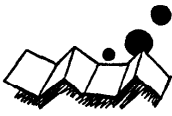


解説



文解析方式†

辻井 潤 一††

1. はじめに

入力文の『構造』を決定する問題は、計算言語学の中心的な課題である。最終的に求める『構造』を理解の結果を表示する構造と考えると、意味処理・文脈処理を含めた計算言語学のほとんどの研究は、この文解析に集約される。本稿では、話題を限定するために、文解析の方式的な側面に焦点をあて、この分野の研究が現在どのような状況にあるか、また、どのような基本的問題が残されているかを展望する。

2. 文解析手法の現況

1979年にやはり今回と同様に本学会誌で計算言語学の特集が生まれ、その中では、

- (1) 拡張文脈自由形文法 (A-CFG)^{19), 63)}
- (2) 拡張遷移文法 (ATN)⁷⁴⁾
- (3) 係受け解析手法^{35), 42)}
- (4) 木構造変換文法²⁵⁾

が構文解析の基本的な枠組として紹介されている¹⁵⁾。

7年後の現在、これらの基本的な枠組は、機械翻訳システムなどの現実的なシステムに適用され、すでにそれぞれの枠組で大規模な文法が開発されている^{1), 6), 8), 42), 47)}。また、それに伴って、各枠組を計算機プログラムとして効率よく実現する手法³⁸⁾や各枠組の長所と短所なども明確になり、標準的な教科書にまとめられるようになってきている^{28), 29), 68)}。しかしながら、大規模な文法の開発は、文解析の基本的な問題がすでに解決されたことを意味するわけではない。むしろ、各枠組での技術的な問題の解決に伴って、枠組の差を越えて、より基本的な問題が明瞭に認識されるようになってきている。1980年代の文解析方式の研究は、次の二つの問題点をめぐって行われてきた。

- (1) 拘束条件と優先的な解釈：よく知られている

ように、統語的な条件だけでは、自然言語の文は、非常に多くの解釈が可能であり、文解析プログラムは、この中から、意味的・文脈的条件などを使って『正しい』解釈を選択しなければならない。上記の各枠組においても、統語規則を適用する段階で意味的・文脈的な条件を検査し、その条件に合格した解釈だけを採用する機構が組み込まれている。しかしながら、この種の検査機構は、対象とする分野が強く限定され、意味的・文脈的な条件が制限的な条件（拘束条件）として使える場合には有効であるが、分野が広がるに従って必要となる意味的な記述が膨大となること、また、日常的に現れる言語の比喩的な用法に対処できないなど、多くの問題をもっている。

(2) 規則記述の枠組：(1)に述べたように、上記の枠組では意味処理や文脈処理と結合するための機構が用意されている。しかしながら、いずれの枠組においても、その機構は一般的な処理手続きの呼出しといった非常に汎用性の高いものであったために、柔軟な処理が可能な反面、文法記述を不透明かつ不完全なものにしてきた。たとえば、A-CFGの枠組では、書換規則適用時に任意のプログラムによる（意味的、文脈的な）適用条件の検査ができるが、どのような検査をするかは、各システムに依存していた。このために、システムによっては、CFGによる統語規則よりも、規則記述の表面には現れない索性間の整合性検査に、文法より重要な側面が埋め込まれることになった。また、ATNにおける BUILDQ の操作のように、統語的な構造とは無関係に任意の『意味』的構造を作ること許したために、統語構造とそれ以外の構造（たとえば、意味構造）との相互関連が非常に ad-hoc なものとなっていた。言い換えると、上記の枠組は、いずれも、名詞・動詞・動詞句といった統語カテゴリを中心とした文法規則を記述するための枠組であり、それ以外の意味的・辞書的な規則性については、システムごとに ad-hoc に対処していた。

† Analysis Methods for a Natural Language Sentence by Jun-ichi TSUJII (Department of Electrical Eng., Kyoto University).

†† 京都大学工学部電気工学第二教室

3. 文法規則の記述と優先的な読み

計算言語学の目的の一つは、表面上の単語列とその単語列の表現する『意味』とを結び付ける体系を作ることであり、文解析研究の目的は、与えられた単語列からその『意味』的構造を出力するシステムを、また、文生成はその逆の変換をするシステムを、開発することである。自然言語の文は、全くランダムな単語の並びからできているわけではない。文がその『意味』を聞き手に伝達できるためには、『単語の列とその意味とを結び付ける』規則が話し手と聞き手の間に共通に存在していなければならない。この単語の列と『意味』とを結び付ける規則性のことを、本稿では文法、また、その規則性を記述したものを文法規則と呼ぶことにする。表面上の単語列を直接『意味』の構造へと写像するような規則の体系を考えるのが良いか、あるいは、その中間に別のレベルの構造を考えるべきかは、それほど自明なことではない。実際、初期の変形文法理論では、表面上の単語の句構造的なまとまりを表示する表層構造、この表層構造と意味的な構造と結び付けるインタフェース構造としての深層構造、そして、『意味』構造という、表面上の単語列を含めると4つの別のレベルの構造を考え、それぞれを別種の写像規則によって結び付けるという構成をとったが、その後、この構成は、理論言語学の分野でさまざまな変更や代案が提案されてきている。計算言語学における文解析の方式を考える場合にも、理論言語学の場合と同様に、

