

かな漢字変換は日本語を入力するためのキー技術である。この技術は、ある程度のレベルですでに実現されているが、文節の切れ目や同音語の誤りなどが多く、さらに精度を向上させ、よりスムーズな日本語入力を可能にする必要がある。今回、われわれは、それぞれの単語の性質やふるまいに着目し、それらを記述したデータやルール(精緻(ち)化文法)を用いることにより、変換精度の向上を実現した。

このシステムは、ビジネス分野において98%の変換率を達成している。この技術は、日本語ワープロRupo JW-V700や、東芝かな漢字変換ソフトウェアRupo ACE<sub>TM</sub>に活用されている。今後、さらに精緻化情報を強化し、幅広い分野での変換率向上が期待されている。

*Kana-to-kanji conversion is the key technology for inputting Japanese. Although this technology has already reached a useful level, there are still many failures in terms of homonym errors and word-division errors. In addition to the need to further raise accuracy, smoother Japanese input must also be facilitated.*

*Recently, we have improved conversion accuracy by using data and the rules (subcategorization features) that describe them, paying attention to the usage and features of each word. This system, which has attained an accuracy of 98% for business documents, has been utilized in the Rupo JW-V700 Japanese word processor and the Rupo ACE<sub>TM</sub> kana-to-kanji conversion software. From now on, further subcategorization features will be added and it is expected that the accuracy of conversion will be improved for various types of documents.*

### 1 まえがき

情報処理装置において、スムーズに日本語を入力することは基本的な必須(す)機能である。当社は、わが国初の日本語ワープロJW-10で文節変換を搭載した。その後、1985年のJW-8D IIでは“全文まるごと変換”を採用して、変換キーを不要にし、文節を意識せずに入力できるシステムを開発した。さらにその後、RupoシリーズにおいてAI(人工知能)・連想変換、ニューロかな漢字変換を開発し、第一次変換率(読みを入力して、かな漢字変換を実行した直後の結果が正解に変換される率)の向上に注力してきた。これにより、第一次変換率90%台を実現したが(JW-05HG)、日本語入力手段として十分であるとはいえなかった。また、入力する分野によっては、変換率が低いものもあり、これらの強化が必要であった。特に、ビジネス分野の文書に対する性能の改善が望まれていた。

そこで、これらの問題を解決するために、辞書に登録する語彙(い)、AI情報の強化とともに、変換ロジックの抜本的な見直しを実施して、新たに精緻化文法を採用した(JW-06Hより)。これにより、最終的にビジネス分野での第一次変換率98%を実現した。近年開発したRupoシリーズの第一次変換率の推移を図1に示す。

ここでは、新たに採用した精緻化文法について簡単に説

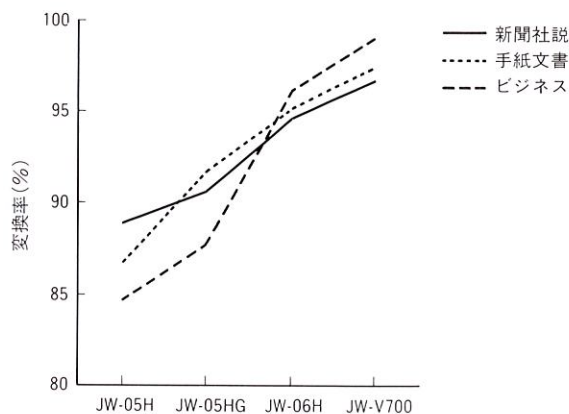


図1. 変換率の推移 近年開発したRupoシリーズの第一次変換率の推移を分野別に示す。精緻化文法を採用したJW-06Hから飛躍的に上がっているのがわかる。

Changes in accuracy of conversion

明し、この方式を採用した最新機種(パーソナルワープロRupo JW-V700、東芝かな漢字変換ソフトウェアRupo ACE<sub>TM</sub>)の概要を述べる。

