



Title	日本語理解システム実現のための省略の分析
Author(s)	桃内, 佳雄; Momouchi, Yoshio; 宮本, 衛市 他
Citation	北海道大學工學部研究報告, 108, 93-104
Issue Date	1982-05-31
Doc URL	<a href="https://hdl.handle.net/2115/41717">https://hdl.handle.net/2115/41717</a>
Type	departmental bulletin paper
File Information	108_93-104.pdf



## 日本語理解システム実現のための省略の分析

桃内佳雄 宮本衛市 竹村伸一

(昭和56年12月26日受理)

### Some Considerations on Ellipsis for the Implementation of Japanese Language Understanding System

Yoshio MOMOUCHI, Eiichi MIYAMOTO and Shin-ichi TAKEMURA

(Received December 26, 1981)

#### Abstract

We examine the elliptic expressions and present a partial taxonomy of ellipsis. Elliptic expressions are linguistic elements omitted from texts as something understood without appearing in texts. We investigate several kinds of information used to restore elliptic expressions. This information comes from the context, knowledge, which is linguistic or nonlinguistic, and situations. We show that information from syntactic structures, word meanings or coherence relations between sentences plays an important role to restore elliptic expressions. It is our hope that these considerations are an underlying study to construct a coherent Japanese language understanding system. As an experimental system, we have implemented a language processing system. It processes a simple type of ellipsis, which requires information from syntactic structures of the immediately preceding sentence to restore the elliptic expression. We describe this process with an example.

#### 1. ま え が き

計算機の高度化の一つに、計算機の知的機能を向上させ人間と同様の能力を持つようにしたいということがある。計算機による自然言語理解システムの実現はこのことと密接に関連している。我々人間は、日常生活において自然言語によって相互のコミュニケーションを行っているが、計算機に自然言語を理解させることはそれほど簡単なことではない。自然言語を日本語に限ったとしても、計算機による高度な日本語理解システムを実現するためには、計算機科学だけでなく、言語学、心理学、論理学、哲学などの研究成果をとりこみ、さらに、計算機科学の立場からの日本語の構造の解明が重要となる。本報告は、計算機科学の立場からの、一文をこえた文章の理解の問題、特に文章における省略の問題の基礎的考察と省略処理を含む実験的な言語処理システムの構成についての報告である。

文章理解において、文章の意味を正しく理解するためには、文章を構成している個々の構成要素についての理解と、構成要素間の結合関係についての理解が文章全体の理解へと正しく統合されなければならない。この統合の過程において文章中の構成要素間で種々の情報の結合が行われる。Halliday と Hasan<sup>1</sup> (1976) により英語の文章における次のような結合関係についての詳細な

考察が行われている。① 指示 (Reference), ② 代用 (Substitution), ③ 省略 (Ellipsis), ④ 接続 (Conjunction), ⑤ 語彙結合 (Lexical cohesion)。本報告においては、これらの結合関係のうち、省略、特に日本語の文章における省略についていくつかの考察を行い、日本語理解システム実現のための一つの基礎を与える。

省略は、言語の意味に近いレベルの情報から、語用論的知識、統語的構造、さらには文そのものに近いきわめて表層レベルの情報までも含んだ総合的な立場から分析されなければならないことが田中<sup>2)</sup>(1979)により指摘されている。三上<sup>3)</sup>(1970)は省略の法則について、久野<sup>4)</sup>(1978)は省略における談話法則について考察を行っている。森岡<sup>5)</sup>(1980)は省略を次のように規定している。

- (1) 言語の省略とは、話し手、聞き手あるいは観察者によって、ある要素が省かれていると実感される現象である。
- (2) 省略されていると意識される要素は、久野<sup>4)</sup>氏の「省略の根本原則」にあるとおり、言語的あるいは非言語的文脈から復元可能でなければならない。
- (3) 省略されているという意識は、他の言語表現と比較することによって生ずる。

本報告においても、省略をこのように比較的ゆるやかな規準で考えることとする。文章を構成する構成要素の単位に着目して省略の部分的な分類を行い、次に、省略されている要素を復元するために必要な情報について基本的な考察を行い、さらに、そのような情報がどこから得られるかという視点を保ちながら、分離先行表現の問題、文と文との接続関係と省略との関連について論じ、最後に、簡単な省略処理を含む、実験的な言語処理プログラムについて述べる。

