

省略を含む日本語テキストの復元システム†

唐 沢 博** 堂 坂 浩 二††
小 川 均†† 田 村 進 一††

日本語における省略現象は多様であるが、本研究では助詞の省略、前文とのアナロジーに基づく省略、予測文脈に基づく省略の3種類に制限し、そのような省略を含む日本語テキストを解析する手法を開発した。解析は構文のおよび意味的な関係を同時に調べながら進められ、さらに文脈的な情報が適切な解析を助ける。構文解析と意味解析は、オブジェクト指向モデルを導入することによって融合し、メッセージ交換を通じて解析が実現するようにした。助詞の省略は意味解析の過程で復元される。メッセージ交換が完結しない場合は、文脈情報に依存した省略が行われていると考える。そこでまず前文とのアナロジーに基づく復元のプロセスが起動され、格構造の比較を通じて必要な文要素を借りてくることにより復元を試みる。アナロジーがとれない場合は予測文脈に基づく復元のプロセスが起動され、MOPsによって表現された文脈記憶構造から推測される情報を適用することによって復元を試みる。また省略を含む文では、その解釈の多義性を解消することが重要である。そのために多世界モデルを導入して多義性に対処した。以上の考えに基づくコンピュータシステムをProlog上に実現した。

1. ま え が き

日本語テキストにおいては省略表現が多用される傾向にある。その目的は冗長な表出を極力避けて可能な限りコミュニケーションを円滑にしようとするのだと考えられる。この目的はマンマシン・インタフェースにもそのまま適用しうる。筆者らは先に英文テキストについて同様の考察を行った¹⁾。そこで明らかとなった課題は、補完体系と呼ぶ省略情報の復元手法が深層構造におけるスクリプト的な概念の利用を含んでいなかったために、人間ならば明らかに復元しようような内容のテキストに対して無力であったことである²⁾。本研究の動機は、この課題を解決することにある。日本語は英語よりも深層に近い体系をしているという指摘があり³⁾、ゆえに深層レベルの復元力が強力に効いてくるものと考えた。このような理由から対象言語を日本語とした。日本語における省略は一見任意に行われているようにも見え、扱いが複雑であり、それゆえ自然言語処理研究においても一般化された省略現象の研究はほとんど例がない。本論文では、省略現象を3種類に大別して省略・復元モデルの中に位置付け、そのモデルに基づき復元手法を開発した。省略文の解析は、構文・意味・談話の各知識を同時に使いながら進

める。解析に伴う解釈の多義性は、多世界モデルと解釈可能性に対する評価基準の導入で解決した。

2. 日本語文における省略現象

日本語文における省略現象にはさまざまなパターンが見られるが、本研究では(1)助詞の省略、(2)前文とのアナロジーに基づく省略、(3)予測文脈に基づく省略の3種類の省略現象を対象とする。これらは図1に示されるように言語現象の表層から深層に渡って位置付けることができる。表出すべき概念が最終的な表層構造をとるまでの過程で冗長な情報が捨てられていく。一方、聞き手(読み手)は、省略を含む表層構造について構文・意味・談話に関する知識をもとにして、省略されている情報を復元し発話者の意図していた概念を得る。以上のようなモデルに基づいて省略現象を扱う。以下に3種類の省略現象について説明する。

2.1 助詞の省略

係り受け関係に曖昧性が少ない場合、助詞が省略されることがある。例えば、

『太郎 りんご 好きだ。』

という省略文は、

『太郎はりんごが好きだ。』

と解釈できる。

2.2 前文とのアナロジーに基づく省略

前文をもとにして新しい情報のみ表出するような場合に見られる省略であり、

『太郎はコーヒーを飲んだ。』

† A System for Recovering Ellipsis in a Japanese Text by HIROSHI KARASAWA, KOUJI DOUSAKA, HITOSHI OGAWA and SHINICHI TAMURA (Department of Information & Computer Sciences, Faculty of Engineering Science, Osaka University).

†† 大阪大学基礎工学部情報工学科

* 現在 京都教育大学教育学部教育実践研究指導センター

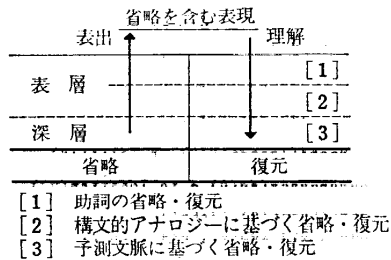


図1 省略・復元モデル
Fig. 1 Ellipsis-recovery model.