- (1) 写像の対象としてどれだけの構造を考えるか、
- (2) 各構造をどのような形式で表現するか、
- (3) この構造間の写像関係の規則（文法規則）をどのような形式で表現するか

を明確にする必要がある。80年代の計算言語学研究の特徴の一つは、それまで一応独立に進展してきた理論言語学との関係が緊密になってきたことであるが、特に、文解析研究における理論言語学の影響は、非常に顕著なものになっている。言い換えると、上記(1)～(3)の問題について、両分野の研究者がさまざまな代案を提案し、それが前章で述べた既存の文解析方式に代わる新しい文解析の方式へと結実するようになってきている。この問題については、5章で議論する。

文解析のもう一つの重要な問題は、一つの文には非常に多くの構造的な解釈が可能であることから生じ

る。たとえば、単語列と統語構造の対応は、よく知られているように1対Nの対応であり、しかも、このNが一般には非常に大きなものとなる。藤崎³⁷⁾は、久野の英語解析文法(2118の規則をもつ)を5528の文に適用した結果、平均48.5個の解釈が出力されたと報告している。平均単語数が10.85語という比較的短い文でも、この程度の構造的あいまいさが生じる。この解釈の組合わせの爆発の問題は、処理時間・記憶領域の点で問題になるだけでなく、この中からどのようにして正解を選択するかという、応用上も深刻な問題を引き起す。

この問題に対する一つの解決策は、たとえば、意味的・文脈的な拘束条件(Constraint)を導入し、統語規則に従って構造を決定する際に、これらの条件も同時に調べることによって、可能な解釈の個数を減少させることである。述語とその格要素との意味的整合性を調べることは、現在の標準的な手法になっている^{17),35),36)}。

しかしながら、意味的・文脈的条件は、必ずしも構造を決定する際の絶対的な拘束条件としては機能しないことも多い。たとえば、よく意味的情報(あるいは、外界知識)の有用性を示す例としてあげられる

Tom saw Mary with a telescope. (1)

においても、『見る』という動作と道具としての『望遠鏡』との親和性から図-1(a)の解釈を優先できるが、(b)の解釈(『望遠鏡を持っているMary』の解釈)を排除することはできない。意味的・知識的な情

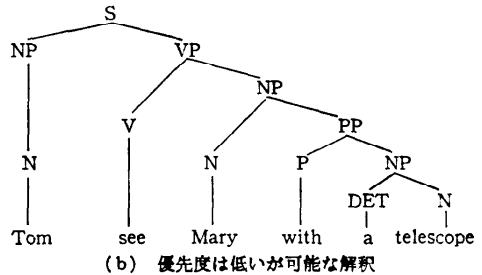
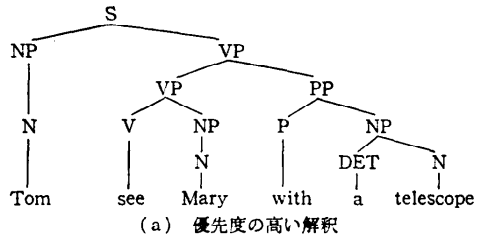


図-1 解釈のあいまいさ

報を活用することによって、with 自体のもつさまざまな意味的可能性、たとえば、『Tom は望遠鏡と一緒に Mary を見た』といった解釈を排除することはできる（これも、実際にはかなり難しい）が、

Tom saw Mary with a telescope. She bought it yesterday to see the comet.

のような例を考えると、(b)の構造を意味的条件から棄却することはできない。『見る』と『望遠鏡』の意味的親和性は、せいぜい解釈(a)を優先するだけで、解釈(b)を禁止するものではない。同様に、

Tom was given a book by his uncle. (2)

も、'by his uncle' を 'to give' の行為者とする解釈がもっとも自然であるだけで、文脈によっては、これを場所格（そばで）とする解釈や 'book' の修飾句（おじが書いた）とする解釈が正しい場合もある。

このように、一文単位の文解析は、それ自体不完全であり、(言語的、非言語的な)周囲文脈の完全な理解なしには、解釈を一意に決定することはできない。しかしながら、一方では、上記(1)、(2)文を単独に見たときには、我々は一応解釈を一意に決定する。たとえば、より広い文脈を参照することによって最終的には一つの整合性のあるテキスト理解に至るにしても³³⁾、文脈中の文がそれぞれ40以上の解釈をもち、その組み合わせだけの異なった文脈があると考えるのは、計算機的にも、人間の処理の心理的実在性の面からも考え難い。我々は、いずれかの時点で、可能な構造解釈から、その時点で最も自然と考えられる解釈を選択しなければならない。この選択をいつ、どの時点で、どのような規則や機構によって行うかという問題は、前述の構造間の写像関係を規定する問題とは別の問題となる。Wilks の指摘した読みの優先 (Preference) の問題^{72), 73)}は、現時点においても依然として文解析の基本的な課題として残されている⁶⁷⁾。次章では、まずこの問題がどのように取り扱われてきたかを見ることにする。