## 2. 省略の部分的な分類

省略についての最初の問題は基本的な省略の類型を分類、整理することであろう。種々の観点からの分類が考えられるが、文章を文字列と考えると、省略されるのは文章の構成要素である文字列であるから、計算機が、省略された文字列を理解するための基礎として、どのような文章の構成要素が省略され、それはどのような文字列から構成されるのかという観点からの省略の一つの分類を試みる。

省略を含む文を省略文、省略の位置を省略位置、省略されている表現を省略表現とよぶことにする。たとえば、次の例において、

{父は茶の間へはいらなかった。φ隣の間に入った。}

第二文が省略文で、その省略位置がφで示されている。φには実際には何の表現も示されない。このφの位置における空の表現をゼロの照応表現とよぶことにし、改めて、φをゼロの照応表現を表わすものとする。φに復元されるべき省略表現は、“父は”である。また、このゼロの照応表現は第一文の“父”と照応しており、このように文脈中の、照応表現と照応する表現を先行表現とよぶことにする。先行表現と照応表現の間の関係を照応関係とよぶ。先行表現は、必ずしも、照応表現の先行文脈中のみ現れるとは限らないが、後続文脈中に現れるものも含めて先行表現とよぶことにする。文脈中の先行表現は省略表現を復元するための重要な情報である。文脈に依存した省略は、先行表現とゼロの照応表現の間の結合関係について考察するものであるということができよう。省略は文脈のみでなく、言語的あるいは非言語的な知識、および場(状況)などにも依存する。特に、場に依存した省略は会話における省略に多い。

文章の構成要素の単位として、文章、文、句、語、語構成要素を設定し、省略表現を文(文

章), 句, 語, 語構成要素と段階的に考えることによって, 省略の一つの分類が可能となる。次にそのような方針による省略の部分的な分類が例とともに示される。文は句点(。)で区切られる単位であり, 文章は一つ以上の文の並びである。句, 語, 語構成要素については, 分類とその例とともに説明が与えられる。例において[ ]で括られている表現が省略表現であり, ... は文脈の省略を示している。一部の例には, 先行表現を含む文脈も同時に示される。

[I] 省略表現が語を構成する語構成要素である場合

(1) 語素の省略

{電子計算機... [電子] 計算機... }

(先行文脈中に“電子計算機”という表現が存在していて, “電子計算機”を構成する語素である“電子”という表現が省略された。)

{n 個の整数データ... [整数] データ... }

(2) 連体修飾部の省略

{三つ葉とゆずの皮を [茶わんむしの] 上にのせ... }

{ [アジの] 頭を落とし... }

(ここでは, “茶わんむしの上” とか “アジの頭” などの表現はすべて語と考えている。

すなわち, “頭”, “上” などの名詞そのものも語であるが, 連体修飾句+名詞という形の表現も語と考えている。)

(3) 被連体修飾部の省略

{小さい方の数で大きい方 [の数] を割って... }

{最初に加える数は2です。二番目 [に加える数] は4です。}

(“の数”, “に加える数” などの表現を被連体修飾部とよんでいる。)

[II] 省略表現が句を構成する語である場合

(1) 助詞の省略

{彼 [は], ほんとうにアメリカへ行ったのですか。}

(助詞は語と考えている。“彼は” は句である。)

(2) 助動詞の省略

{太郎は学校へ行き, 次郎は図書館へ行った。}

(“行き” という表現において過去の時制に関する情報が省略されていると考える。この場合は, “た” という助動詞の省略。)

(3) 動詞, 形容詞の省略

{ [行く] かもしれない。}

[III] 省略表現が文を構成する句である場合

(1) 主題部の省略

{サバは三枚におろしてうす塩をします。 [サバは] しばらくして酢につけます。}

{鯨はけものだ。 [鯨は] 魚ではない。}

(“サバは”, “鯨は” という表現は主題部, これは主題に対する表現+係助詞“は”からなる。)

{今日は太郎が欠席した。 [太郎は] 風邪をひいたらしい。}

{太郎がわざわざ花子を訪ねて行ったのに, [花子は] 会ってくれなかった。}

(この二例では, 先行表現が先行文脈中の格要素に対応する表現である。)

