解　説

IPS

自然言語処理技術の応用

1. 日本語ワードプロセッサにおける自然言語処理†

斎藤裕美††　野上宏康†††

1. まえがき

文書の入力，編集，記憶保存，および印刷の各基本機能を有するワードプロセッサ（ワープロ専用機のほかパーソナルコンピュータ上のワープロソフトも含める）は，広く一般に使われるようになった．商用の日本語ワードプロセッサが登場して15年近くになるが，その間半導体などのハードウェア要素の小型化，高速化，低価格化の進歩に合わせて，ソフトウェアでも改良強化が続けられ，今日では実に多種多様で高度な機能が手軽に実現できるようになった．

このように文章の作成あるいは作成した文章を取り扱うワードプロセッサの内部処理には，個々の文字コードやコード列のデータに対してことばとしての特徴をとらえて動作する自然言語処理部分がある．自然言語処理とは常にあいまいな結果をもたらす部分でもある．スペルチェッカやスペルコレクタが主である欧文の世界とは異なり，日本語ワードプロセッサでは漢字を含む表現をいかに入力するかということが大きな課題であった．したがって仮名やローマ字キーを使って，より自然で使いやすい仮名漢字変換を実現するためには自然言語としての日本語の特性を強く意識した仕掛けが必要である．

以下本稿では，日本語ワードプロセッサにおける自然言語処理として仮名漢字変換を中心に取り上げ，各種の文法処理，意味処理に基づく一般的な変換方式，辞書のあらましなどについて述べる．また文字や表現の誤りを直すための支援機能や日本語表現の特徴をとらえて文書構造に適した紙面の体裁を整える例にも触れる．

2. 仮名漢字変換

昭和53年に発表された初の商用ワードプロセッサでは文節式と漢字指定式の入力方式が提供され，そこで採用された活用語尾や付属語の接続性を解析する機構は多数ある現在の装置でも基本的に共通したものである．文節内文法解析と呼ばれるこの解析機構は，仮名書き表記の入力から変換後の対話操作を最小限にするために不可欠なものであり，その後，現在に至るまで文節や漢字などの文字種の指定を省略できる連文節入力方式が実用化され，また変換率の向上のための数々の工夫がなされてきている．

ここで仮名漢字変換の研究開発に関しての経緯を簡単に整理する．まず大学で研究が始められ，入力文の文節分かちや同音語（**2.3**参照）の現象などの基本的，一般的な問題点の整理と考察を経て，文節解析，単語辞書の構成と照合法，さらに連語情報を利用した同音語識別率向上の方法などに関する提案が昭和40年代前半になされ[1]，一部のメーカがこのシステムを試作した[2]．また特定分野についての実用化の試みとして，ローマ字の外電からの漢字混じり文変換システムが通信社で作成された[3]．さらにニュース文を使用した試作実験システムの発表があり，ここでまとめられた文節内文法解析のための具体的でコンパクトな単位分類と接続表の内容が解析アルゴリズムとともに注目を集めた[4]．また単語の意味分類情報を使った同音語判別手法も報告されている．関連語情報や単語の意味分類情報の利用を含んだシステムの試作例もあった[5]．

初期の日本語ワードプロセッサには単に文字列表記の変換操作にとどめその意味では自然言語処理には関わらない方式も存在したが，一部の装置

† Natural Language Processing for Japanese Word Processor by Hiroyoshi SAITO (Information Systems Engineering Laboratory, Toshiba Corporation) and Hiroyasu NOGAMI (Research and Development Center, Toshiba Corporation).

†† (株)東芝情報処理・機器技術研究所
††† (株)東芝研究開発センター

では文節内文法解析を基本とし，固有名詞や数値表現部分に限って意味的な関連語情報を適用していた[6]．いずれにおいてもユーザ辞書への登録機構や，能率良く同音語選択ができるように最終使用語優先の学習機構を組み込んだことが特徴であった．これらは分かち単位の入力を前提にしていたものであり，日本語解析の基本である文節内文法解析の手法はほぼ確立された．次の課題は文節指定のいらないベタ書き仮名入力をいかに誤りなく自動分かちするか，および同音語識別のために必要な意味的処理をいかに実用化するかであった．