次郎も。』
という文章において第2文は、

『次郎もコーヒーを飲んだ。』
と解釈できる。この場合、第2文に潜在する『コーヒ
ーを飲んだ』という情報は、表出の時点で冗長である
として省略されたものである。

2.3 予測文脈に基づく省略

文章の内容が典型的な日常の場面を伴っているよう
な場合、聞き手の常識的な推論に依存した省略が多用
される。現文章に固有の情報のみ表出されればよいの
で、個々の文だけを取り出すと明らかな非文や曖昧性
の非常に大きな文が見いだせる。

『太郎は次郎とレストランへ行った。
次郎はメニューを見つめて迷っていた。

太郎はカレーライスだ。』
という文章では、「レストランで食事をする」という
場面から予測される文脈下でのみ第3文の意味が一意
に定まる。

3. システムの構成

図1に示された省略・復元モデルの復元過程部分を
システムとして実現した⁴⁾。省略を含む日本語テキ
ストが入力され解析される。解析部では構文および意味
の解析が同時に進行し談話レベルの知識や文脈情報も
利用される。そして解析結果は深層構造として得られ
る。得られた深層構造が適切であり、かつ表層文にお
いて省略されていた情報が深層構造中に見いだせると
き、省略が復元されたと考える。図2にシステム構成
を示した。

4. 省略文の解析

省略を含む文の解析では統語構造の不完全さや曖昧
性がつきまとうので、意味解析を同時に進める必要が
ある。この問題を解決するためにオブジェクト指向モ
デルを用いた。

- 1) 各単語をオブジェクトに対応させる。

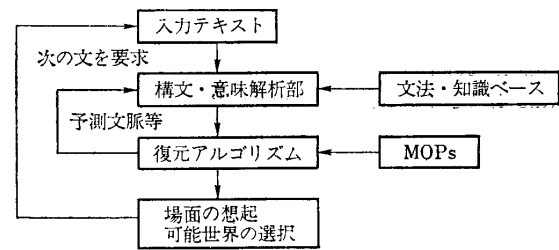


図2 システム構成
Fig. 2 Construction of the system.

- 2) 構文・意味解析の知識を各オブジェクトに分散
する。
- 3) メッセージの交換を通じて構文・意味解析をす
る。

この手法は、関連のある知識がひとかたまりに記述
でき、柔軟な記述が可能であるという利点をもつ。

4.1 メッセージ交換による解析

入力テキストは文単位に分割され、各文単位が分か
ち書き処理によって単語に分割されて、各単語が対応
するオブジェクトに置きかえられているものとする。
その後、以下に示す手続きに従って解析が遂行され
る。

- 1) 各オブジェクトのメッセージの送信先が決定さ
れる。すなわち、オブジェクトの並び (M はオブジ
ェクト数) を $\{w_m\} = w_1 w_2 \dots w_m \dots w_M$ とおいたとき、
オブジェクト w_m のメッセージ送信先は w_{m+1} であ
る。
- 2) モニタが w_1 に解析開始を指示するメッセ
ージを送信する。
- 3) w_m は w_{m-1} からのメッセージを構文・意味知
識を用いて理解し、その結果を w_{m+1} へ送信する。
- 4) 最終的に w_M はモニタへメッセージを返す。
メッセージ交換には次のような規則を設ける。
 - a) w_m の受信したメッセージが受理可能な場合は
そのメッセージを最初に出したオブジェクト w_n (n
 $< m$) と w_m との間に係り受けの関係が成立するもの
とする。メッセージが受理可能とは、 w_n と w_m とが
構文・意味的に係り受けの関係にある場合に限る。
 - b) 任意のオブジェクト間の係り受け関係は交差し
ない。
 - c) w_m の受信したメッセージが受理不可能な場合
は、そのメッセージを w_m 固有のメッセージとともに
 w_{m+1} に送信する。

以上のプロセスは、構文解析の観点から見れば bot-
tom up parsing に等価である。図3には本研究で用
いた文法規則が示されている。Rule 6 は助詞が省略

