

解説



日本語の構造とその解析†

首藤 公昭^{††} 吉村 賢治^{††}

1. はじめに

漢字入出力技術の進歩、記憶装置の大容量、低価格化、処理の高速化、ソフトウェア技術の進歩など、日本語の機械処理を支える計算機側の環境はかなり整ってきており、今後一層の充実が予想される現状から日本語の機械処理における問題点は、言語現象を明らかにする計算言語学的側面に絞られてきた感がある。現在、日本語ワードプロセッサの普及が進み、日本語を対象とする機械翻訳システムの商品化も行われるなど、日本語処理への期待がますます膨らんでいるが、言語の現象は複雑、膨大で、形態レベル、構文レベルの現象も必ずしも十分には捉えられていないのが実情である。意味処理、文脈処理などを取り入れた将来の各種日本語処理システムを実現するためには、意味、文脈に至る前段階として、できるだけしっかりした形態、構文処理技術を確立しておかなければならない。また、この技術は意味処理との整合性が良く、日本語の膠着性、非成層性 (non-configurationality)^{††} といった欧米語と異なる特徴をふまえたもので、かつ広範な表層的現象に耐えうるものであることが望まれる。

以上の点を念頭におき、本稿では、主として日本語の形態素解析、構文解析の段階の基礎的事項に焦点を絞り、筆者らが行ってきた基礎研究および関連の研究を紹介し、問題点を論じる。これらの段階の上に積み上げられる意味解析の細部や文脈解析などについては本特集の他稿にゆだねたい。

2. 単語とその機能

膠着性の強い日本語を機械処理の対象とする際の基本的な問題は文の最小構成単位をどう捉えるかである。単語という概念の定義や品詞についても国語学で古くから議論されているところであるが²⁾⁻⁵⁾、現在な

お細部については不明瞭な点が多い。既存の国語辞典類でも不統一な点が多く、そのままでは機械処理にとって使いやすいものではない。

通常の単語が持つ意味上の基本機能としては、大きく、

① 概念情報 (名詞性、動詞性、…に細分される。) を与える機能

② 概念間の関係 (〈主体〉、〈対象〉、…) の情報 (関係情報) を与える機能

③ 話者の判断、態度や、テンス、アスペクトなどの情報 (助述情報) を与える機能

④ その他の付加的意味情報を与える機能

があり、文中の他の単語との係わり方に関する文法的機能は、

(A) 隣接した単語だけと係わりを持つ形態的機能

(B) 隣接した単語に限らず、一般に他の単語と係わりを持つ構文的機能

に大別される。通常の国文法は以上の機能を細分、総合して品詞などの概念を得、語形変化情報 (活用の有無、種類など) を加えて文構造を明らかにしようとするものであるが、機械処理、特に意味を考慮した形態、構文処理にそのまま使用できるほどの整合性は備えていない。機械処理の立場からは以上の機能を処理の方式に照らして明確に指標化し、これによって通常、国語辞典などに収録されている単語や実際の文中に現れる文字列を改めて見なおす作業が必要となる。

ここでは、このようなことも考慮に入れ、広範な表層的現象を処理対象とする一般的な日本語の構文、意味解析を行う立場で、単語 (的に扱う表現) の認定について、特に複合語と慣用句を中心に考察する。なお以下では、①、②、③の機能を持つ表現 (通常の単語を含む) をそれぞれ概念表現、関係表現、助述表現と呼ぶ。

2.1 複合語

通常、造語成分と呼ばれているものを構成要素とす

† The Structure of Japanese Language and Its Mechanical Analysis by Koshi SHUDO and Kenji YOSHIMURA (Department of Electronic Engineering, Fukuoka University).

†† 福岡大学工学部電子工学科

る複合語(派生語)をすべて単語とみなして辞書に登録することは量的にも、また、システムの拡張性の点からも難しい。また、これらすべてを切り離し、ルールによって複合語を解析することも例外的ルールをいわずらに増やす結果となる。したがって、これらは、含まれる造語成分の造語機能の大きさと一般性を判断し、機能が大き、一般性のある機能によって造語されている(造語の仕組みがシステムの枠組でルール化できる)と認められる場合は、造語成分とその他の部分を分離してそれぞれ別の単語として登録することになる。(たとえば、「回路・的」、「大・会社」、「名詞・性」、「大人・向き」、「同・大学」)この段階の作業で実際に造語成分として取り扱うべき表現が明確になる。