ベタ入力文の解析に関しては，2文節最長一致法[7]，語長と頻度の評価関数による方法[8]，文節数最小法[9]がそれぞれ研究され，その後さらに操作上の工夫が施されて商品にも取り入れられて連文節変換，一括自動変換，全文変換などの呼び名で現在標準的に採用されるようになった[10),11]．いっぽう同音語の正解率を向上させるための意味的手法としては，あらかじめ記憶された共起関係データ（同時出現性の高い関連語の例）に一致する候補を優先させる方式が広まり，AI辞書変換などと呼ばれて実現されている．このほかに最近では話題の変化を判断してその話題に即した優先語を決めていく方式も一部で実用化されている．以下，仮名漢字変換のための日本語文解析の基本となる文節内文法解析，入力操作の改善のための連文節変換，および同音語の判別率強化のための取組み方に分けて説明する．

2.1 文節内文法解析

日本語文は文節を任意の個数連結した形を成しており，おのおのの文節を構成している単語どうしには規則的な関係がある．文節内の単語構成は**図-1**のように定義でき，さらに各単語間の接続に関して活用語尾や付属語の種類に応じた一定の制限がある．この関係規則を基に入力文に対して単語辞書から単語を引き当てることを形態素解析と呼び，仮名漢字変換とは形態素解析で得られた単語について辞書に示された漢字表記を出力することである．文節内形態素解析は単位分類と接続表を使用して行う[4]．単位分類は接続性を規定する単語を機能別に分類したもので二種類ある．一方は後続単語に働き掛ける性質で分類したもので用言や助動詞は活用変化ごとに異なる単位とし，他方は先行単語が共通という観点で活用変化は考慮せずに付属語全般を分類したものである．接続表は前者の単位分類番号を行，後者の分類番号を列として先行単語と後続単語との接続性を定めた行列である．

$C_{ij}=1$：行 i が列 j に接続可能

$C_{ij}=0$：行 i が列 j に接続不可

また行番号に対しては次のような文節終端条件も示している．

$T(i)=1$：文節終端可能

$T(i)=0$：文節終端不可

自立語と付属語，さらに付属語と付属語の構成を解析するには，入力文字列の先頭側からまず辞書照合で得られる自立語候補を切り出し，求めた行番号 i と続く付属語候補に対する列番号 j との間で接続表を調べる．さらに順次後続する付属語候補との関係を調べていき，入力文字列終端まで接続性が満たされ，しかも最終単位が文節終端条件を満たしているとき解析成功となる．**図-2**は“ながれたが”を解析するときの接続表の例である．

以上の文節内解析の具体的手法は，形式言語理論における正規文法に相当するものであり，次の規則で文節を表すことができる．

A→aB，B→b

（A, B：非終端記号　a, b：終端記号）

これらの解析では入力文字列に対して可能なすべての解釈を試み，成功したものを選び出して出

〈文節〉:=〈通常文節〉|〈数詞文節〉|〈固有名詞文節〉
〈通常文節〉:=([接頭辞]自立語[接尾辞]*)*[付属語]*
〈数詞文節〉:=[前置助数詞]数詞[後置助数詞][接尾辞]*[付属語]*
〈固有名詞文節〉:=[接頭辞]固有名詞[接尾辞]*[付属語]*
[ ] 印は省略可　　* 印は繰返しが可

図-1 文節の定義

i \ j	助動詞「た」	格助詞「が」	接続助詞「が」	助動詞「れる」	$T(i)$
一段活用動詞連用形	1	0	0	0	1
一段活用動詞未然形	0	0	0	0	0
名　詞	0	1	0	0	1
助動詞「た」終止形	0	0	1	0	1
助動詞「た」連体形	0	0	0	0	1
格助詞「が」	0	0	0	0	1
接続助詞「が」	0	0	0	0	1

（入力：“ながれたが”に関係する項目の例
「ながれ“流れ”」→一段活用動詞連用形，一段活用動詞未然形
「ながれ“流れ”」，「な“名，菜”」→名詞）

図-2 接続表の例

力する．付属語部分や活用変化語尾は非漢字であるので，複数の解釈があるときはそれぞれの自立語部分が同音語候補となる．また，接頭，接尾は名詞や固有名詞など接続可能な特定の単語と結合した後，同様に後続の付属語との接続性が調べられる．