## (2) 格要素部の省略

{あなたは、昨日山田さんにあいましたか。ええ[私は昨日山田さんに]会いました。}

(“私は”は主題部，“昨日”，“山田さんに”は格要素部。格要素部は格要素に対する表現＋(格助詞または空表現)からなる。)

## (3) 動詞句，形容語句の省略

{ポン酢じょうゆは小鉢受けに[用意し]，薬味は小皿に用意する。}

{油の温度は170℃前後[とする]。}

(“とする”のような表現も動詞句と考える。)

## 〔Ⅳ〕省略表現が文章を構成する文(文章)である場合

{ふたが閉まらないときは：カセットのストッパーがついたままか，逆向きに入っているときです。[カセットのストッパーをとるか，カセットの向きを変えるか，あるいはその両方をして下さい。]}

(この例では常識的な推論により省略表現を復元しなければならない。また省略表現そのものにもかなりの自由度がある。)

省略表現を文章の構成要素を単位としてこのように段階的に構成要素間の関係としてとらえることは、何か省略されているということ判定する手続き、何が省略されているかを同定する手続きを作るうえでよい助けとなるであろう。検討された資料(例)が部分的であり、詳細化の可能性を残しているという意味でこの分類は部分的なものであるが、基本的な分類の枠組は充分示されたと考えられる。

### 3. 省略表現を復元するための情報

何か省略されているという判定が行われたとして、では、省略表現は何であるのかを次に決定しなければならない。そのために利用可能な情報はどのようなもので、またそれはどのような情報源から得られるのかを明らかにして、それを計算機が利用可能な形に定式化することが問題である。3, 4, 5章ではこのような情報の種々相について考察する。

省略表現を復元するための情報源として、文脈、知識、場(状況)などが考えられる。文脈との照応により省略表現を復元する場合には、先行表現が本質的に重要な情報源となることは前述した。文脈中の先行表現以外の表現も様々な知識と関連して省略表現を復元するための情報を提供する。次に、いくつかの例をとおしてどのような形で、情報が得られるのかを検討する。

{ぼくはいそいでそとにでました。[ぼくは]にじのほうへむかって、はしっていきました。}

この例では、“むかう”，“はしる”という動詞の格要素として動作主が必要であるという文法的な知識が省略表現を復元するための重要な情報を与える。このような知識は、まえがきで述べた省略を規定する条件(1)中の‘省かれていると実感される’ことに対する一つの手がかりともなっている。

{2に2を加えると，[その和は]4になる。}

この例では，“2に2を加える”という表現から，加え算をすると結果としてその和が求められるという非言語的な，おそらく小学校で習ったであろう知識が働いて，それが省略表現を復元するための重要な情報となる。

{[これは]なんだろう。へんなものだね。}

この例では、ある状況の中で眼前にある何かを指示して“なんだろう”と言っていると考えると、“これは”という指示の省略表現は、そのような状況（場）が主な情報源となって復元されると考えられる。

{[私は]東京へ行きたい。}

この例では、第一例のように格要素が必要であるという文法的な知識に加えて、“たい”という表現が重要な役割を果たしており、この“たい”という表現によって、省略表現が“私は”であると決定するための手がかりを得ることができる。

{[あなたは]東京へ行きたいですか。}

この例も、うえの例と同様に、“たい”と“か”という表現によって省略表現が“あなたは”であることが確定してしまう。すなわち、これらの例では、“たい”、“か”などの表現に関する意味についての知識が省略表現の復元のための重要な情報を与えている。

文脈からの情報をどのような形で格納するか、文法的な知識、語の意味に関する知識をどのような形で表現し、格納し、どのような時に利用するか、非言語的な知識としてどのようなものが必要で、それをどのような形で利用するか、などという、計算機処理のためのより具体的な手続きについての考察も同時に進められなければならない。

#### 4. 分離先行表現

寺津・山梨<sup>6)</sup>(1979)を参照しながら省略における分離先行表現について考察する。これは寺津・山梨<sup>6)</sup>では分離先行詞とよばれているものである。次の例について考えてみよう。