4. 読みの偏好：文解析の制御的な側面

可能な解釈の中から最も妥当なものを選択する手法としては、A-CFG の枠組を採用するシステムでは、規則ごとに数値的な優先度を付ける、組み立てられた部分構造の良さを評価関数で評価する、規則を優先度に従って配列しておくといった方法、あるいは、TEXAS 大学の METAL のように、規則集合を分割し、優先度の高い規則だけでは解析結果が得られない

場合に、より低位の規則集合の適用を試みるといった方法が取られた。また、ATN の場合には、優先度の高い解釈が最初に出力されやすいように各状態からの遷移の枝を配列しておくのが普通である。しかしながら、このような単純な手法では、多くの場合に正しい解釈が優先されるようにシステムを調整することは困難である。この困難を解消する手法としては、次のようなアプローチが取られている。

(1) 統計的な手法：藤崎³⁷⁾は、入力集合から各規則適用の確率を学習し、それをもとに文の統語的解釈に対する尤度を計算することによって、正しい解釈を選択する手法を提案している。訓練集合から確率を学習した文法を適用すると、

We do not utilize outside art services directly.
に対して、

(a) 'art services directly' を 'which' の省略された関係節 ('art' が主語、'services' が他動詞、先行詞が 'outside') とする解釈

(b) 'outside' 以下を前置詞句とする解釈

(c) 'outside art services' を名詞句とする解釈
の3とおりの解釈が出力されるが、それぞれの尤度は0.356, 0.003, 0.641 となり、この場合には、正しい解釈の尤度が最も高くなる。文集合から規則適用の優先度を求める考え方は、長尾らも採用している³⁰⁾。あらかじめ与えられた規則の優先度だけを調整することのような方法に限界があることは確かである（たとえば、前記の『見る』と『望遠鏡』の意味的な親和性は、あらかじめ用意された統語的な規則の優先度を変更するだけでは捉えられない）が、すでに作成された文法規則を、その範囲内で最適な状態にするには有効であろう。また、英語の多品詞語のように、解釈の間に大きな頻度差がある場合（たとえば、'paper' の名詞解釈と動詞解釈）や、非文法的な文の出現を認める場合（実際、非文法的な定義自体があいまいであり、頻度が非常に低い表現と非文法的表現の境界はあまり明瞭ではない）などでは、解釈の選択に統計的な数値を使用することは有効であろう。このようなアプローチの延長上には、訓練集合から規則そのものを学習的に変更することも考えられる¹⁴⁾。

(2) 解釈の保留：統語的な制限条件だけでは、多くの解釈可能性が残るということは、統語解析の段階では、構造を一意に決定するのに十分な情報が参照できないことを意味する。したがって、統語解析の段階では、係先を厳密には決定せず、他の情報（意味的、

文脈の情報)が参照可能となる,より後の処理段階までこの決定を延期することが考えられる. Woods の ATN による解析では,前置詞句の修飾先を取りあえず句構造中の最も高い位置にある節点としておき,文全体の(おおまかな)構造が決定された後で LISP プログラムによって前置詞句の係り先を決定している⁷⁴⁾. また,井佐原は,

『花子は池に沈んだ人を助けた犬を見た』

に対して,通常の統語解析では12個の異なる木構造が出力されてしまうという困難を指摘し,統語解析ではあいまいさを保存した図-2のような構造だけを出力する右方枝分かれの文法を採用している⁴⁾. この手法では,図-2の構造から厳密な構造を決定するのは,意味解析の仕事となる.この『決定の保留』の考え方は,上の二つのシステム以外にも,新田らのQP-構造⁹⁾などさまざまなシステムに見られる.ただ,この考え方はそれ単独では問題の解決にはならず,解釈を決定する『後の処理』とはどのようなもので,かつ,その部門での規則記述がどの程度明示的に記述できるかが問題となる.

(3) 決定的な解析^{50),59)}: (2)の議論は,必ずしも,文全体についての解析が終了するまで詳細な構造決定を保留すべきであるという議論にはならない.決定に十分な情報が参照できるようになれば,その時点で構造を決定していけば良い. Marcus⁵⁹⁾は,

(i) 人間が文を読み進む過程では,解釈を一意に決定していること,

(ii) したがって,それまでに決定した仮説的な構造を破棄しなければならない

The cotton clothing is made of grows in Mississippi.

といった文 (garden-path sentence) に対しては,人間

は強い認知的困難を感じることを,

(iii) それまでの解析手法では,非常に多くの仮説的な構造を同時に保持しているか,あるいは,一つの仮説に従って解析を進めている場合には非常に多くのバックトラック動作を繰り返すことになっており,いずれの方式も(i), (ii)のような人間の処理過程を反映しているとはいえないこと,

から,ある一定の先読み機能と構造決定の保留機構を持つパーザ (wait-and-see parser) を提案し,入力文を左から右に解析していく過程で一意に構造を決定していけることを示した. A-CFG や ATN の規則が,英語や日本語における可能な言語表現の集合を規定する規則であるのに対して,このパーザ (Parsifal) での規則記述は,ある環境の下で最も妥当と思われる構造を決定する規則になっている. さらに, Marcus は,この決定性の仮説から出発し,パーザの出力する統語構造を厳密な構成要素間の関係を規定する木構造で表現することの不都合を指摘して, (2) の『決定の保留』の考え方を導入することも提案している⁶⁰⁾.