2.2 連文節変換

複数の解釈によって同音語が生じるのは次のような場合がある．

（1） 単独自立語

a かがく→科学/化学/価額
あつい→厚い/暑い/熱い

b ひとで→人手/人で/火とで
よく→良く/欲/翌

（2） 接辞語の結合

しんぶんや
→新分野/新聞屋/新聞や

（3） 文節の結合

ていあんしたいけん
→提案したい件/提案した意見/提案し体験

ほけんがいしゃに
→保険会社に/保険が医者に

これらのうち(1)(2)は文節内文法解析で同音語候補として扱われる現象であるが，連文節の解釈では(3)のようなあいまい性にも対処する必要がある．文節を指定しない連文節変換では，自動的に文節境界を推定しなくてはならないが，解釈し得る文節形を単純に並べただけでは多数の不自然な表現形を生じてしまう．

ていあんしたいけん
→提案慕い件，低案下意見

自動分かちの有力な手法としては以下に示すようなものが提案されている．

A．文節最長一致法

最長の文字列に対応する文節解釈を優先して区分していく．比較的処理が簡単であるが，後続文節の文頭の読みであっても常に先行文節の付属語系列部分としか解釈されないという欠点があり，他の方式より精度が劣る．

B． 2文節最長一致法

連続2文節の長さが最長となる解釈を求め，そのときの先行文節を決定し，順に後続の文節を先頭として次の文節との間で同様の評価を繰り返していく方法である．最長2文節の中の文節境界が複数存在するときには優先度基準にしたがって決定する．優劣のないときは一般に先行文節の付属語が少ないすなわち前方側が短いものが正解となることが多い．文節最長一致法に比べ，1文節先まで含めて解釈の妥当性を評価できるので精度が向上している．入力文字列がどんなに長くても一度に行う処理量は一定であることで以下に示す文節数最小法より処理コストの点で有利であり，一般に多く使用されている．

C．最尤評価法

単語長と頻度に基づく評価関数を定めてその計算値に従って分割位置を判断していく方法で，長さの短い単語でも頻度が大きいものは優先されるようにしたものである．具体的な評価関数はさまざまな実例文に対する実験から最適な値に設定する．

単語長と頻度のほかに，前後の単語間の接続の重みも数値化して評価関数に取り入れた方法もある[12]．

D．文節数最小法

入力文字列を構成する文節の総数が最小になる解釈にしたがって区分する方法であり，平均的に最長文節の並びを求めることに相当する．総当たりで探索を行って判定するのでより多くの記憶領域を必要とするが成功率は高い．ほとんどの場合，最小文節数になる区切り方と1文節多い区切り方までの解析候補の中に正解があるので，それらの中から単語並びなどに関する評価基準を決めて出力候補の優先度付けを行う方法がとられる．

E．共通区切り探索法

2文節最長一致法を拡大して連続 n 文節の長さが最長となる解釈を求め，そのときの前方 $n-1$ 単位の文節に区分していくが，最長となる文節系列が複数存在する場合にはそれらの中で共通する文節境界で区分する．すなわち必ずしも文節単位に一つずつ区切らなくても確実性の高い文節境界のみを決定することとし，文節境界の判断があいまいなところは前後まとめて拡張同音語構造として扱うことができる．n を増やしていくと文節数最小法のように処理コストが大きくなるので現実では $n=4$ あたりで行っている[13]．

図-3にこれらの例を示す．

単純な文節最長一致法以外の上記の方式においては，文節区切りを決定する上で経験的に得られ

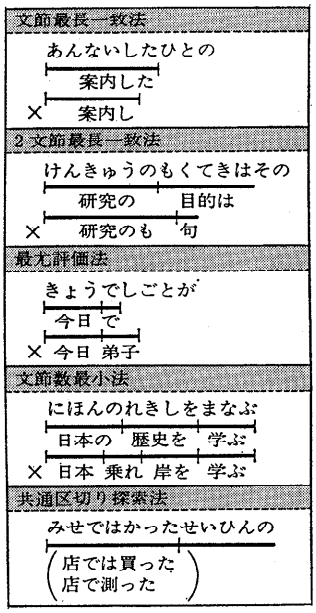


図-3　連文節解析の例

るいくつかの現象を規則化して精度を保つための工夫が試みられている.

- 1字名詞候補は優先度を下げる
  増えてきたと/聞く（×増えてきたとき/区）
- 漢字熟語の結合を優先
  行政/改革（×行政か/威嚇）
- 名詞に直接動詞がつきにくい
  彼は/知っていると（×彼/走っていると）

変換の操作面からみると，連文節解析でない方式では文節ごとに変換キーを押す必要があったが，現在の多くの装置では句読点の入力で，あるいは一度に入力できる仮名の最大数以内での変換指示により連文節解析が起動される．また仮名列の入力と並行して順次自動区分されて変換結果の候補を表示していくものもあり，その場合は句読点のない文でも変換指示が不要である.