{太郎は次郎に似ている。しかし[太郎と次郎は]おたがいにそのことに気づいていない。}

省略表現は“太郎と次郎は”であり、照応関係にある先行表現が、先行文脈中で“太郎”と“次郎”という分離した二つの表現となっている。このような先行表現を分離先行表現とよぶ。次の二つの例を比較してみる。

{太郎は次郎に似ている。しかし[太郎は]そのことに気づいていない。}

{太郎は次郎に似ている。しかし[太郎と次郎は]おたがいにそのことに気づいていない。}

第一例の省略表現は“太郎は”であり、第二例の省略表現は“太郎と次郎は”であって分離先行表現を持つ。二つの例で異なる部分は第二例での“おたがいに”という表現のみである。この“おたがいに”という表現が分離先行表現と照応関係にあるということの手がかりを与えている。次の例では、“けんかをする”という表現の言語的な意味が分離先行表現との照応関係を決定する手がかりとなっている。

{あひるが「～」といったら、カニが「・・・」といったので、[あひるとカニは]けんかをしました。}

ただし、この例では、省略表現として[あひるはカニと]という表現も考えられ、これを[あひるは]と[カニと]と分離して、それぞれ先行表現は一つずつと考えることもできる。次の例もこれと同じ型の例である。

{ぼくは、まい日、まさみちとけんかをします。・・・だけど、[ぼくとまさみちは/ぼくはまさみちと]すぐになかよくなって、テレビをみたり、かいじゅうごっこをしてあそびます。}

分離先行表現との照応関係に対する手がかりを与える表現として二種類の表現が分析された。“おたがいに”と同類のものとして、“いっしょに”、“ともに”、“二人で”などが考えられ、“けんかをする”と同類のものとして、“議論をする”などが考えられる。このように、分離先行表



これから次のような命題を推論する。

got (Taro, Math Test, Good Mark)

got (Jiro,  $\emptyset$ , Bad Mark)

この二つの命題から二つの文の間の Contrast Relation を引きだし、 $\emptyset$  に対応するものとして、‘Math Test’ を同定し、結局もとの文での  $\phi$  における省略表現として“数学のテスト”を同定する。このように二つの文の間の接続関係を解明してゆく中で照応関係が解明されてゆき、また逆に、照応関係を解明しながら二つの文の間の接続関係が解明されてゆく過程を、Hobbs は明らかにした。この Hobbs による考察の一部をより形式的に整理してみると次のようになる。連続した二つの文  $S_0, S_1$  とからそれぞれ命題  $P_0$  と  $P_1$  を推論する。 $P_0$  と  $P_1$  は同じ述語を持ち、従って同数の引数要素を持つものとする。 $S_0$  と  $S_1$  の接続関係 Expansion Relations の一部を、この  $P_0$  と  $P_1$  の引数要素の間の意味関係によって定めることができる。それを記号的に示すと次のようになる。

$$\begin{array}{ccc} S_0 & & S_1 \\ \Downarrow & & \Downarrow \\ P_0 & & P_1 \\ P_0 : P(a_1, a_2, \dots, a_n) & & \\ P_1 : P(b_1, b_2, \dots, b_n) & & \end{array}$$

- ①  $a_i$  と  $b_i$  が同じである。 ( $i = 1, 2, \dots, n$ ) (Elaboration)
- ②  $a_i$  と  $b_i$  が似ている。 ( $i = 1, 2, \dots, n$ ) (Parallel)
- ③  $a_i$  が  $b_i$  の下位概念か  $a_i$  と  $b_i$  は同じである。 ( $i = 1, 2, \dots, n$ ) (Generalization)
- ④  $a_i$  が  $b_i$  の上位概念か  $a_i$  と  $b_i$  は同じである。 ( $i = 1, 2, \dots, n$ ) (Exemplification)
- ⑤  $a_i$  と  $b_i$  が相反する。 ( $i \in \{1, 2, \dots, n\}$ ) (Contrast)

概念間の意味関係として重要なものに Partonymy (部分全体関係) とよばれるものがあり、うえの関係にさらに次のような関係をつけ加えることができる。

- ⑥  $a_i$  が  $b_i$  の部分概念か  $a_i$  と  $b_i$  は似ている。 ( $i = 1, 2, \dots, n$ ) (Whole)
- ⑦  $a_i$  が  $b_i$  の全体概念か  $a_i$  と  $b_i$  は似ている。 ( $i = 1, 2, \dots, n$ ) (Part)