- (Rule 1) S → VC 1.
 (Rule 2) VC 1 → VC.
 (Rule 3) VC 1 → NC, VC 1.
 (Rule 4) VC 1 → NC, Analogical-VC.
 (Rule 5) VC 1 → MOPs-VC.
 (Rule 6) NC → NP.
 (Rule 7) NC → NP, Case-marker.
 (Rule 8) NC → S, NC. (埋め込み文)

S: 文

VC: 述語節

NC: 名詞節

NP: 名詞句

Case-marker: 助詞

Analogical VC: 前文とのアナロジーによって得られた述語節

MOPs-VC: 予測文脈によって得られた述語節

図 3 文法規則

Fig. 3 Grammar rules.

```

tag,
(oblig_msg, ((1, Actor), single), ((2, Object), single)),
(no_oblig_msg, ((3, Loc), single)),
(syn_feature, (sub, 1), (obj, 2)),
msg,
  np,
    ((1, Actor),
      (syn, (case(Actor, (ga, wa)), 1, 0),
        (case(Actor, 0, 0.5)),
        (sem, (concept(Actor, living), 1, 0),
          (not(concept(Actor, living)), 0, 1))),
    ((2, Object),
      (syn, (case(Object, (wo)), 1, 0),
        (case(Object, 0, 0.5)),
        (sem, (concept(Object, liquid), 1, 0),
          (not(concept(Object, liquid)), 0, 1))),
    ((3, Loc),
      (syn, (case(Loc, (de)), 1, 0),
        (case(Loc, 0, 0.5)),
        (sem, (concept(Loc, location), 1, 0),
          (not(concept(Loc, location)), 0, 1))) ),
(sem_representation,
  (drink,
    (actor, Actor),
    (object, Object),
    (loc, Loc) ) ) ).
  
```

図 4 オブジェクト「飲む」の例

Fig. 4 Example of object 'drink'.

されている場合に適用される。助詞の省略はメッセージ交換の過程で復元される。

4.2 オブジェクトの表現形式

各オブジェクトは受理可能なメッセージのパターンや構文・意味的な制約に関する知識、そして生成すべき概念構造をもっている。一方、メッセージの授受に関する手続きは各オブジェクトに共通なため、オブジェクトとは独立に存在している。オブジェクト「飲む」の例を図4に示す。

5. 深層構造

本研究では予測文脈情報を得るために MOPs モデル⁵⁾を用いているので、深層構造は概念依存構造⁶⁾とした。ただし、すべての概念をプリミティブ・レベルに変換するのは非効率なので推論過程で必要が生じた場合のみ変換するように考慮した。

```

      P      D      | -> New York
John <=> go <- | -<
  
```

(a) 概念構造

```

concept(go(0), act, go).
concept(human(0), pp, human).
concept(location(0), pp, location).
dependency(go(0), actor, human(0)).
dependency(go(0), d_to, location(0)).
tense(go(0), past).
name(human(0), 'John').
name(location(0), 'New York').
  
```

(b) Prolog による表現

図 5 Prolog 上での深層構造表現

Fig. 5 Deep structures in Prolog.

5.1 Prolog 上での深層構造の表現

Prolog の term としてさまざまな概念を表現する。例えば図5(a)の概念構造は、Prolog システム上では図5(b)のように表現される。

6. 前文とのアナロジーに基づく復元

前文の構文情報を利用して新しい情報のみ表出するような省略では次のような現象が生じる。

i) 係るべき文要素が省かれている。

例: 主題や必須格の欠如

ii) 係られるべき文要素が省かれている。

例: 述語節の欠如

メッセージ交換による解析において、i) はオブジェクトが受理すべきメッセージの欠如、ii) はメッセージを受理すべきオブジェクトの欠如として取り扱うことができる。

6.1 メッセージの欠如

これは(1)主題の省略、(2)穴あけ構文に大別できる。主題は古い情報を表している代表的なものであり、「主語+は…」の形をした文における主語が主題を示すものとし、前文に主題が提示されていればそれを借りてきて、主語の欠如を補う。

例)『太郎は学生だ。

(太郎は) 自転車で登校する。

(太郎は) 背が高い。』

一方、穴あけ構文の場合は前文と現行の文の格構造を比較し、同一の構造が認められた場合に限り前文の格情報を借りてくる。

6.2 オブジェクトの欠如

一般に被修飾語や述語節が省略されると、どのオブジェクトにも受理されないメッセージが発生する。特に動詞は格関係のまとめ役を演じるので、動詞が省略されると多くのメッセージが遊離する結果となる。幸いにも、省略された動詞は前文を参照することによって得られることが多い。したがって、省略された動詞