なお，実際のワードプロセッサにおいては，これらの文節自動分かちでは完全には避けられない区切り誤りに対して強制的に区切り位置を変更させるための操作機構を提供している.

2.3　同音語の判別

国立国語研究所によれば同音語には次の種類がある[14].

（1）　表記にゆれのある短単位
　　歓び/喜び
（2）　送り仮名にゆれのある短単位
　　表す/表わす
（3）　同音異義語
　　自己/事故，厚い/暑い/熱い
（4）　同音類義語
　　機具/器具，追求/追及/追究

これらのうち，(1)(2)は使用者の好みに関わることで，また(4)の判別は困難であり，機械処理としては(3)の判別が課題であるとされている.

また，開発側の立場からは，実例の特徴別に次のような種類の対応が考えられる[15].

（a）　体言と用言の格関係で同音の用言を判定
（b）　体言と用言の格関係で同音の体言を判定
（c）　慣用表現として優先
（d）　文章内の話題性から判断
（e）　非一般語に対しては分野情報で判断
（f）　極端な頻度差があるものは頻度情報で判定

これら(a)から(d)までのように同音語の決定を実現するには単語の意味情報に取り組む必要がある．アクセントのような音声情報は利用できない．以前から慣用表現などの連語情報や単語の意味分類コードを基にした単語間の親和性を判断して優先付けする方法などが提案されていたが，そのための知識の記憶データや処理の負荷が多いので初期のワードプロセッサでは実用化が難しくせいぜい頻度情報で候補を優先付ける程度であった．その後，より高い率で正解候補を第1変換候補にあげるための手段として，主に単語共起関係を利用した方法が，高密度メモリや CPU の高速化に支えられ一般的に実現されるようになった.

2.3.1　共起語の利用

多くの用例から同時出現単語を選び出して単語対形式の辞書をあらかじめ用意しておき，変換結果の文節系列の同音語候補の中からこの辞書に登録された単語対と一致したものを優先出力することにより正解候補を第1順位に引き上げようとするものである[16),17]．単語対になるのは関連語すなわち一般に続けて使われる可能性が高いもので，意味的な制約関係にある修飾語と被修飾語のペアや複合語を構成する単語並びあるいは慣用句を構成する単語の組合せなどが相当する．図-4は共起語による変換例である.

実際の共起語辞書では数万から数10万例の規模の具体的な2単語の対データを提供する．文法

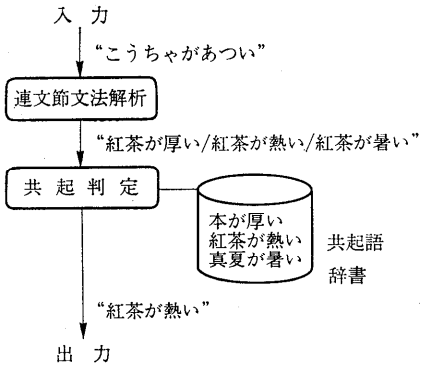


図-4 共起関係による同音語判定

解析を経て文法上の矛盾のない文節列の同音語候補の文字列と共起語辞書との間で照合処理を行うが，そのときに対象とする文節列の範囲は，前後2文節の単語とするのが比較的簡単な手続きで実現できる．実際の新聞データに基づいた調査で同音語の約7割は前後の語の情報で判別可能という報告もある[14]．

共起語辞書の構成として，2単語の対にさらに単語間の格関係を表す情報を付加することでマッチングをより限定させることができる．格関係の情報としては表層の格助詞を用いる．体言と用言の用例であれば，名詞—助詞—動詞のような形式で助詞まで含めて一致性を判定する．このことにより，たとえば，「名が通る」という用例を単に「名　通る」の対データだけで提供したときに，「なかをとおる」に対して「名かを通る」のような誤った結果になることを防ぐことができる．実際の用例では唯一の助詞のみが適応され得るとは限らないので，許容される助詞の組合せで分類しておくと記憶データサイズの増加を抑えることができる[18]．