(ただし、③、④ではすくなくとも一対の引数要素についてそれぞれ真に下位概念または上位概念であり、⑥、⑦ではすくなくとも一対の引数要素についてそれぞれ真に部分概念または全体概念である。)

Whole と Part という接続関係が新しく設定されたが、これを Hobbs による Contrast と同様に規定すると次のようになるであろう。

Whole :  $S_0$  と  $S_1$  の接続関係が Whole であるとは、命題  $P_0$  と  $P_1$  をそれぞれ  $S_0$  と  $S_1$  から推論することができ、 $P_0$  と  $P_1$  は同じ述語を持ち、 $P_0$  のいくつかの引数要素が  $P_1$  の対応する引数要素に対して部分の関係にあり、他の対応する引数要素の対は似ている時である。

Part :  $S_0$  と  $S_1$  の接続関係が Part であるとは、命題  $P_0$  と  $P_1$  をそれぞれ  $S_0$  と  $S_1$  から推論することができ、 $P_0$  と  $P_1$  は同じ述語を持ち、 $P_0$  のいくつかの引数要素が  $P_1$  の対応する引数要素に対して全体の関係にあり、他の対応する引数要素の対は似ている時である。

それぞれの例を示すと次のようである。

{札幌には小学校が  $m$  校あります。石狩地方では、 $\phi$   $n$  校あります。}

↓

↓

there-is (Sapporo, Elementary School,  $m$ )      there-is (Ishikari,  $\emptyset$ ,  $n$ )

$P_0$  と  $P_1$  から Whole Relation を引きだし,  $\emptyset$  を ‘Elementary School’ すなわち省略表現を “小学校” と同定する。

{太郎は足にけがをしました。φすねの骨を折ったのです。}

↓

injured (Taro, Foot)

↓

injured ( $\emptyset$ , Leg)

$P_0$  と  $P_1$  から Part Relation を引きだし,  $\emptyset$  を ‘Taro’ すなわち省略表現を “太郎” と同定する。‘似ている’ は ‘似ている概念’, ‘同じ’ は ‘同じ概念’, ‘相反する’ は ‘相反する概念’ の意味で用いており, ‘似ている’ は ‘同じ’ を含むものと考えている。

文と文との間の接続関係に関する情報が省略表現を復元するための情報として有効に用いられることが示された。前述したようにこの関係は逆の方向にも働き, 省略表現が復元されることによって文と文との間の接続関係が明らかになるであろう。文から命題を引き出す操作は, 計算機に行わせようとすると非常にむずかしい問題であるように思われる。また, 接続関係の判定の手続き化も, 本質的にはパターンマッチングになるであろうが, それほど簡単なことではない。しかし, 文間の接続関係に関する情報が有用であるということが示されたと思われるので, このような情報を有効にとりこむしくみを, 計算機処理の手続きの中にどのように実現するかということについて次に考えなければならぬであろう。

## 6. 省略処理

計算機により省略処理を実現する時に必要と思われる基本的な処理手続きの概略を前章までの考察結果を考えあわせながら検討してみることにする。まず最初に考えなければならないことは省略文をどのように位置づけるかということである。処理対象である言語の文法が与えられているものとし, 省略文は文としてこの文法では規定されないものとする。つまり, 省略処理はこの言語の文法には含まれないものとする。基本的な処理手続きの枠組は次のようになるであろう。

(1) 省略文かどうかの判定: 省略文であるということは, 最終的には省略表現の復元が正しく行われて省略処理が終了した時点で判定されると考えられるので, 正しくは, 省略文である可能性があるかどうかの判定となる。まず文法的に正しい文と正しくない文に区別され, この文法的に正しくない文が省略文である可能性がある文と真に文法的に正しくない文に区別される。言語の文法によって規定される文章の構成要素間の統語構造が重要な手がかりを与える。2章での構成要素を基本的な単位とする省略の分析がこの判定の手続きの具体化に役立つであろう。