また、含まれる造語成分がほかの複合語を派生する事例が少ないと判断できる場合や、通常の造語機能と異なった用法と認められる場合、あるいは複合語としての使用頻度が高いものは単語としてそのまま登録する。(たとえば、「ほの白い」、「神経質」、「目的」、「大人」、「逆境」)これらの規準の設定は、残念ながらある程度恣意的にならざるを得ない。たとえば、機械翻訳においては訳語の有無、訳語句の合成方法などが規準となり得るなど、規準の設定は処理の目的、方法に強く関わってくるため、一般的かつ厳密に論じることは難しい。

複合語には、「臨教審」、「車検」、「京浜」などのような、省略を伴って派生しているものがあるが、これらは現時点では単語として登録せざるを得ない。以上の考え方で接頭語、接尾語の機能も改めて見直す必要がある。

2.2 慣用句

複数単語からなる列で、単語の並びが固定化して用いられ、まとまった形態、構文、意味上の機能を持ち、さらに、この機能の上でそれ以上分割することのできないものを一語性慣用句(あるいは単に慣用句)と呼ぶ。これらの表現はかなり頻繁に使用されており、広範な現象を対象として意味処理まで行う際には単語的な要素として収集・整理しておく必要がある。また、慣用句には特有の文法機能を持つものがある。たとえば、「かもしれない」は名詞にも用言にも接続可能で、活用する助述表現であるが、通常の単語にはこの種のものは見られない。構成要素語の文法機能から慣用句の文法機能が導けない場合もある。たとえば、関係表現「について」の場合、「について・から」という連

接は不可であり、接続助詞「て」とは異なる接続機能を持つ。通常の単語にこれらの慣用句を加え、文法機能の総括的整理を行い、同一水準の機能設定によって、自然な形で(例外処理でなく)慣用句を文法体系に組み込むことが望ましい。

慣用句としては、たとえば次のようなものがある。

A) 概念表現

「背中合わせ」、「木の間隠れに」、
「まわりくどい」、「小気味よい」、
「薄気味悪い」、「くそまじめ」、「ばか正直」、
「生意気」、「しゃべりたてる」、「うち滅ぼす」、
「わめきちらす」、「出し抜く」、「開き直る」、
「遅ればせながら」、「可能な限り」、
「人煙まれな」、「文字どおりの」、「この結果」、
「そのことにより」、……

B) 関係表現

「によって」、「に基づいて」、「における」、
「などといった」、「にもかかわらず」、
「ばかりではなく」、「に伴って」、
「のあいまに」、……

C) 助述表現

「ねばならない」、「にちがいない」、
「のようである」、「かのようだ」、
「べきだ」……

D) 接尾語的慣用句

「である」、「か否か」、「なのか」、……

このように慣用句を収集・整理しておくことは言語における例外的現象の一部を辞書に吸収し、処理手続きの負担を減らす効果がある。特に、非概念性の、B)、C)、D)は量的に膨大ではなく、これらを網羅しておくことは処理可能な表現の枠を拡大することにつながる。筆者らは、日本語文の枠組を与える表現として、当面、機械翻訳をめやすとして、B)約1,000語、C)約3,000語、D)約200語を収集・整理している^{6),7)}。

2.3 その他の問題

「行う」と「行なう」、「関わる」と「係わる」、「プロセッサ」と「プロセッサー」など送り仮名、漢字、外来語については表記のゆれがみられ、辞書構成上の問題となる。また、「美しき人」、「行かねばならぬので」といった古語的表現や、「行ってる」、「お使い下さいませんか」などの丁寧語や会話特有の表現を処理可能とする際の問題もある。

3. 名詞連続

漢字造語成分や、漢字名詞からなる列が、日常頻繁に用いられているがその構造はまだあまり解明されていない。これらの解析には、(1)単なる接続で構成された形態的構造体として荒く捉える立場と、(2)概念間の意味的關係によって構成された構文的、意味的構造体と捉える立場がある。現在のべた書き文を対象としたかな漢字変換では(1)の立場を採るものが多い。森らは、辞書の見出し語とのマッチングと分割数が少ない結果を優先するヒューリスティクスによる漢字列の単位切りを試みた⁹⁾。長尾らは、さらに一字漢字造語成分の造語力を統計データから数値化し、これを用いた漢字連続の単位切りを行って良好な結果を得ている⁹⁾。また、マルコフ・モデルに基づいて単位切りを試みる武田ら¹⁰⁾や、松延ら¹¹⁾の研究もある。語構成や漢字連続などに関しては国語学サイドからの調査・研究もいくつかある^{12), 13)}。