また，「この本は少し厚い」のように共起語の対の間に他の修飾語が入ることもあるので，適用文節列の範囲を前後2文節から3，4文節当たりまで拡張した方法もある．

共起語辞書によりあらゆる同音語の判定を正しく行おうとすれば予想し得るさまざまな用例として数百万あるいはそれ以上のオーダの膨大なデータの規模の登録が必要になる．その対策として類義の単語をカテゴリにまとめ共起関係の知識を個々の単語の代わりに単語カテゴリ番号で記述する方法が考えられる．1000種類の意味カテゴリを用い，同一文節構造内の同音語判定のみでなくあいまいな文節構造の判定にも適用した研究では，平均3.7文節の文に対して文節数最小法のみの解析に比べて1/7〜1/10の候補に絞り込めることが報告されている[19]．

共起語データは仮名漢字変換をはじめ各種の言語処理にも有効な用途が考えられ，大量の例文からのデータ収集に関する研究報告や実際の収集データが公開されている[20),21)]．

2.3.2 格文法解析の利用

語と語の係り受け関係が求まれば，あらかじめ各単語にもたせた単語間の結合性に関する判定知識をより正確に適用させることができる．

日本語文に対する係り受け解析に取り組む手法としては，一般に動詞の格支配の考え方が利用される．たとえば「読む」という動詞が，“(人) が (書物) を (場所) で”のようにそれぞれの格を特徴付ける助詞を介して特定の意味範囲の名詞と結び付けられる現象を利用したものである．各名詞に意味範囲を示す分類符号を与え，動詞には適合可能なすべての意味範囲の分類（一般に複数存在）とその動詞と結び付くときの格関係を記述しておく必要がある．また解析時には入力文の構文構造を分析し，得られる多数の単語候補の組合せの中からこれらの格支配の条件に基づいて最適な解釈を求め，正しい文節境界の判定や各同音語の判定を行う．このように入力文を詳細に分析するには大規模な意味辞書や複雑な処理手順を要するので，ハードウェア上の制約もあって本格的な実用化は今後の課題である．関連の研究発表を参考文献にあげておく[22)〜25)]．なお単語の意味分類を体系化した一般的な資料には，分類語彙(い)表[26)]や類語辞典[27)]などがある．また，(株)日本電子化辞書研究所 (EDR) では，概念間の結び付きを表す概念記述と呼ばれる辞書，概念を階層化した概念体系と呼ばれる辞書や単語共起辞書といった大規模の知識データを開発している[28)]．

2.3.3 文脈情報の利用

上記に述べた方式は与えられた文や句を正しく日本語として解釈するための手法であり，いわば静的な取組み方であるのに対し，文章全体の特徴や話題に関する情報を得て同音語の優先度判定に利用する動的な取組み方の例を紹介する．

動的な対応としては使用者が最も最近選択した候補を同音語の中で優先させる学習機構が一般に

備えられている．これは文章中で同一の単語が再び出現する可能性が高いことや送りがなどの異表記をもつ同義語の表記を統一させる上でも重要なことである．しかし，文章中で初めて出現する単語についても的中率を確保したいとか文脈に応じて目的の単語が同音語の中で変化する例においては単純な学習のみでは対応できない．

その試みとして対象文の属する分野を推定しその分野で使用可能性の高い単語を優先させる方式がある[29]．辞書の各単語は分野情報が記述されており，同音語が選択されていくに従いおのおのの単語の分野情報を取り出して加算していき，次の変換時には加算値の最も高い分野の単語を優先出力させるものである．

また，対象文章の分野を判断するのに特徴的なキーワードによって行う方法もある[30]．特定の文脈や分野で出現頻度が高いとみられる単語を集めてグループ化しておき各グループの中で同音語をもたない単語をキーワードとしておく．この方式は単語の共起性という観点でみたとき，同一グループ中の単語相互の共起性によって同音語を判定しようとするものである．従来の2語間の共起性だけでは表せなかった話題性に関わる同音語の判定能力を補ったものである．

文脈に追随しながら同音語を判定する方式として，神経回路網（ニューラルネットワーク）の連想機能を用いた方式も実現された[31]．この方式では単語をノードに，単語間の関係をリンクに対応させたネットワークで文脈を表現し，次のようにして同音語の優先度を判定する．