(2) 省略表現の復元: 文法的な知識 (情報) に加えて, 文脈からの情報, 非言語的な知識 (情報) が適当な形で格納されているものとする。計算機による省略処理の実現においてはこれらの知識 (情報) をどのような形で格納すべきかを考えることは, どのような知識 (情報) を格納したらいいかということとあわせて重要な問題である。またそれらの知識 (情報) の利用方法も手続き化されなければならない。

(2.1) 省略表現の予測: 省略表現の予測は主に文法的な制約によって行われるであろう。3, 4, 5章での考察が関連する。

(2.2) 先行表現の探索: 先行表現をもつかどうかの判定はそれほど簡単なことではない。3, 4章での考察におけるように語彙的な手がかりから先行表現をもつということはある程度, 確定することができる場合がある。5章での考察のように文間の接続関係の同定が成功すれば, 先行

表現の同定が比較的容易に行われる場合もある。先行表現の探索は格納されている文脈情報に対して行われる。

(2.3) 省略表現の復元：先行表現が存在する場合はそれをもとに省略表現を復元する。先行表現が存在しない場合は、3章で考察したように非言語的な知識、推論を用いなければならないであろう。また、状況に関する情報も利用しなければならないかもしれない。

実際に計算機による実現を考えるにあたっては、段階的に処理を詳細化してゆくことを考えるべきで、たとえば、省略表現としてあるレベルの構成要素のみを許すようにするとか、文脈情報のみ、特に、有限の先行文脈からの情報のみを用いて省略表現の復元が可能であるようにするとか、省略文であるかどうかの判定は文の構文構造に関する手がかりのみから判定できるようにするとか、その省略処理の対象となる範囲を段階的に限定しながらプログラムの作成を進めてゆくのが良い方法であると考えられる。次章ではこのような観点から一つの実験的な省略処理を行う言語処理プログラムの構成について検討する。

## 7. 簡単な省略処理を含む言語処理プログラム

本章では、簡単な省略処理を含む一つの実験的な言語処理プログラムについて述べる。このプログラムでの簡単な省略処理では、文脈からの情報（省略文の直前の文の構文構造からの情報）を用いて省略表現が復元される。省略することのできる表現は、語または句であり、省略が行われているかどうかの判定はあらかじめ定義された文法にもとづく構文解析が成功したかどうかによって依存して行われる。

計算機による文章理解システムの構成に関する研究の中でも省略処理についての考察が行われている。日本語文章に対する省略処理の機構の計算機プログラムによる実現については、長尾・辻井・田中<sup>9)</sup>(1977)の研究がある。空白必須格の補足、属性名詞・属性値名詞に関する省略処理がそれぞれ、名詞、仮説名詞を格納している M スタック、KM スタックを用いて処理されている。これらは、格要素部、連体修飾部の省略の処理に対応する。