(4) 構造決定規則を中心にする手法: Parsifal では,決定的な統語解析とは独立に意味的な整合性を検査する機構が存在することが前提とされているが,この前提は機械翻訳のような現実的なシステムにおいては成立しない. また, Parsifal では取り扱えない構造 (たとえば,各種の並列構造) も実際のテキストには数多く表れる. このような現実のテキストを対象とするシステムでは,人間の心理プロセスとの対応という観点から導入された Parsifal のさまざまな制限 (バッファ・サイズの制限など) は不要であり,構造を決定する規則は,さまざまな種類の情報を活用できるようにしておく方が良い. このような考え方から, Mu-プロジェクト²⁷⁾の文解析²¹⁾においては,たとえば,並

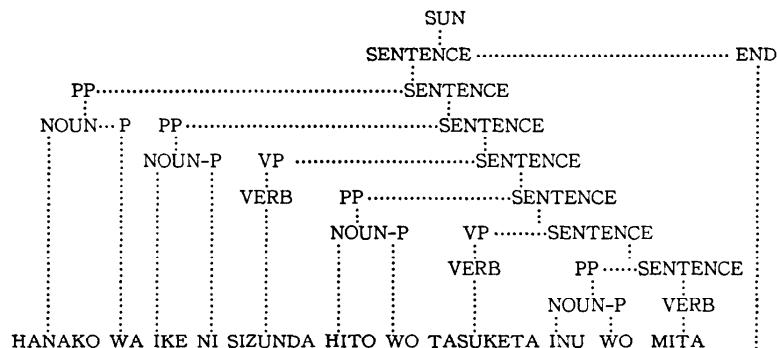


図-2 あいまいさを残した木⁴⁾

列名詞句のスコープ決定を、

(i) 助詞『と』が呼応している場合には、二つ目の『と』をスコープの区切りとせよ。

(ex: 関数の Z-変換とグレイコードとの関連)

(ii) 助詞『など』が存在すれば、それをスコープの区切りとせよ。

(ex: 迷路法, 探索法, チャネル法などの配線手法)

(iii) 同一名詞が存在すれば、それをスコープの区切りとせよ。

(ex: 散乱行列と伝達行列の次元数)

といったさまざまな種類の情報を参照する発見的な規則で行っている^{51,26)}。これらの規則は、いずれも、『指定の条件が満たされれば、スコープをそのように決定せよ』というタイプの規則であり、通常の句構造規則で規定される『日本語の文が満足しなければならない』拘束条件を規定する規則ではない²²⁾。Mu-プロジェクトでは、文解析を最も妥当な解釈を決定していく過程と捉え、文法をこの種の構造決定規則の集合として記述している。Mu の文解析文法は、

(i) 決定規則にはそれぞれの間に優先度の差があり、複数の規則が異なった構造を推奨する場合には、優先度の高い規則の決定に従うこと²²⁾、

(ii) 大域的な構造の把握を行った後に、詳細な構造を決定する『決定の保留』の機構を用いる、すなわち、処理単位間に順序を設けること⁴⁶⁾、から、規則集合の適用順序をコントロール・グラフ(部分文法ネットワーク)で指定する手続き的な文法となっている³¹⁾。このタイプの文法は、さまざまな種類の発見的構造決定規則を活用できる反面、規則の規模が大きくなること、宣言的な規則記述に比べて規則の可読性が落ちることなど、文法の保守容易性の点で問題をもっている。

(5) 拘束条件の記述と一般的な優先の原則: Parsifal や Mu の文法は、決定的な文解析 (deterministic parsing) を行うために文法記述そのものを構造決定規則の集合で記述した。これに対して、Shieber⁶⁴⁾、往住²⁴⁾、Wanner⁷¹⁾ らは、規則記述は CFG 規則や ATN のままにしておき、人間の文解析に見られる『読みの優先』の原則を規則適用の機構中に組み込む手法を取っている。たとえば、Shieber は、CFG の Shift-Reduce パーザを基本機構とし、これに、

(i) 読み込み (shift) と還元 (reduce) の二つの操作が可能の場合には、還元操作を優先する

(ii) 複数の還元操作が可能な場合は、最も長い範囲をまとめる規則を優先せよ

といったメタ・レベルの原則を導入することによって、人間の読みの偏好が説明できることを示している。しかしながら、この種の原則は、語彙的・意味論的・語用論的な偏好が特に存在しない場合のデフォルト的な読みの偏好 (たとえば、(i), (ii) は、それぞれ、『複数の修飾先がある場合には、最も近くのを優先せよ』、『広い意味での最長一致的な規則適用を行え』という単純な原則である) を説明するだけであり、往住²⁴⁾のあげる