各単語の選択されやすさ，すなわち優先される確度を対応するノードがもつ活性値で表す．また二つの単語の関係の度合いすなわち確率的に同時に現れる確度を重み付きリンクで示す．ある単語が選択されて確定するとそのノードの活性値が強制的に高められ，またその上昇分がリンクを伝播して関係するノードの活性値も高められる．具体的にはノード j の活性値 O_j は，ノード i と j とのリンクの重みを w_{ji} としたとき，以下のような計算で求める．

$$O_j \leftarrow f(n_j)$$

$$n_j \leftarrow (1-\delta)n_j + \delta(\Sigma w_{ji}O_i + I_j)$$

ただし　f：単調増加関数

　　δ：$0<\delta<1$ の実数

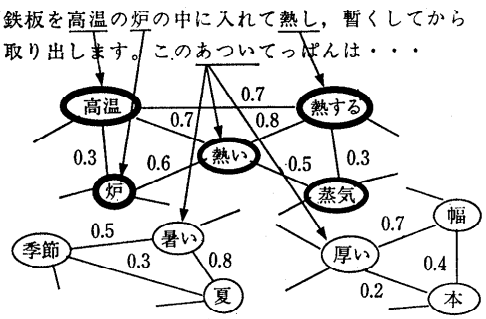


図-5　語の連想ネットワークの例

I_j：ノード j への外部入力

$$w_{ji}=w_{ij},\quad w_{ii}=0$$

ここで，関数 f の形式と δ は経験的に求めた最適値を適用する．I_j はノード j に対応する単語が選択されると正のある値になり，徐々に減衰するようにする．また文脈の移り変わりに応じてスムーズに活性値の分布が移行するように，リンクのない単語には仮想的に負のリンクを設定している[32]．このような動作によって，入力中の文章の文脈に強く関係する単語の活性値が高まり，その語が未出現であっても文脈に適した単語として優先して出力することができる．**図-5**はその例であり，ここでは「熱い鉄板」が文脈に即して選ばれる．この例では，共起語辞書のみで判定しようとしても「熱い鉄板」と「厚い鉄板」とが選ばれるので，このような文脈の情報を利用しないと両者間での優先順位がつけられない．

2.4 辞書管理

仮名漢字変換の機構で参照される知識データは，単語辞書，文法辞書，学習辞書および**2.3.1**に述べた共起語辞書などのデータファイルからなる．

単語辞書は入力文を解釈するための語彙を提供するもので，一般に国語辞典や事務文書，技術文献などから使用度の高い用語が選定されている．また，地名や人名などの固有名詞も含まれる．収容用語の総数は固有名詞を除いておよそ5，6万語程度あれば一般的使用に対応でき，さらに専門用語など特殊なものは分野別に別途提供したり，ユーザ登録機能を活用して補うことができる．

辞書の記述項目は，読み仮名の見出しに対する正規の表記と，品詞などの文法属性，その他頻度情報などからなる．用言は活用語尾が規則によって派生できるので語幹部分を登録する．同様に連

用形名詞や可能動詞などの他品詞への派生も規則化することで辞書メモリを有効に使える．さらに，読みや品詞が共通な単語をグループ化したり，隣接単語どうしで共通する文字部分を省略して表すなどデータ圧縮上の工夫がある[33]．また，入力と並行して少しでも高速に検索ができるように見出しを逆引き順に並べ替えておく方法もある[34]．

辞書の大語彙化については，実際に77万語の辞書を試作した結果，10万語の辞書と比較して未登録語が約6%減少し約10%正解率の向上（共起語辞書による効果を除く）があったと報告されている[35]．

文法辞書は**2.1**の接続表などの規則であり，学習辞書としては**2.3.3**で述べた同音語内の最新選択語を記憶する機構が一般的である．また，連文節変換の文節分かち結果に対して使用者が強制的に変更した新しい区切り方を記憶し学習に利用する例もある[36],[37]．

3. 文章作成支援機能

仮名漢字変換機能以外で自然言語処理に関連したものとしては以下のような機能がある．

3.1 校正・推敲支援機能

文章の入力，編集操作に起因した誤りを能率よく正して高品位の文書に仕上げるための支援機能である[38],[39]．単語分かちのない日本語文に対しては欧文での単語スペルチェックのような機構はそのまま適用できない．単語を識別するためには仮名漢字変換と同様に形態素解析を行って自動分かちする．そのための単語辞書も仮名漢字変換辞書とは別に提供される．解析結果のあいまい性はあるが辞書未登録語や文法的不ぐあい部分を指摘する．また，文末表現に着目して「です・ます」と「だ・である」の形態の混在を指摘したり書き換えて統一させる．あるいはあらかじめ送りがなの誤りや「人口知能→人工知能」のような誤用と正解との書換え例を収録しておき，文章中の文字列と直接照合させて指摘する実現例もある．長すぎる文や漢字列の警告，括弧の対応誤りの指摘などの形態的特徴をとらえた処理も行っている．