### 2 精緻化文法

当社の変換における文法解析には、当社が開発した精緻



化文法を用いている。これは従来の品詞（名詞、動詞、形容詞など）を形態的、構文的ふるまいに基いてさらに細分した品詞体系を用いたものである。この精緻化文法により、変換精度を飛躍的に向上させることができた。それは、従来の文法では品詞分類が不十分で、文法的機能が異なるものにも同一の品詞を与え、同一の文法処理を行っていたため誤変換を生じていたからである。以下、精緻化文法について、名詞の修飾性と被修飾性に基づく品詞細分の例を用いて説明する。

## 2.1 名詞の細分

**2.1.1 用言修飾性の観点からの細分** 名詞の修飾性に基づく細分の例として、用言修飾性の有無による細分がある。一般的な名詞（機能、大使、工場など）には用言修飾性がないが、次に示す名詞などにはこの性質があり構文的ふるまいが異なる。

時を表す名詞：当日，昨日，来年，将来，長年

数量名詞：大勢，多数，少量，全部，半分

副詞的名詞：空路，他方，内実，概略，最低

機能的な名詞：あまり，とおり，くらい，ため

**2.1.2 名詞に対する被修飾性の観点からの細分** 名詞の被修飾性に基づく細分の例として、名詞に対する被修飾性の有無による細分がある。それぞれの例を次に示す。

もつもの：高騰，機関，事業，事項，専用

もたないもの：口答，窮状，貴社，上記，既製

## 2.2 文法の適用

文法は、入力されたひらがなから生成される多くの変換候補から構文的に正しい候補を選択するために用いられる。ここでは、上述の細分品詞に基づく精緻化文法適用について、従来文法の適用と対比して述べる。

### 2.2.1 名詞と動詞が隣接して出現する候補への適用

具体的に「きのうかいたてがみ」がかな入力された場合で説明する（図2）。この入力に対しては、変換候補として「機能書いた手紙」（候補1）と「昨日書いた手紙」（候補2）が生成される。従来は、「機能」、「昨日」には、品詞として「名詞」を与えていたが、精緻化文法では、「機能」には「用言修飾性のない名詞」、「昨日」には「用言修飾性のある名詞」という異なる品詞を与える。従来の文法規則(1)は、「名詞と動詞が隣接して出現する」場合は受理することを意味している。精緻化文法規則(2)は、「用言修飾性のある名詞と動詞が隣接して出現する」場合は受理することを意味している。それは、用言修飾性のある名詞は隣接する動詞の後方に出現する単語を修飾する可能性があるからである。(3)は、「用言修飾性のない名詞と動詞が隣接して出現する」場合は受理しないかまたは非優先にすることを意味している。それは、この場合の名詞の修飾先は隣接する動詞に限られ、名詞が格助詞を伴わずに隣接する動詞を修飾する可能性は、書き言葉では非常に低いからである。つまり、従来文法で

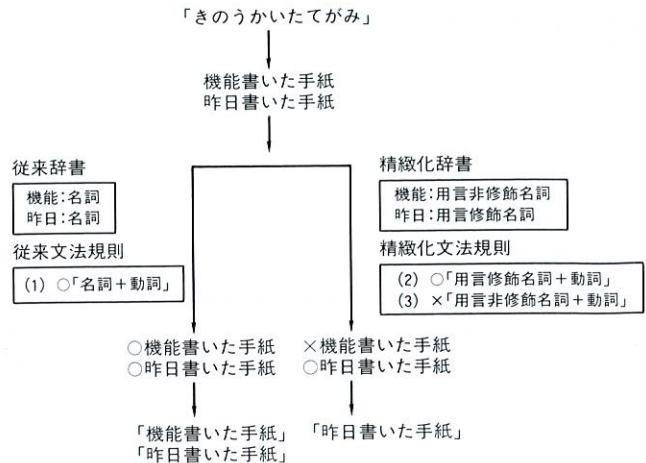


図2. 文法適用の例(1) 用言修飾可否の情報により、正しい結果を得られるようになる。

Example of syntactic analysis (1)