(a) 花子は新聞で私が書評した本を買った。

(b) 花子は丸善で私が書評した本を買った。

に見られる語用論的な要因や(4)の手法のような表層表現上の細かな優先要因を取りこむことはできず、工学的な文解析手法とするにはさらに研究が必要であろう。

ただ、CFG で規定される人間の言語能力と実際に制限されたリソースの下で示される人間の言語運用的な側面とを区別して考え、規則適用の機構に制限を設けることによって、人間の読みの偏好を説明しようとする試みは、処理の実時間性と短期記憶の容量制限をもとにした橋田³⁴⁾の議論など、近年興味ある研究を生み出している。また、言語運用的な側面は、CFG よりも低位の正規文法で記述できるとする Langedoen らの議論もある³⁷⁾。

(6) CFG と Filtering: 構造間の写像関係を規定する文法規則と構造決定のための規則とを分離しようとする考え方は、実際的なシステムにも見られる。櫛¹³⁾は、CFG 規則によって組み立てられた構造をあらかじめ指定された禁止木・排他木 (もしその解釈が可能であれば、他の解釈よりも優先される解釈、優先度の数値が付加されている) と比較することによって、可能な解釈の中からより妥当性の高い解釈 (複数個) を選択する方式を提案している。ただ、この方式は、(5)の一般的な原則よりもキメ細かな指定が可能なる反面、禁止木・排他木の記述が雑多になり、系統的な文法開発が難しくなるように見える。

(7) Connectionist 的なパーザ: 以上の議論は、いずれも表層的な手掛りや統語構造を参照して、人間における読みの偏好を計算機的に実現しようとするものであった。より大域的な文脈的要因や意味的要因を考慮して読みを決定する手法は、まだほとんど手がついでいない。Waltz⁷⁰⁾ らは、多義語の意味が文脈的要

因によって決定される過程を、

He shot some bucks.

を取り上げて、議論している。この文には、'to shoot' と 'buck' の多義性から、『子鹿を射た』と『10ドルを失った』の二つの主要な読みがあるが、Waltz らは、この二つの読みが文脈によってどのように選択されていくかを、単語の意味の間の活性化と抑制のネットワーク・モデルによって模擬している⁷⁾。これは、コネクション・マシンに代表される高度に分散した並列処理の機構を自然言語理解に適用しようとする興味のある試みでもある。彼らの試みは、読みの偏好に対する統語的・意味的・語用論的な要因をすべて分散型のネットワークを使った一様な処理で行おうとするもので、(現在の段階では、まだ未熟であるが)今後の展開が期待される。心理的プロセス、特に、刺激拡散モデル的なものへの興味は、日本のいくつかの研究においても顕著なものになってきている^{21),18)}。

5. 規則記述の枠組：文解析の記述的な側面

前章での議論が、読みの偏好という人間的な処理プロセスを問題にする、強いて言えば、心理学的傾向の強い問題意識であったのに対して、80年代の文解析方式の研究には、理論言語学からの影響も強く見られる。

2章で述べた文解析方式の多くは、名詞・動詞・名詞句・動詞句・文といった『統語カテゴリ』を基礎にして、文法規則を組織化する枠組になっていた。しかしながら、実際の文解析においては、

(i) 文法規則を精密化すればするほど、たとえば、Longman や Hornby 辞書における動詞の細分類のように、各単語に個別な情報が必要になること²²⁾

(ii) 高度な文解析を行うためには、意味や文脈に関する情報を文解析時に参照する必要があることなど、統語カテゴリでは表現し難い情報が必要となり、統語カテゴリを使った規則で表現できる言語の規則性は、意外に少ない。このような認識は、現在の理論言語学にも見られ、1979年以降の文解析研究の中で、このような傾向を代表するものを以下に列挙する。

(1) 意味記述の定式化：それまでの言語学が、生成意味論的なアプローチを除くと、意味の問題をできるだけ排除していたのに対して、Montague は、意味を中心に文法理論を再編成した⁵¹⁾。このモンテギューの文法に基づく文解析プログラムもいくつか開発され

た^{32),39),52)}が、それ自体は文解析方式として特に目新しい展開をもつものではなかった。しかし、意味も形式的に記述できること、また、統語構造と意味記述との対応をシステムティックに規則化できることを示したことは、それまで意味の記述形式や統語構造と意味構造の相互関係を ad-hoc に処理していた計算言語学に及ぼした影響は大きく、その後の計算言語学と理論言語学との相互接近を促す契機となった。

(2) 構造記述間の対応に関する宣言的記述：3章で述べた、表面上の単語列とその意味構造との対応関係を説明する体系を作ること、理論言語学の基本的な問題でもある⁴⁵⁾。初期の変形文法は、この対応を変形規則という木構造の変換によって説明した。変換規則を入力の木構造を別の木構造へ変換する『命令』であると考え、二つの構造を結び付ける『手続き』によって、構造間の写像関係を説明していたと考えることができる。実際、木構造の変形規則による記述は、