単位切りに高い精度が要求される場合や、かな漢字変換において漢字候補を正しく絞り込む場合、あるいは機械翻訳で名詞列の訳出に正確さを期す場合などは(2)の立場をとる必要が生じる。また、将来の意味理解などでは(2)の立場をとることは不可欠となる。名詞の複次結合の仕組みは通常の文や句の構成の仕組みと似ており(5章参照)、解析においては係り受け解析法や句構造文法に基づく方法が考えられるが、名詞の間に関係表現が介在しないだけ意味処理は難しくなる。宮崎は、分割数最小法と比較的簡単な係り受け解析によって名詞列の単位切りを試みている¹⁴⁾。今後は、このように(2)の立場での研究が推進されるものと思われるが、語と語の關係にどのようなものがあり、これをどう規定するかという問題は一般には未解決で、将来の重要な研究課題である。

4. 文節、拡張文節

4.1 文節、拡張文節の構造

日本語の機械処理では、5章で述べる日本語文の特徴から文の構成単位を文節とする橋本文法の考え方²⁾が広く用いられている。文節は基本的には2章で述べた単語の形態的機能による結合体であり、その構造は、概略、次のように示される。

〈文節〉＝〈自立部〉|〈文節〉|〈付属語〉
 〈自立部〉＝〈接頭語〉|〈自立語〉|〈接尾語〉
 ()は省略可であることを示す。

自立語は概念情報を与え、付属語は関係情報、または助述情報を与えるもの、文節は基本的には日本語文の構文・意味解析を行う上で重要な情報を提供する基本単位と考えられるが、現実の文では、必ずしもこのような明確な情報が与えられない文節も多い。(「その本について述べる」、「開発における問題」など)しかし、2章で述べたように慣用句を単語的に扱えば、このような文節をより意味機能の明確な文節内に吸収することができる。この立場で文節の考えを一般化したものに拡張文節がある^{6), 7)}。

文節、あるいは拡張文節の構造は、通常、要素語間の接続を規定した有限オートマトン・モデルで捉えられる。拡張文節の例として筆者らのモデルの概略を図-1に示す。ただし、接頭語の部分を省き、名詞連続の部分は3章で述べた(1)の立場で荒く近似している。図-1のノードは大まかな文法機能上のカテゴリを示し、ノードをつなぐ矢印が単語の接続可能性を示す。初期ノードから出発し、矢印に沿って順次ノードを遷移する経路が原則として拡張文節に対応する。破線は自立部と付属部の境界、*印は活用に関する接続チェックの必要なことを示している。実際には2章で述べた単語の機能の細かな見直しを行って約140個にカテゴリの細分化を行い、図-1を大幅に精密化したルール体系となっている。また、たとえば、関係表現「ために」は〈原因・理由〉と〈目的〉の二つの意味があるが、「行ったために」など、助動詞「た」に接続した場合は、〈目的〉の意味は取らないなど、接続機能と意味機能の關係も調べ、意味の相違を接続ルールに反映させることで、形態素解析の段階でできるだけ多義選択を行えるようにしている。

このモデルによれば、たとえば、次のような語列が拡張文節とみなされる。/は自立部と付属部の境界を表す。

値/を基に
 容易さ/についても
 人である/かもしれず
 求めること/についてであったが
 効果的な/のだろうかなどと
 行っ/たことがあるだろうか
 ゆっくりと/さえ