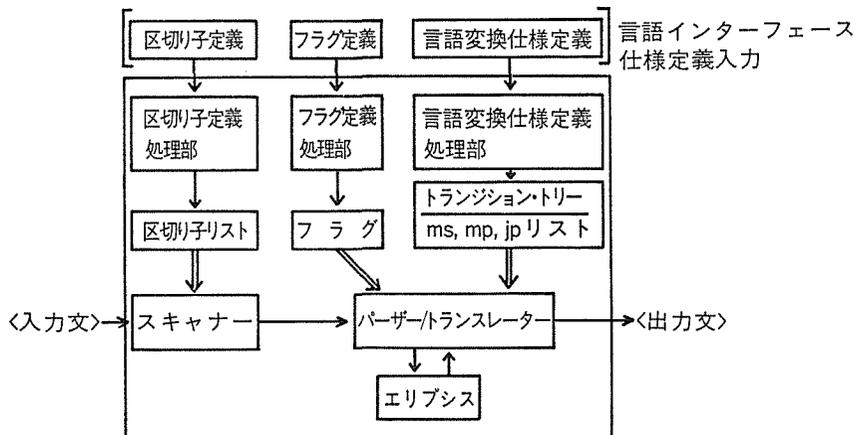


図1 LIFT の構成

詳細についての説明は紙面の都合で省略するが（詳細については参考文献11を参照）以下で説明するプログラムは、LIFT (Language Interface Facility by Translator-generator) とよばれ、Hendrix<sup>10)</sup> (1977) により作成された LIFER を参考にして作成された<sup>11)</sup>。LIFT は、言語インタフェースを作成するシステムである。言語インタフェース仕様定義を入力することによって、LIFT はその仕様に基づいた言語インタフェースを構成する。言語インタフェース仕様定義の処理部、スキャナー、パーザ/トランスレーター、エリプシスからなり、意味文法を基礎とした仕様の形式的記述、ATN を基礎とした構文解析の機構をとりいれている。LIFT は、データベースへの検索、質問応答システムへの質問などにおける、簡単な自然言語のためのインタフェースの構成のための実用的かつ基本的な道具として用いられることを目的として作成された。自然言語が入力言語であるような言語インタフェースにおいては、省略処理は必須であると考えられる。LIFT のエリプシスは、省略処理を行うサブプログラムである。省略処理は正しく処理された直前の文の構文構造の一部を格納しているヒストリースタックを用いて行われる。語または句の省略の一部を処理することができる。

次に示すのは、日本語の一部を英語の一部に変換する簡単な言語インタフェース仕様定義中の言語変換仕様定義の例である。これは言語の変換を定義しているものと考えられ、入力言語と出力の定義を同時に含むもので、| の左側が入力言語の文法の定義、右側が入力言語の文法の定義に依存して付与される出力の定義である。

```
<TG>⇔<DOBUTSU> WA <ZYUTSUBU>. | FLAT <DOBUTSU> IS <ZYUTSUBU>.
<ZYUTSUBU>⇔<DOBUTSURUI> DEARU | <DOBUTSURUI>
<ZYUTSUBU>⇔<DOBUTSURUI> DENAI | FLAT NOT <DOBUTSURUI>
<DOBUTSU>⇔((KUJIRA. (A WHALE)) (ZOU. (AN ELEPHANT)))
<DOBUTSURI>⇔((GYORUI. FISH) (HONYURUI. MAMMAL))
```

この言語変換仕様定義に基づいて LIFT が行う簡単な言語変換の例が次に示される。I: は LIFT への入力、O: はそれに対する LIFT からの出力、EP: は省略処理を行っていることを示す。

```
I: KUJIRA WA GYORUI DENAI.      (鯨は魚類でない。)
O: A WHALE IS NOT FISH.         (A whale is not fish.)
I: HONYURUI DEARU.             (哺乳類である。)
EP:
    KUJIRA WA HONYURUI DEARU.   (鯨は哺乳類である。)
O: A WHALE IS MAMMAL.          (A whale is mammal.)
I: ZO WA.                       (象は。)
EP:
    ZO WA HONYURUI DEARU.       (象は哺乳類である。)
O: AN ELEPHANT IS MAMMAL.      (An elephant is mammal.)
```