(i) 処理の順序(変換規則の適用順序)に敏感である

(ii) 生成方向への適用は容易であるが、認識方向への適用は困難であるという方向性をもつという、手続き的な記述の基本的性質(欠点)をもっている。これに対して、二つ(あるいは、それ以上の)構造の対応は、おのおのの部分構造がどのように対応するかを指定したより宣言的な形式でも記述できる。LFG⁴⁹⁾の取った記述方法は、このような記述方法の一つである。宣言的記述が知識の記述の明示性や可読性に優れていることから、このLFG的な記述は、計算言語学においても非常に利点であり、LFGに基づく多くの文解析プログラムが開発されている^{10),43),44)}。

(3) 多種類の情報の記述：変形規則による記述は、構造間の対応関係を説明することはできるが、いつ、どのような目的で、変形規則が適用されるかについては、なにも説明しない。ある文体(たとえば、it-thatのような分裂構文)が選択されるのは、焦点や情報の新旧といった文脈的な要因など、さまざまな要因に左右されている。このような要因を文解析時点で明示的に抽出しておくことは、機械翻訳や文章要約などに使われる文解析には不可欠である²³⁾。(2)の宣言的記述に、このような多種類の情報レベル間の相互関係の記述を導入しようとするものに、KayらのFUG(Functional Unification Grammar)がある^{55),56)}。こ

の種の情報は、文の生成には不可欠であり、文生成においては従来からシステミック文法のように、言語の構造的側面よりも機能的側面に重点を置いた文法が使われている^{40), 59)}。(2)の文法の宣言的記述が、文の認識にも生成にも使える双方向を目指すすると、このような機能的側面と構造的側面の相互関係に関する記述が無視できないことは明らかであろう。

(4) 辞書的情報の重視：文法規則の精密化は、多くの場合、各単語ごとの記述の精密化に繋がる。計算言語学においては、従来から、各述語ごとの表層格構造を辞書に記述しておき、これを統語構造を決定する際の『拘束規則』として活用することが一般的に行われていたが、A-CFGやATNで規定される統語処理と辞書の記述とをどのように融合するかはシステムごとに異なり、一般的な『理論化』は行われていなかった。この『理論』の欠如は、統語規則と辞書や意味解釈規則の役割分担を不明確にし、また、往々にして、システムが受理できる可能な言語表現の集合を不明確なものにしていた。これに対して、理論言語学から提案されてきたLFGやHG⁶²⁾などの枠組は、このような単語ごとの情報を積極的に取り込んだ見通しの良い理論となっており、ICOTを初め、多くの計算言語学のグループに採用されている^{3), 40), 69)}。また、郡司¹²⁾は、同様の枠組によって、日本語における語順の任意性を取り扱うことを提案している。

上記(2)、(3)、(4)の傾向は、伝統的な変形文法における貧弱な規則記述の形式(CFGの構築規則とそれを補うための変形規則)を見直そうとするものであり、たとえば、GPSG^{11), 53), 54)}やLFGは、規則記述の枠組を豊富にすることによって、従来変形規則で説明していた言語事実を句構造規則や単語ごとの辞書規則によって説明しようとする。このような理論言語学における『規則の記述と計算の枠組に対する再検討』の動きは、統語カテゴリ間の書換規則という非常に限定された共通点しかもたなかった理論言語学と計算言語学の相互関係を、より緊密で有効なものに変えつつある。たとえば、Shieberら^{61), 66)}は、ユニフィケーションを基礎にした文法記述言語PATR-II⁶⁵⁾を中心にして、GPSG・LFG・HGなどの文法間の還元(Reduction)関係を議論しているが、一見異なった規則記述の枠組と計算機構をもつ文法体系間の、計算言語学の立場からの比較は、文解析に必要な情報の種類(辞書に記述すべき情報と一般的な統語規則で表現すべき情報の区別、など)とその計算機構とを明らかに

するだけでなく、計算言語学、理論言語学の両分野での今後の研究を見通しの良いものにするのに役立つであろう。また、ユニフィケーション・ベースの言語CILを文脈理解と文解析のいずれにも適用しようという向井らの試み⁴¹⁾は、文脈まで含めた文解析のための一般的な計算機構とはなにかを考える上で興味深いものとなっている。

6. おわりに

本稿では、前回の計算言語学の特集(1979年)から今日に至るまでの文解析研究の方式的な展開に焦点をあてて議論した。この7年間は、機械翻訳における大規模文法の開発に見られるような既存の文解析方式の成熟の時期であるとともに、既存の方式を革新する新しい方式が登場した時期でもあった。ただ、現時点では、3, 4章で議論した、読みの偏好という内部での処理プロセスの問題(あるいは、可能な解釈から最も妥当な解釈を選択する問題)と5章での理論言語学に触発された文解析方式とは独立して議論されている段階であり、これら二つの側面(心理学的側面と理論言語学的側面)が統合され、現実的な文解析の手法となるためには、将来の研究に待つところが大きい。