現在、このモデルは長期にわたって増補・改良を施されつつある。

以後、従来型の文節と拡張文節を区別なく文節と呼ぶことにする。

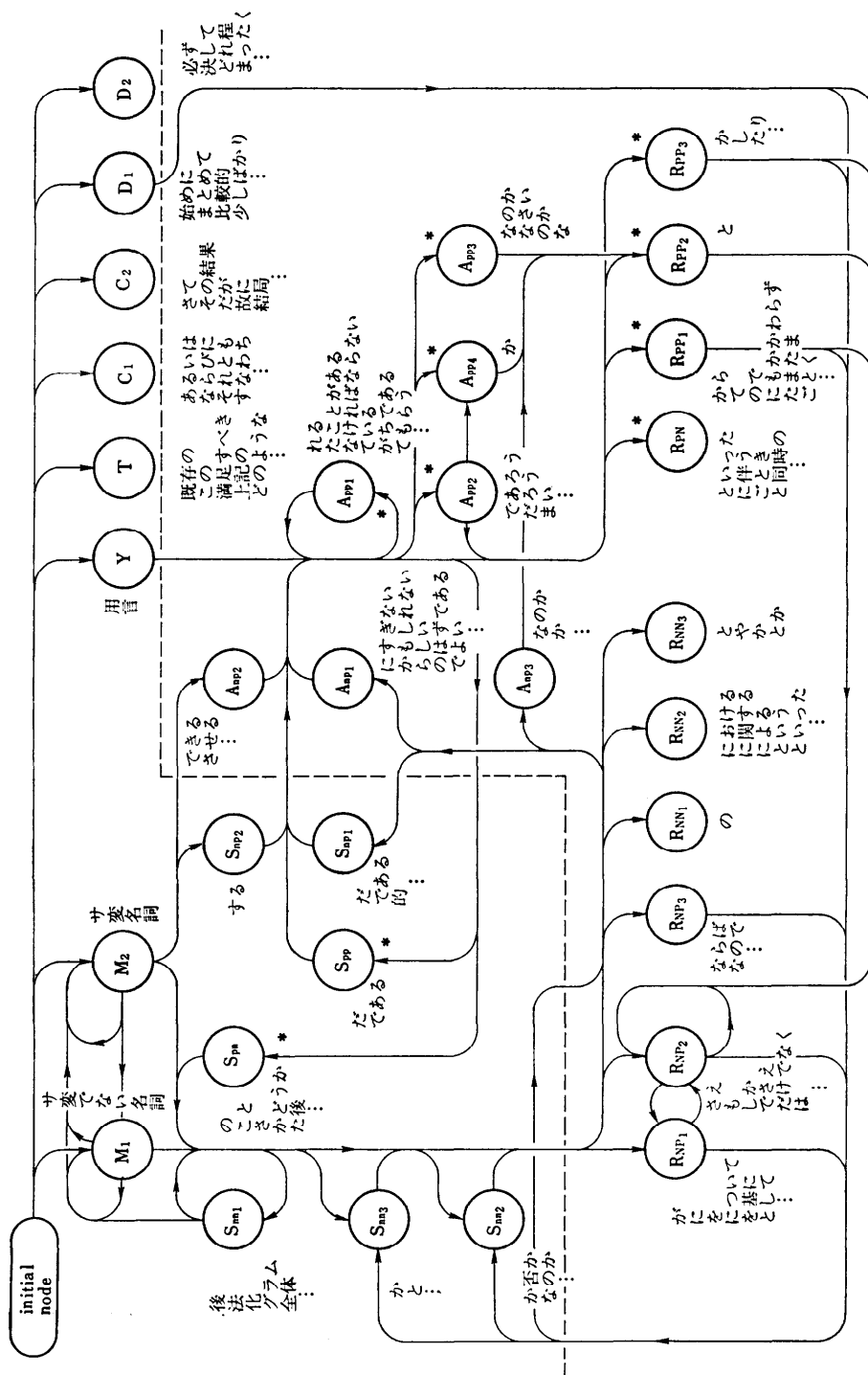


図-1 拡張文節の構造

4.2 形態素解析

文節の構造規定に基づき、文節、文節連続、あるいは文を単語に分割し各単語の品詞などを認定することを通常、形態素解析と呼んでいる。自然言語の解析における大きな問題は解析結果の曖昧さをいかに減らすかということであり、形態素解析でもこのために種々の工夫がなされる。形態素解析のアルゴリズムには大きく縦型探索と横型探索とがある。縦型探索のアルゴリズムとしては、単語の候補が複数個存在する場合に一致した長さの長いものから優先して解析を行う最長一致法や長さの評価を文節で行う牧野らの文節最長一致法¹⁵⁾などがある。また、牧野らは、文節最長一致法における解析の精度を改善するために縦型探索の枠組で二文節内の局所的な横型探索を行い、二文節の長さで文節の候補を評価する二文節最長一致法¹⁶⁾を提案している。横型探索のアルゴリズムとしては、入力文を構成する可能性を持つすべての単語からなる表を作成し、その表を参照することにより文節の数が最も少なくなる解析結果から優先して出力する吉村らの文節数最小法¹⁷⁾がある。これらのアルゴリズムはかな表記のべた書き日本語文を対象としたもので、多くの解析結果の中から正しいものを早い時期に求めるために、それぞれ、単語の長さ、文節の長さ、連続する二文節の長さ、解析結果の文節の数に関するヒューリスティクスを用いている。一般に領域計算量と最初の解を求めるための時間計算量においては縦型探索のアルゴリズムのほうが優れているが、弱いヒューリスティクスを用いて正解を求めるときの時間計算量では横型探索のほうが優れている。