は候補1および候補2に文法的な差は生じないが、精緻化文法では文法的な差が生じ候補2を優先することができる。

上述と同様の処理で、「誤りの内容(ないよう)願います」、「尊敬勝つ(且つ)信頼して」、「意向(以降)作った」などの誤変換を回避することができる。

### 2.2.2 名詞と名詞が隣接して出現する候補への適用

「かかくこうとうで」がかな入力された場合の処理を図3に示す。候補1および候補2は、従来文法ではいずれも差は生じないが、精緻化文法では候補2だけを受理することができる。同様の処理で、「規制窮状が(貴請求状が)」、「学会既製が(学会規制が)」などの誤変換を回避することができる。

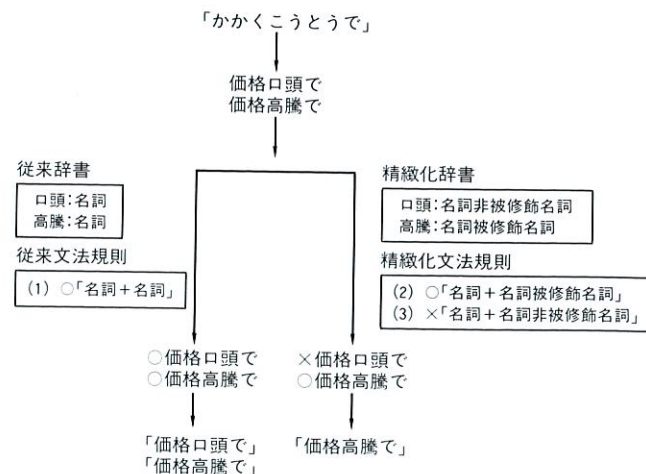


図3. 文法適用の例(2) 名詞からの被修飾性可否の情報により、正しい結果を得られるようになる。

Example of syntactic analysis (2)



### 3 Rupo JW-V700 のかな漢字変換の概要

JW-V700 では、100 万語を超える大語彙ハイパー AI 辞書とともに、今回開発した精緻化文法を採用した最新のかな漢字変換を搭載している。また、ニューロかな漢字変換をはじめとした、変換率向上のための多彩な機能を備えている。

#### 3.1 高精度のかな漢字変換

JW-V700 ではビジネス分野において 98% の第一次変換率を実現している。また、第一候補で得られない場合でも、次候補操作 4 回以内に 99% 以上の率で正解候補を得ることができる。近年開発した Rupo シリーズの、ビジネス分野での第  $n$  次候補での変換率のグラフを図 4 に示す。精緻化文法を採用した JW-06H から第一次変換率、およびその後の第  $n$  次変換率が向上していることがわかる。

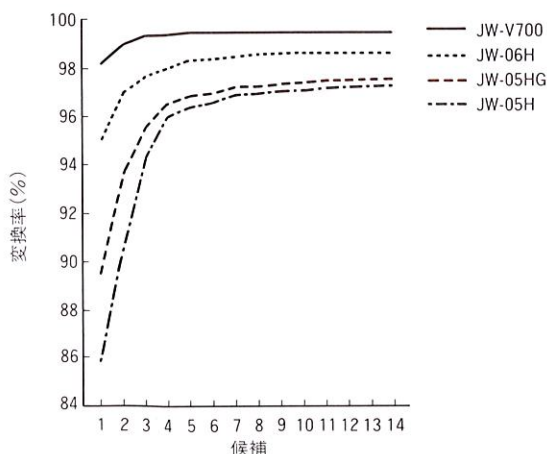


図 4. 第  $n$  次候補における変換率 JW-V700 は第 5 次候補まで (次候補キー操作 4 回以内) にほぼ確実に正解が得られることがわかる。  
Accuracy of  $n$ -th conversion

#### 3.2 ニューロかな漢字変換

ニューロかな漢字変換機能について述べる。これは、ニューラルネットワークを利用することにより、入力中の文章に使用されている言葉から文脈を判断し、文脈に合った言葉を優先出力することで、最適な言葉に変換、文脈の変化に臨機応変に対応する機能である。