参考文献

- 1) 天野真家：英日機械翻訳用パーサについて、情報処理学会NL研, 32-1(1982)。
- 2) 安西裕一郎：Towards Emotional Architecture for Natural Language Processing, Proc. of International Symposium on Language Artificial Intelligence, North-Holland, to appear.
- 3) 安食敏弘他：状況と素性具現化に基づく自然言語解析, 日本ソフトウェア科学会第2回大会論文集2A-1, pp. 33-36(1985)。
- 4) 井佐原均他：日本語埋め込み文の構文解析における諸問題, 情報処理学会NL研, 26-4(1981)。
- 5) 石川雅彦他：英語並列句のあいまいさとその処理, 情報処理学会NL研, 50-2(1985)。
- 6) 市山他：日本語処理の観点から見たプッシュダウンオートマタの再評価, 情報処理学会自然言語処理シンポジウム論文集, pp. 23-27(1983)。
- 7) 乾敏郎他：知識の構造と獲得過程に関するメモ, 科研費『談話行動のモデル化に関する認知科学的研究』報告書(代表者 辻井潤一)(1986)。
- 8) 内田裕士他：日英機械翻訳システムATLAS/U, 情報処理学会NL研, 29-3(1985)。
- 9) 岡島惇他：自然言語処理における中間語表現と多義性解消のしやすさとの関係, 情報処理学会NLシンポジウム論文集, pp. 85-90(1983)。

- 10) 工藤育男他: LFG のスキーマを用いた機械翻訳, 情報処理学会 NL 研, 48-8 (1985).
- 11) 郡司隆男: ソフトウェア文書のための日本語処理の研究-3, 4, 情報処理振興事業協会技術センター(1983).
- 12) 郡司隆男: Subcategorization and Word Order, Proc. of International Symposium on Language Artificial Intelligence, North-Holland, to appear.
- 13) 榑 博史他: フィルタリング操作を伴うパーズングの一形式, 情報処理学会 NL 研, 43-1 (1984).
- 14) 佐藤理史他: 文法推論に基づいた翻訳文法の学習方式, 情報処理学会 AI 研, 46-11 (1986).
- 15) 佐藤泰介: 日本語の構文解析, 情報処理, Vol. 20, No. 10, pp. 865-872 (Oct. 1979).
- 16) 今野 聡他: ボトムアップ構文解析システム BUP での左外置変形, 情報処理学会 NL 研, 44-1 (1984).
- 17) 島津 明他: 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 NL 研, 29-1 (1982).
- 18) 島津 明他: 助詞『の』が結ぶ名詞の意味関係の Sub-categorization, 情報処理学会 NL 研, 53-1 (1986).
- 19) 田中穂積: 拡張 LINGOL—自然言語処理のためのプログラミング・システム, 電子技術総合研究所 (1979).
- 20) 田中穂積: 意味表現用言語 SRL の機械翻訳への応用, 情報処理学会 NL 研, 31-5 (1982).
- 21) 辻井潤一: 日本語構文解析, 情報処理学会 NL 研, 38-5 (1983).
- 22) 辻井潤一他: Analysis Grammar of Japanese in the Mu-Project—A Procedural Approach to Analysis Grammar, Proc. of COLING 84, pp. 267-274 (1984).
- 23) 辻井潤一: Future Directions of Machine Translation, Proc. of COLING 86, to appear.
- 24) 往住彰文: 日本語文の構造的な多義性を決定論的に解決するための一般的規則について, 日本認知科学会第二回大会論文集 E-5, pp. 59-60 (1985).
- 25) 長尾 真他: 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).
- 26) 長尾 真他: 科学技術論文における並列句とその解析, 情報処理学会 NL 研, 36-4 (1983).
- 27) 長尾 真他: 科学技術庁機械翻訳プロジェクトの概要, 情報処理, Vol. 26, No. 10, pp. 1203-1213 (Oct. 1985).
- 28) 長尾 真: 言語工学, 昭光堂 (1983).
- 29) 長尾 真編: 日本語情報処理, 第5章『構文解析』, 電子通信学会 (1984).
- 30) 長尾 真他: A Parser which Learns the Application Order of Rewriting Rules, Proc. of COLING 82, pp. 253-258 (1982).
- 31) 中村順一他: Grammar Writing System GRADE of Mu-Machine Translation Project and its Characteristic, Proc. of COLING 84, pp. 338-343 (1984).
- 32) 西田豊明: Studies on the Application of Formal Semantics to English-Japanese Machine Translation, Ph. D. Dissertation, Kyoto Univ. (1983).
- 33) 西田豊明: 言語と記憶, 電子通信学会言語処理とコミュニケーション, NLC 86-2 (1986).
- 34) 橋田浩一他: 言語運用のモデルに基づく実時間の構文解析, 情報処理学会 NL 研, 37-3 (1983).
- 35) 日高 達: 格文法による日本の構文解析, 情報処理学会 NL シンポジウム論文集, pp. 41-46 (1983).
- 36) 平井 誠: 格構造の変換規則と簡単な事象の分類を用いた日本語文解析, 情報処理学会 NL 研, 33-3 (1982).
- 37) 藤崎哲之助: A Stochastic Approach to Sentence Parsing, Proc. of COLING 84, pp. 16-19 (1984).
- 38) 松本裕二他: BUP の高速化, 情報処理学会 NL 研, 39-5 (1983).
- 39) 松本裕二他: 日本語モンテギュー文法の実働化と質問応答への応用, 情報処理学会 NL 研, 31-7 (1982).
- 40) 三吉秀夫他: 日本語文法記述—JPSG, ソフトウェア科学会『論理と自然言語』ワークショップ, LNL 86-2 (1986).
- 41) 向井国昭: Unification over Complex Indeterminate in Prolog, ICOT TR 113 (1985).
- 42) 村木一至: 日英機械翻訳システム VENUS の解析メカニズム, 情報処理学会 NL 研, 40-7 (1983).
- 43) 安川秀樹: 文法関係の形式的記述について, 情報処理学会 NL 研, 39-5 (1983).
- 44) 安川秀樹: LFG System in Prolog, Proc. of COLING 84, pp. 358-361 (1984).
- 45) 山梨正明: 計算機からみた意味表現, 情報処理, Vol. 20, No. 10, pp. 873-879 (Oct. 1979).
- 46) 山本利文: Mu-プロジェクトの英日機械翻訳システムにおける多品詞語の解消, 情報処理学会 NL 研 53-7 (1986).
- 47) 吉田 将: 機械翻訳のための構文解析手法, 情報処理, Vol. 26, No. 10, pp. 1157-1164 (Oct. 1985).
- 48) Bateman, J.: Conversation Generation—A Theoretical Watershed?, in New Development in Systemic Linguistics (eds: Fawcett and Young), Frances Printer, London, to appear.
- 49) Bresnan, J. ed.: The Mental Representation of Grammatical Relations, The MIT Press (1982).
- 50) Carter, A. W. et al.: Simplifying Deterministic Parsing, Proc. of COLING 84, pp. 239-242 (1984).