なお，文章の校正・推敲支援機能についての詳細は，本特集別稿を参照されたい．

3.2 文書レイアウト支援

標題，章，節，箇条書きなどの文書の論理的な構造を抽出して，統一した文書レイアウトに変換する機能が実現されている[40],[41]．文書構造の論理関係は，形態的な特徴により大見出し（章），中見出し（節），小見出し（項）といった階層構造として把握する．各見出しの決定はあらかじめ登録されているキーワードとの照合や見出しの先頭にある数字，記号などに基づいて行う．たとえば，「1. はじめに」で始まる行があったときに，キーワード「はじめに」を抽出し，その前部分にある「1.」とともに見出しを構成していると判断する．次に出現する見出しは「2.…」の形態であると予測され，その行までの記述部分は最初の見出しに対する段落部分（内容部）であると判断される．同様に「2. 1…」で始まる表記があれば一段下位の段落になる．解析された構造に従い，一定数の左余白を挿入して見出し部分や段落部分の配置を定めたり，また統一的に見出し部分に下線や書体変更などの属性を付加して文書の体裁が整えられる．見出しを判断するキーワードや配置に関する定義情報は，ユーザ登録ができるようにして多様な文書様式に対処している．

4. おわりに

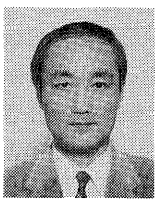
自然言語処理の技術の応用として，主に商用ワードプロセッサで実際に採用されている機構について概説した．中でも日本語文章入力のための仮名漢字変換に関する取組みの方式を中心に紹介した．現在でもより高機能を目指した仮名漢字変換の研究が各所で行われている．変換正解率を高めることは，不自然な結果を同音語候補の中からできるだけ除外できるようにさせることでもある．そのためには文法や意味の知識をいっそう精密に作り上げていく必要がある．また，電子化された文書ファイルがますます増大する中で，本稿で述べたことのほか，たとえば表現の意味内容に基づいた検索やあるいは外国語の文書作成のための支援機能などは今後の研究開発に期待されるところである．使い勝手のよいさまざまな応用機能がワードプロセッサのような身近な機器に取り入れられていくよう今後の技術発展に期待したい．

参考文献

1) 栗原，稲永：カナ漢字変換（I），九大工学集報，Vol. 42, No. 6, pp. 880-884 (1970).
2) 松下，山崎，佐藤：漢字かな混り文変換システム，