図 5 にネットワークの概念を示す。同じ話題のなかで使われるような、意味の関連した単語どうしをネットワークで表現する。ある単語が入力中に選択されると、その単語 (ニューロン) に刺激を与え、リンクのつながっている各ニューロンへ刺激を伝播 (ば) させる。その結果、その単語の属するグループ全体の活性化度が上がり、相対的に他のグループに属する単語の活性化度が下がる。このため、同音異義

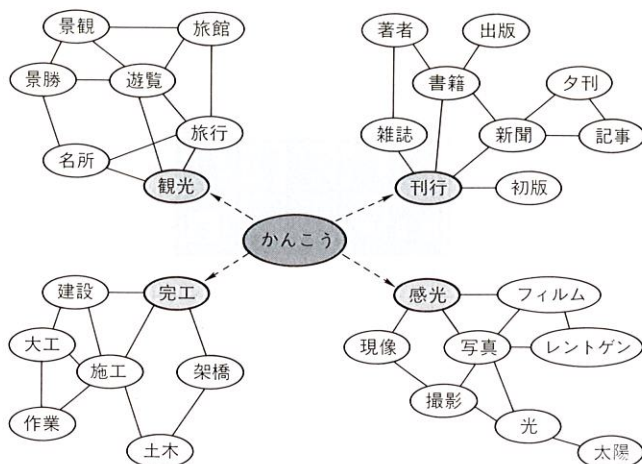


図 5. ニューロかな漢字変換の概念 関連する語をニューラルネットワークで構成し、それらの使用状況によってグループの活性化値を制御し、関連のある同音語を優先する。

Concept of kana-to-kanji conversion using neural networks

語の中から活性化値の高いグループに属する単語を優先することにより、文脈に沿った変換が可能になる。

### 4 東芝かな漢字変換ソフトウェア Rupo ACE<sub>TM</sub> の概要

パーソナルワープロ Rupo JW シリーズのかな漢字変換を基本に、Windows<sup>®</sup> (注1) の標準インタフェース上で実現した日本語入力システム Rupo ACE<sub>TM</sub> について概要を紹介する。

#### 4.1 高い変換性能

当社独自の全文まるごと変換方式はもちろんのこと、AI 用例を強化したハイパー AI 辞書、さらに新たに精緻化文法を取り入れたことで、パソコン用日本語入力システムの中でも最高の変換性能を実現した。

#### 4.2 わかりやすい操作性

変換性能だけでなく、だれでもすぐに使えるくふうを随所に施している。そのなかでも、入力の状態に応じたガイド行表示と使用頻度の高い機能をアイコンとしてまとめたモードウィンドウは、その代表例といえる。図 6 は、その例である。

#### 4.3 誤入力自動補正

ローマ字入力やかな入力時に、おかしやすい入力ミスを自動的に補正して正しい変換結果を得ることができるため、思考を妨げないスムーズな入力環境が実現できた。図 7 は、その例である。