- 51) Dowty, D.R. : A Guide to Montague's PTQ, Indiana University linguistic Club (1978).
- 52) Friedman, J. : Evaluating English Sentences in a Logical Model: A Process Version of Montague Grammar, Proc. of COLING 78, B1-16 (1978).
- 53) Gazder, G. et al. : Recent Computer Implementation of Phrase Structure Grammars, Computational Linguistics, Vol. 10, No. 3-4, pp. 212-214 (1984)
- 54) Gazder, G. et al. : Generalized Phrase Structure Grammars, Oxford, Basil Blackwell (1985).
- 55) Kay, M. : Functional Grammar, Xerox, PARC memo (1978).
- 56) Kay, M. : Functional Unification Grammar : A Formalism for Machine Translation, Proc. of COLING 84, pp. 75-78 (1984).
- 57) Langendoen, D. et al. : The Representation of Constituent Structure for Finite-State Parsing, Proc. of COLING 84, pp. 24-27 (1984).
- 58) Mann, W. C. et al. : Nigel, A Systemic Grammar for Text Generation, USC report, ISI/RR-83-105 (1983).
- 59) Marcus, M. : A Theory of Syntactic Recognition for Natural Language, MIT Press (1980).
- 60) Marcus, M. : Deterministic Parsing and Description Theory, Proc. of ALVEY/ICL Workshop on Linguistic Theory and Computer Application, UMIST (1985).
- 61) Pereira, F.C. N. et al. : The Semantics of Grammar Formalisms Seen as Computer Languages, Proc. of COLING 84, pp. 123-129 (1984).
- 62) Pollard, C. J. : Generalized Phrase Structure Grammars, Head Grammars and Natural Language, Ph. D. Dissertation, Stanford University (1984).
- 63) Robinson, J. J. : DIAGRAM, A Grammar for Dialogue, SRI Technical Notes 205 (1980).
- 64) Shieber, S. M. : Sentence Disambiguation by a Shift-Reduce Parsing Technique, 21 ACL (1983).
- 65) Shieber, S. M. : The Design of a Computer Language for Linguistic Information, Proc. of COLING 84, pp. 362-366, (1984).
- 66) Shieber, S. M. : Separating Linguistic Analysis from Linguistic Theories, Proc. of ALVEY/ICL Workshop on Linguistic Theory and Computer Application, UMIST (1985).
- 67) Schubert, L. K. : On Parsing Preference, Proc. of COLING 84, pp. 247-250 (1984).
- 68) Tennant, H. : Natural Language Processing, Petrocelli Books (1981) (邦訳: 自然言語処理入門, 森 健一他, 産業図書).
- 69) Vorasucha, V. et al. : Thai Syntax Analysis Using GPSG, 日本ソフトウェア科学会第2回大会論文集 2A-4, pp. 45-48 (1985).
- 70) Waltz, D. L. et al. : Massively Parallel Parsing : A Strongly Interactive Model of Natural Language Interpretation, Cognitive Science, Vol. 9, pp 51-74 (1985).
- 71) Wanner, E. : The ATN and the Sausage machine, Cognition, 8 (1980).
- 72) Wilks, Y. : Parsing English II, in Computational Semantics (eds: Charniak & Wilks), North-Holland (1976).
- 73) Wilks, Y. : Making Preference More Powerful, Artif. Intell. Vol. 11, No. 3, pp. 197-223 (1978).
- 74) Woods, W. A. : Transition Network Grammars for Natural Language Analysis, CACM Vol. 13, pp. 591-606 (1970).

(昭和 61 年 6 月 27 日 受付)