これらのアルゴリズムは漢字かな混じり日本語文の形態素解析にそのまま適用することもできるが、この場合には文節の認定に字種情報を利用することができる⁹⁾。現在、筆者らはヒューリスティクスとして文節数最小法を一般化したコスト最小法を、文法モデルとして4.1で述べた拡張文節モデルを用いた形態素解析システムを作成している。このシステムでは解析表の展開を多段階に制衡することにより効率の向上が図られている。二文節最長一致法が縦型探索の枠組で局所的な横型探索を行っているのに対して、このアルゴリズムは横型探索と縦型探索を融合したものになっている。縦型探索ではバックトラックの際に生じる単語辞書の重複した検索が処理の能率を著しく悪くするが、このアルゴリズムではこのような重複した検索を回避する工夫もなされている¹⁸⁾。

5. 文節構造から文構造へ

形態素解析によって得られた構成単位の情報から、与えられた文法規則に基づいて文の構造を決定する処理を構文解析という。語順の制約が強い英語などの文法規則はCF(文脈自由型)文法に基づくことが多い。文法規則がCF文法で記述されているときの構文解析には、能率の良いCocke-kasami-YoungerのアルゴリズムやEarlyのアルゴリズムなど¹⁹⁾を利用することができるが、実際にはCF文法をなんらかの手段で強化して用いることが多い。英語などの構文解析システムとしてはWoodsのATNパーサ²⁰⁾やPrattのLINGOL²¹⁾がよく知られている。日本語の構文解析システムとしてはATNにパターンマッチングと柔軟なバックトラックの機能を追加した長尾らのPLATON²²⁾やLINGOLに予測制御機能と分かち書きの機能を追加した田中らの拡張LINGOL²³⁾がある。また、論理型のプログラミング言語PROLOGを用いるPereiraらのDCG²⁴⁾を日本語の解析に適した上昇型に改良したものに松本らのBUP²⁵⁾がある。しかし、最近実用化されつつある機械翻訳システム²⁶⁾における日本語の構文解析はほとんどが日本語の特徴を考慮した係り受け解析に基礎をおいて、上記のCF文法に基づくシステムは大規模な日本語解析にはあまり利用されていない。

日本語文の構成単位を文節と考えれば、日本語文の一般的特徴として次の点があげられる。(以下では単一の文節も句と考えている。)

(1) 文を構成する句には、その末尾に構文・意味上の代表となる主要素(head)が必ず存在する(一種の内心性, endocentricity)。

(2) 句と句の結合は文頭側の句が文末側の句に意味的に依存する関係が成立するときに許される。この関係は両句の主要素間の係り受け関係として近似的に捉えられ、多くの場合、係り側の文節の末尾の関係表現(あるいは活用形)によってその手がかりが与えられる。

(3) 一つの句に依存する数個の句の間には階層的關係が見られず、配列の順序はかなり自由である(非成層性)。

(4) 現実の文においては、述語性の句が必ずしも特定の依存句(主語、目的語など)を要求しない。したがって、欧米語と比べて、文、非文の区別が明瞭でない。

このように文節を構成単位と見れば、日本語文は依存文法²⁷⁾の考え方を取り入れやすい構造を持っている。以上の特徴に基礎をおき、各文節（自立部）間の矛盾のない係り受けのパターンを検出するのが係り受け解析法である^{28), 29)}。係り受け関係は係りの語が受けの語になんらかの意味的な属性値を付与する関係と考えられ、この解析法はCF文法に基づく方法に比べて意味解析に、より密着した解析法といえる。二単語間の係り受けを文法レベルで荒く捉えることは比較的容易であるが構文解析結果の曖昧さが一般には膨大となり、なんらかの意味的な制約を設けて解析結果を絞る必要がある。また、機械翻訳や意味理解においては意味による係り受けの規定は不可欠となる。以下ではこの点を認識しながら主な問題点について述べる。