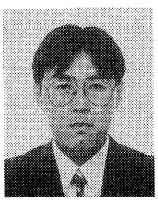
情報処理, Vol. 15, No. 1, pp. 2-9 (Jan. 1974).
3) 藤井, 原, 木村：カナから漢字への変換, 電気四学会連合大会, 177. pp. 619-622 (1971).
4) 相沢, 江原：計算機によるカナ漢字変換, NHK技術研究, Vol. 25, No. 5, pp. 261-298 (1973).
5) 木村, 遠藤, 小橋：日本語文入力用カナ漢字変換システムの試作, 情報処理, Vol. 17, No. 11, pp. 1009-1016 (Nov. 1976).
6) 森, 他：計算機への日本語情報入力, 電子通信学会技術報告, EC 78-23, pp. 33-41 (1978).
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) 内田, 杉山：自由入力形式のカナ漢字変換, 情報処理学会自然言語処理研究会, 27-3 (1981).
9) 吉村, 日高, 吉田：文節数最小法を用いたべた書き日本語文の形態素解析, 情報処理学会論文誌, Vol. 24, No. 1, pp. 40-46 (Jan. 1983).
10) 小渕, 他：一括自動かな漢字変換, シャープ技法, No. 33, pp. 134-138 (1985).
11) 首藤, 柳沢, 檜垣：自動かな漢字変換, NEC 技法, Vol. 40, No. 4 (1987).
12) 藤田, 林：長さ, 頻度, 接続重みを用いたかな漢字変換方式, RICOH TECHNICAL REPORT, No. 15, pp. 13-19 (1986).
13) 斎藤, 他：ベター入力かな漢字変換：情報処理学会第30回全国大会, 1G-7 (1985).
14) 中野：同音語の判別, 情報処理学会自然言語処理研究会, 33-4 (1982).
15) 山崎, 橋本, 平塚：同音語処理の一方式について, 情報処理学会第29回全国大会, 5J-7 (1984).
16) 牧野, 木澤：べた書き文の仮名漢字変換システムとその同音語処理, 情報処理学会論文誌, Vol. 22, No. 1, pp. 59-67 (Jan. 1981).
17) 空閑, 他：かな漢字変換における意味処理の一方法, シャープ技報, No. 23, pp. 35-40 (1982).
18) 長崎, 他：表層格を用いた共起変換の一実現方法, 情報処理学会第 39 回全国大会, 4F-7 (1989).
19) 本間, 山階, 小橋：連語解析を用いたべた書きかな漢字変換, 情報処理学会論文誌, Vol. 27, No. 11, pp. 1062-1067 (Nov. 1986).
20) 田中：語と語の関係解析用資料—“を”を中心とした解析編・資料(I)(II), 文部省科学研究費特定研究「言語情報の高度化」総括班 (1987).
21) 田中, 吉田：共起データの今後の研究について, 情報処理学会第40回全国大会, 5F-1 (1990).
22) 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. 679-687 (July 1986).
23) 山田, 大山：結合価を用いたかな漢字変換, 情報処理学会第36回全国大会, 2T-4 (1988).
24) 後藤, 他：深層格を用いた仮名漢字変換, 情報処理学会第41回全国大会, 1S-8 (1990).
25) 山内, 他：表層格情報を用いたべた書き仮名漢字変換, 情報処理学会第43回全国大会, 7H-1 (1991).
26) 国立国語研究所：分類語彙表, 秀英出版 (1973).
27) 大野, 浜西：類語新辞典, 角川書店 (1981).
28) EDR 電子化辞書仕様説明書, (株)日本電子化辞書研究所 (1993).
29) 増田, 市村, 竹中：日本語ワードプロセッサにおける, かな漢字変換 III：情報処理学会第22回全国大会, 1I-3 (1981).
30) 山本, 久保田：共起グループを用いたかな漢字変換：情報処理学会第44回全国大会, 4P-11 (1992).
31) 鈴岡, 他：神経回路網の連想機能を用いたかな漢字変換方式, 情報処理学会第40回全国大会, 1C-3 (1990).
32) 小林, 中里, 長崎：ニューロかな漢字変換の実現, 東芝レビュー, Vol. 47, No. 11, pp. 868-870 (1992).
33) 西, 他：かな漢字変換辞書圧縮方式, 情報処理学会日本文入力方式研究会, 17-1 (1984).
34) 武田, 他：ベター入力かな漢字変換方式のワープロへの実現, 情報処理学会第30回全国大会, 1G-8 (1985).
35) 山田, 大山：大語彙かな漢字変換方式の変換率評価, 情報処理学会自然言語処理研究会, 87-5 (1992).
36) 古和田, 他：かな漢字変換における学習機能, 情報処理学会第39回全国大会, 3J-1 (1989).
37) 横田, 他：かな漢字変換における文節切り直し学習機能, 情報処理学会第39回全国大会, 3J-1 (1989).
38) 浅見：使いものになるか, シーズ先行で校正支援機能付きワープロが4社から, 日経エレクトロニクス, No. 443, pp. 183-188 (1988).
39) 田中, 山添, 松原：OASYS トータル文書処理システム, FUJITSU, Vol. 41, No. 3, pp. 264-270 (1990).
40) Iwai, I. et al.: A Document Layout System Using Automatic Document Architecture Extraction, CHI '89, pp. 369-374 (1989).
41) 原, 田野崎, 橋本：日本語ワードプロセッサ Rupo における自動文書編集機能, 東芝レビュー, Vol. 47, No. 11, pp. 874-876 (1992).

(平成5年6月3日受付)



斎藤 裕美（正会員）

昭和25年生. 昭和50年早稲田大学大学院理工学研究科電気工学専攻修士課程修了. 同年東京芝浦電気(株)(現, (株)東芝)入社. 以来仮名漢字変換, 推敲支援, 機械翻訳などの自然言語処理システムの研究開発に従事. 現在同社情報処理・機器技術研究所主査. 電子情報通信学会会員.



野上 宏康（正会員）

昭和34年生. 昭和58年九州大学大学院総合理工学研究科情報システム学修士課程修了. 同年東京芝浦電気(株)(現, (株)東芝)入社. 以来仮名漢字変換, 機械翻訳などの自然言語処理システムの研究開発に従事. 現在同社研究開発センター情報・通信システム研究所研究主務. 電子情報通信学会, 人工知能学会各会員.