#### 4.4 単文節共起変換 (単文節 AI 変換)

従来の AI 変換では、ユーザが変換単位を長く入力しないと AI 辞書の効果が得られなかった。しかし、ユーザによ

(注 1) Windows は、Microsoft 社の商標。

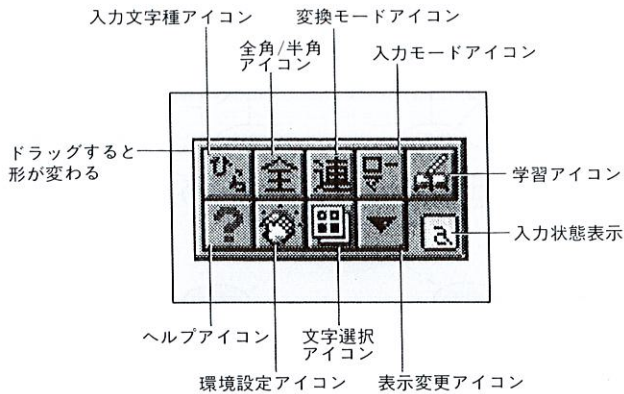


図6. モードウィンドウ よく使う機能をアイコンにしてまとめたものである。

Mode window

- ローマ字入力モードのときに有効
    - きってをあつめる ⇨ きってをあつめる ⇨ 切手を集める
    - にゆりよくのほうほう ⇨ にゆりよくのほうほう ⇨ 入力の方法
    - ほんやくする ⇨ ほんやくする ⇨ 翻訳する
  - かな入力モードのときに有効
    - まるちめていあ ⇨ まるちめていあ ⇨ マルチメディア
  - モードにかかわらず有効
    - いづれ ⇨ いづれ ⇨ いづれ
    - しみゆれーしょん ⇨ しみゆれーしょん ⇨ シミュレーション
- [入力された文字列]      [補正された文字列]      [変換結果]

図7. 誤入力自動補正の例 誤入力自動補正が適用される入力の例を示す。

Example of automatic correction

てはこまめに変換キーを押しながら変換していく人もいる。このような入力方法をとっても、高精度かな漢字変換エンジンの効果が得られるようにした。

例：とりがいつせいに [変換] きれいになく [変換]  
 例：とりが [変換] いつせいに [変換] きれいになく [変換]

変換結果

[従来] 鳥が一斉にきれいに泣く  
 [今回] 鳥が一斉にきれいに鳴く

#### 4.5 充実したユーティリティ

自分好みの環境を構築できるのはもちろんのこと、操作性、ユーザ辞書、ローマ字変換における他社 FEP (フロントエンドプロセッサ) との相違も解消でき、簡単に東芝かな漢字に移行することが可能である。図8に、それぞれアイコン表示されている状態を示す。

- 環境設定 各ユーティリティで設定した環境をスタイルファイルに保存したり、切り換えを行う。
- 入力環境設定 入力、変換に関する設定や辞書ファイルの設定を行う。
- キーカスタマイズ キーに対する機能の割当てや

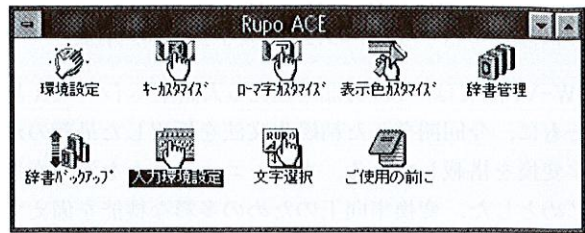


図8. ユーティリティのアイコン表示画面 各ユーティリティのアイコンが RuPo ACE™ のグループウィンドウに表示されている状態を示す。

Icon pattern of utility

機能に対するキーの割当ての設定を行う。

- 表示色カスタマイズ 画面ごとに文字と背景色の設定を行う。
- ローマ字カスタマイズ ローマ字入力力で文字を入力するときのキーの組合せに対する読みの設定を行う。
- 辞書管理 単語の一覧表示や単語登録、削除や辞書を合成することが可能である。
- 辞書バックアップ 重要な辞書を保存したり、新規辞書の作成を行う。

## 5 あとがき

精緻化文法を採用した高精度かな漢字変換について概要を述べるとともに、それを搭載した商品を紹介した。

ビジネス分野 98% という高い精度でかな漢字変換できるとともに、それをサポートする豊富な機能により、効率よく日本語が入力できる。今後は、一般的な文書作成だけでなく、メールなどで活用できるような、多様な分野で扱えるよう、高精度の範囲を拡大していく必要がある。



野上 宏康 Hiroyasu Nogami

情報・通信システム研究所研究第五担当主任研究員。  
 自然言語処理の研究・開発に従事。情報処理学会、電子情報通信学会、人工知能学会会員。  
 Communication & Information Systems Research Labs.



中里 茂美 Shigemi Nakasato

マルチメディア技術研究所開発第五部主務。  
 自然言語処理の研究・開発に従事。情報処理学会会員。  
 Multimedia Engineering Lab.



唐崎 幸弘 Yukihiko Karasaki

青梅工場ワープロ設計部主務。  
 自然言語処理ソフトウェアの開発・設計に従事。電子情報通信学会会員。  
 Ome Works