5.1 連用修飾関係

連用修飾関係は広い意味の格関係と考えることができる。Fillmoreの格文法では8種類の格で基本的な考え方が示された³⁰⁾が、実際の日本語文では格助詞以外にも2章で述べたごとく多くの連用修飾を指示する関係表現が用いられている。また、一つの関係表現（格助詞）が微妙に異なった意味にも用いられる。（たとえば、「庖丁で切る」、「顔色で判断する」、「その仕事で生活する」、「この速度で走る」）したがって、機械処理においては、どこまで細かく関係表現の意味を区別し、格として柱を立てるかが最も基本的かつ重要な問題である。しかし、これは現時点では難しい問題で、処理の目的、方法などによって設計者がかなり直観的に定めざるを得ない状況である。島津らは機械翻訳を目的として大きく6グループ、細かく29種の格の種類をあげた³¹⁾。筆者らは、同様の目的で対象となる関係表現を網羅的に抽出・分類し、100種前後の意味の種類をあげている⁷⁾など。通常、我々は格助詞を用いる代わりに慣用句的な関係表現を用いて係りの意味を明確にすることがある。（たとえば、「会議で休む」の代わりに「会議を理由に休む」など）この点に着目し、これらの慣用句を積極的に用いて日本語を規格化する研究も行われている³²⁾。辻井らは多岐にわたる格助詞の意味の相違をより基本的な条件の組によって明確化する試みを行っている³³⁾。このような比較的少数の普遍的な条件を定めることができれば、工学的にも言語学的にも非常に有意義であろう。また、稲永らは体言と用言の係り受けを単語（あるいは一字漢字）のレベルで捉え、膨大な係り受け辞書を作成している³⁴⁾。これは一種の用例辞書であり、将来の応用が目ま

れる。

係り受け解析法は基本的には語と語の2項関係を把握する方法であり、そのままでは、「象が箸で飯を食う」などの意味上の不自然さを捉えることが難しい。そこで、日本語の特徴(3)を考慮し、述語に対して、それぞれの格を埋める名詞の資格をn項関係で同時に規定する格フレーム³⁰⁾や結合価文法³⁵⁾の考え方がよく用いられる。この考え方によれば一つの述語が同一の格を複数個取ることはないという条件（単文異格の原則）のチェックも容易になる。しかし、構文解析における格フレームの構成法やその共起制限の効果については一般的な評価はまだなされていない。そのほか、細かくは、「明日5時に行く」は格の重複と見るのか否かといった問題や、「彼が手が長い」などの形式上の二重主語の問題、あるいは、「彼はかせいだ金を使った」の「かせいだ」と「使った」の両方の主語が「彼」である場合の係りの二重性の問題など、構文・意味解析における格の取り扱いについては検討を要する問題が数多く残されている。

5.2 連体修飾関係

連体修飾の場合も二単語間の係り受けでは句の結合関係が正しく捉えられない場合がある。たとえば、「目の青い人形」の意味は「青い」と「人形」の関係だけでは捉えられず、係りの句「目の青い」全体と「人形」との意味関係を把握することが必要となる。また、係りの句全体の意味が分かっても受けの名詞との関係が捉えにくい「魚を焼く煙」、「1カ月働いたお金」などの例もある。また、「工場が進出する見通し」、「鋼板をプレスする工程」、「価格が安定する効果」などのように係りの意味が格関係を基本として定めにくい場合もある。助詞「の」を介した名詞の結合や、3章で述べた名詞連続においても同様に名詞間の意味上の関係に隔たりがあって簡単には関係を規定しておけない問題が多い。これらの問題を解決するためには、多くの場合かなり詳細なソーラスや対象世界のモデル、これらを用いた推論の機構が必要となり、今後の重要課題である。「の」の意味に関しては島津らの考察がある³⁶⁾。

連体修飾を文法上マークするものとしては、上記のごとく活用語の連体形や助詞「の」が基本的なものであるが、それ以外に「における」、「に基づく」、「関する」などの慣用句もよく用いられる。前述の拡張文節モデルではこれらの表現も単語として組み込まれている。

5.3 慣用句

2章で形態素レベルで扱われる慣用句について述べたが、二語性、三語性の慣用句については、構文レベルで取り扱うことになる。これらは、「手を焼く」、「頭をかかえる」などの文形式をとるものや、「決して……ない」、「どんなに……でも……ない」などの構文的枠組に関するものがある。日本語は、特徴(3)、(4)から、構文解析において下降型の予測機構が取り入れにくい言語であるが、この種の強い係り受け関係を利用して解析の曖昧さを減らすことが考えられる。これらの表現を網羅的に収集・整理しておくことも重要である。陳述の副詞と助動詞との呼応については栗原らの調査がある³⁷⁾。