省略処理がどのように行われるかを簡略に示すために、うえの例中の次のような二つの文の連続について考える。

<1> 鯨は魚類でない。

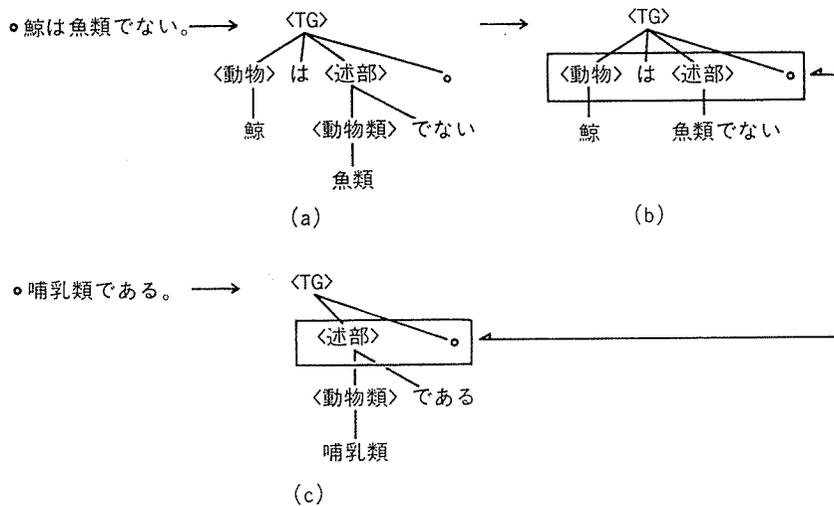


図2 省略処理における構文構造からの情報の利用

<2> 哺乳類である。

<2> の文は省略文であり、省略表現は“鯨は”である。LIFT は<2> の文を、次に、<3> の文のように復元する。

<3> 鯨は哺乳類である。

<1> の文を入力すると LIFT はこれをあらかじめ与えられている文法定義に基づいて構文解析することによって図 2(a) のような構文構造 (木) を作り、この構造を再構成して図 2(b) のような構造をヒストリースタックに格納する。次に<2> の文を入力すると、パーザ/トランスレータによる構文解析は失敗し、文法的に正しい文か、処理可能な省略文のみが入力されると考えると、このことは入力文が省略文であるということになり、エリプシスによる省略処理を行う。省略処理は、(b) と入力文の構文構造 (c) を照合することによって、入力文に欠けている部分を、(b) から補う。この場合、照合は□で囲まれた部分について行い、欠けている部分の補充は終端記号の文字列によって行う。この例では、“<述部>。”という部分パターンが一致し、“<動物>は”という部分パターンが不足していることになり、“<動物>は”という部分パターンに対応する終端記号列である“鯨は”を補うことにより<3> の文が生成される。このように LIFT における省略処理は、正しく処理された前文の構文構造から情報を得て、その前文と同じ構文構造をもつ文の省略の処理を行うものである。簡単な省略処理ではあるが、(1) 省略が行われているということの判定、(2) 文脈情報の表現と格納方法、(3) 文法的な知識の利用、(4) 省略表現の復元の機構などがコンパクトに計算機プログラムの中に実現されている。文脈情報が構文構造としてとらえられているが、これはもちろん、あらかじめ与えられている文法に関する知識が利用されてはじめて可能なのであり、文法的な知識と文脈からの情報の合成物としての情報であるというのが正確な言い方であろう。

## 8. あとがき

計算機による日本語理解システムの実現のための一つの基礎を与えることを目的として、文章を構成する構成要素の単位に基づく省略の分類を行い、省略表現を復元するための情報と文脈、知識、場との関連についていくつかの考察を行った。また、実験的な観点から、省略処理の計算機による実現の可能性を探るために、統語構造からの情報によって簡単な自然言語文の省略処理を行う計算機プログラムを作成し、省略処理の機構の手續き化についての基礎的な検討を行った。

省略には本報告では考察できなかった現象もあり、さらにそれらの分類整理が必要である。また、本報告においても、文の統語構造、文間の接続関係、語の意味、非言語的な知識などが省略において果す役割について考察を進めてきたが、省略表現を復元するために用いられる情報について、さらにいろいろな角度からの考察が行われなければならないであろう。このような実際の省略現象のさらに進んだ分析と平行して、省略処理を計算機プログラムとして実現するための方法の詳細な考察も進められなければならない。

## 参 考 文 献

- 1) M. A. K. Halliday, R. Hasan: *Cohesion in English*, Longman (1976)
- 2) 田中穂積: 談話理解の構造, 情報処理, 20-10 (1979)
- 3) 三上 章: 文法小論集, くろしお出版 (1970)
- 4) 久野 暲: 談話の文法, 大修館 (1978)
- 5) 森岡健二: 伝達論からみた省略, 言語生活, 339 (1980)
- 6) 寺津典子, 山梨正明: 日本語における照応現象について (その一), 「計算機による日本語談話行動の総合モデル化」文部省科研費研究報告 (代表者 石綿敏雄) (1979)
- 7) J. R. Hobbs: Why is discourse coherent?, *SRI Int. TN 176* (1978)
- 8) J. R. Hobbs: Coherence and coreference, *Cognitive Science* 3-1 (1979)
- 9) 長尾 真, 辻井潤一, 田中一敏: 意味および文脈を考慮した日本語文の処理, 情報処理, 17-1 (1976)
- 10) G. G. Hendrix: *The LIFER manual*, *SRI Int. TN 138* (1977)
- 11) 中川貴之, 桃内佳雄, 宮本衛市: LIFT (Language Interface Facility by Translator-generator) による簡単な言語インターフェースの構成, 情報処理学会記号処理研究会報告, SP 12-6 (1980)