: 抽象的關係を意味する動詞の用法—一言語情報処理のための動詞句の分析・その3, 国研報告 51, pp. 63-160, 秀英出版, 東京 (1974).
  - 19) 石綿敏雄: 動詞を中心とした語彙の分類, 国研報告 51, pp. 161-188, 秀英出版, 東京 (1974).
  - 20) 金田一他: 新明解国語辞典, 三省堂, 東京 (1971).

(昭和54年12月27日受付)

(昭和55年9月18日採録)