5.4 並列構造

自然言語の文構造を乱すものとして並列(等位)構造があり、言語学においても古くから問題視されてきた。日本語の構文解析においても並列構造については特別の扱いが必要となる。日本語の並列構造は並列要素の種類によって大きく、(I)文形式の並列、(II)名詞句の並列、(III)その他の並列(「動作時消費電力 600 mW、待機時が 22 mW である」など)に分けられる。並列構造の解析に関しては(I)と(II)を対象として解析のためのヒューリスティクスを求めた長尾らの研究³⁸⁾や筆者らの(I)、(II)、(III)を対象とした大量データに基づく調査・考察がある³⁹⁾。並列構造の認定では、上記(III)の例のように並列要素が文法上相当異なった構造を持つ場合もあり、並列要素間の意味上の類似性をどう検定するかが本質的に重要であるが、当面は並列に固有の表現(「と」、「や」、「など」、「などと」と)と並列要素間の文法上および浅い意味上の類似性を検出し、なんらかのヒューリスティクスを援用して解析せざるを得ない。本格的な概念の類似性の研究は今後の大きな課題である。名詞句の並列には大きく、名詞句を列挙して共通な文要素の重複を避ける用法(「彼はりんごと柿を食べた」など)と、新しい複合概念を表す用法(「彼と彼女の関係」など)とがある。後者の用法の際に共起しやすい語句(「関係」、「組み合わせる」、「等しい」など)を整理しておき、構文解析に利用することが考えられる³⁹⁾。入力文に現れる並列句が上記のどちらの用法なのかを解析することも意味処理上は重要となる。

5.5 助述情報

テンス、アスペクト、モーダリティなどの助述情報は文脈処理などにおいて推論の制御に関わる重要な役

割を持つ。これらの情報は通常の助動詞だけでなく、「なければならぬ」、「かもしれない」、「べきだ」など数多くの慣用句によってももたらされ、基本的なものとして100種程度が数えられている⁷⁾。拡張文節モデルはこの種の表現を単語的に組み込んだものとなっており、このモデルによれば助述情報の抽出は基本的には単語の切り出しと同時的に行うことができる。しかし、助述表現が文形式の並列構造と共起した場合にはその意味機能が文中のどの領域まで及ぶかという問題(たとえば、「明日は晴天で人出が多いだろう」の「だろう」のスコープが「明日は晴天だ」まで及ぶか否かなど)が起り、これは文脈解析ともからむ今後の課題である。

6. むすび

本稿では、日本語文の解析において基礎的と思われる事項を概観した。ここで述べた問題点には20年近く前から指摘されていたものも多く、また、これらは現在山積する問題点の一部にすぎない。このように言語の壁は確かに厚いが、時間をかけた研究の蓄積によって現在のヒューリスティクスなどを多用した荒い近似処理を徐々に近似度の高いものにしていくことは可能であると思われる。このためにも認知科学、論理学、理論言語学、心理言語学、人工知能など、広範な領域と連繋した基礎研究の重要性、および言語現象を一つ一つ丹念に調査、分析する作業も含めた、言語を工学の見方で現象面から捉える基礎研究の重要性を指摘しておきたい。

参考文献

- 1) 今井邦彦: チョムスキー小事典, 大修館(1986).
- 2) 橋本進吉: 国文法体系論, 岩波書店(1959).
- 3) 時枝誠記: 国語学原論, 岩波書店(1941).
- 4) 松下大三郎: 標準日本語文法, 勉誠社(1974).
- 5) 山田孝雄: 日本文法学概論, 宝文館(1936).
- 6) 首藤, 檜原, 吉田: 日本語の機械処理のための文節構造モデル, 電子通信学会論文誌, Vol. J62-D, No. 12, pp. 872-879 (1979).
- 7) 首藤, 檜原: 日本語の文構造のわく組みを与える表現, 福岡大学総合研究所報, No. 63, pp. 1-52 (1983).
- 8) 森, 河田, 天野, 武田: 計算機への日本語情報入力, 信学技報, EC 78-23, pp. 33-41 (1978).
- 9) 長尾, 辻井, 山上, 建部: 国語辞書の記憶と日本語文の自動分割, 情報処理, Vol. 19, No. 6, pp. 514-521 (June 1978).
- 10) 武田, 鈴木, 藤崎: 漢字複合語自動分割の一手法, 情報処理学会第30回全国大会論文集, 5G-6,

- pp. 1687-1688 (1985).
- 11) 松延, 日高, 吉田: 単語の造語モデルを用いた複合語の自動分割法, 情報処理学会第 32 回全国大会論文集, 2 T-7, pp. 1651-1652 (1986).
 - 12) 阪倉篤義: 語構成の研究, 角川書店 (1966).
 - 13) 野村雅昭: 造語法, 岩波講座日本語 9, pp. 245-284 (1977).
 - 14) 宮崎正弘: 係り受け解析を用いた複合語の自動分割法, 情報処理学会論文誌, Vol. 25, No. 6, pp. 970-979 (June 1984).
 - 15) 牧野, 木澤: べた書き文のカナ漢字変換, 信学技報, PRL 77-22, pp. 65-72 (1977).
 - 16) 牧野, 木澤: べた書き文の分かち書きと仮名漢字変換一二文節最長一致法による分かち書き一, 情報処理学会論文誌, Vol. 20, No. 4, pp. 337-345 (Apr. 1979).
 - 17) 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).
 - 18) 吉村, 首藤: 制御機構を有する表方式の形態素解析, 福岡大学工学集報, Vol. 36, pp. 37-42 (1986).
 - 19) Aho, A. V. and Ullman, J. D.: *The Theory of Parsing, Translation, and Compiling, Vol. 1: Parsing*, Prentice-Hall (1972).
 - 20) Wood, A. N.: Transition Network Grammars for Natural Language Analysis, *Comm. ACM*, Vol. 13, No. 10, pp. 591-602 (1970).
 - 21) Pratt, V. R.: A Linguistic Oriented Programming Language, *IJCAI 3*, pp. 372-381 (1973).
 - 22) 長尾, 辻井: 自然言語処理のためのプログラミング言語 PLATON, 情報処理, Vol. 15, No. 9, pp. 654-661 (Sep. 1974).
 - 23) 田中, 佐藤, 元吉: 自然言語処理のためのプログラミングシステム—拡張 LINGOL について—, 電子通信学会論文誌, Vol. J 60-D, No. 12, pp. 1061-1068 (1977).
 - 24) Pereira, F. C. N. and Warren, D. H. D.: Definite Clause Grammar for Language Analysis—A Survey of the Formalism and a Comparison with Augmented Transition Networks, *Artif. Intell.*, Vol. 13, pp. 231-278 (1980).
 - 25) Matsumoto, Y., Tanaka, H., Hirakawa, H., Miyoshi, H. and Yasukawa, H.: BUP: A Bottom-Up Parser Embedded in Prolog, *New Generation Computing*, Vol. 1, No. 2, pp. 145-158 (1983).
 - 26) 情報処理, 大特集機械翻訳, Vol. 26, No. 10 (1985).
 - 27) Hays, D. G.: Dependency Theory: A Formalism and Some Observations, *Language*, Vol. 40, No. 4, pp. 551-525 (1964).
 - 28) 吉田 将: 二文節間の係り受けを基礎とした日本語文の構文分析, 電子通信学会論文誌, Vol. 55-D, No. 4, pp. 238-244 (1972).
 - 29) 首藤, 檜原, 津田: 意味理解を目的とした日本語の構文解析アルゴリズム, 福岡大学工学集報, No. 28, pp. 41-49 (1982).
 - 30) Fillmore, C. J. (田中春美ほか訳): 格文法の原理, 三省堂 (1975).
 - 31) 島津, 内藤, 野村: 格構造モデルに基づいた日本語文の分析と解析, 情報処理学会自然言語処理研究会資料, 29-1, pp. 1-8 (1982).
 - 32) 吉田, 田中: 科学技術文書のための日本語文法の制限について—付属語の表現の語彙および語義の制限を中心として—, 情報処理学会自然言語処理研究会資料, 37-4, pp. 1-8 (1983).
 - 33) 辻井, 山梨: 格とその認定規準, 情報処理学会自然言語処理研究会資料, 52-3, pp. 1-7 (1985).
 - 34) 稲永, 吉田: 日本語処理のための機械辞書, 情報処理, Vol. 23, No. 2, pp. 140-146 (Feb. 1982).
 - 35) 水谷, 石綿, 荻野, 賀来, 草薙: 文法と意味 I, 朝倉日本語新講座 3 (1983).
 - 36) 島津, 内藤, 野村: 助詞「の」が結ぶ名詞の意味関係の Subcategorization, 情報処理学会自然言語処理研究会資料, 53-1, pp. 1-8 (1986).
 - 37) 栗原, 吉田, 鶴丸: 日本語文の分析—副詞の処理について(1)—, 九州大学工学集報, Vol. 41, No. 5, pp. 855-860 (1968).
 - 38) 長尾, 辻井, 田中, 石川: 科学技術論文における並列句とその解析, 情報処理学会自然言語処理研究会資料, 36-4, pp. 1-9 (1983).
 - 39) 首藤, 吉村, 津田: 日本語技術文における並列構造, 情報処理学会論文誌, Vol. 27, No. 2, pp. 183-190 (1986).

(昭和 61 年 6 月 3 日受付)