は次のようなアルゴリズムによって復元することができる。

step-1 前文までの解析結果をさかのぼって本動詞を含む述語節を借りてくる。

step-2 現在遊離しているメッセージを借りてきた動詞に送ってみて受理可能か否か調べる。受理されない場合はさらにさかのぼって本動詞を含む述語節を借りてきてこの step を試みる。借りてくる述語節が尽きた場合は予測文脈に基づく復元処理に切り換える。

step-3 借りてきた動詞が適切であり、かつその動詞の受理すべきメッセージが不足している場合は前節で述べた処理プロセスを適用する。

7. 予測文脈に基づく復元

聞き手が話しの内容展開を予測できる場合は、わずかな情報から多量の情報を復元できる。そこで話し手は聞き手の復元能力に依存した省略を多用することになる。このような状況で表出される文は、断片的な文要素のみであったり、曖昧性の大きな文になる傾向が強い。そのような文を意味解析まで成功させるためには、その文が属する文脈中の、陽に表出されない大量の概念を復元しなければならない。この要求を満たすためにはスクリプト⁶⁾が適しているが、本研究ではスクリプトよりも能力の高い MOPs モデルを導入した。

7.1 MOPs の文脈予測能力

MOPs は記憶モデルの一種であり階層構造的な場面の表現ができる点でスクリプトよりも優れている⁷⁾。MOPs 理論に従えば、人間が新たなテキストを理解するとき、過去の経験の場面とアナロジーをとって一致点や不一致点を判別しながら予測文脈を引き出していく。また、関連する複数の場面も呼び起こしていく。MOPs のこのような特性は予測文脈に基づく復元に対して効果的である。

7.2 MOPs による理解と復元

MOPs の記憶構造は、その活性状態が現在の文脈を反映している。入力文の概念構造がその活性部分と矛盾なく適合するとき適切な理解が達成されたと考えれば、活性状態にある MOP のいずれかに言及することにより、その他の活性化 MOP が、言及されなかった概念に該当することになる。こうして、予測可能なために言及が省略された概念が記憶構造内に復元される結果となる。

```
mop(mop_name,
    tag_list,
    remind_list,
    oblig_list,
    condition_list,
    scene_list).
```

図 6 MOP の構造
Fig. 6 Construction of MOP.

```
[scene(1,
  lead_to([scene(1,1),...,scene(1,n)]),
  memory_pointer(mem_str_name,tag_list)),
  :
  scene(m,
  lead_to([scene(m,1),...,scene(m,k)]),
  memory_pointer(mem_str_name,tag_list))]
```

図 7 scene-list の構造
Fig. 7 Construction of scene-list.

```
mop(go_to_restaurant,
  (actor=Actor, food=Food, location=Location),
  0,
  (Actor, Location),
  (concept(Actor, pp, human),
  concept(Food, pp, food),
  concept(Location, pp, restaurant)),
  (scene(1,
    lead_to(0),
    memory_pointer(go,
      (actor=Actor, d_to=Location))),
  scene(2,
    lead_to(0),
    memory_pointer(eat,
      (actor=Actor, object=Food, loc=Location)))).
```

図 8 「レストランへ行く」 MOP の例
Fig. 8 Example of MOP 'go_to_restaurant'.

7.3 Prolog 上での MOPs の表現

MOPs は Prolog の term として図 6 のように表現する。

- 1) mop_name は MOP の識別子である。
 - 2) tag_list は MOP が他の記憶構造から呼ばれるときのためのタグのリストであり、各要素は、タグ名 = タグ変数の形をしている。他の MOP はタグ名によって情報を結び付ける。タグ変数は MOP 内で使われる局所変数である。
 - 3) remind_list は、この MOP を呼び起こす MOP 名のリストである。
 - 4) oblig_list は MOP が起動されるのに必要なタグに対応するタグ変数のリストである。
 - 5) condition_list はタグが満たすべき条件を記述する。
 - 6) scene_list は連続する場面のリストから成り、図 7 の構造をしている。すなわち、lead_to は次に生起する場面へのポインタであり、memory_pointer はその場面で呼び起こすべき記憶構造 (mem_str_name) へのポインタである。
- 図 8 に「レストランに行く」 MOP の例を示した。次に、MOP の起動条件を示す。

- ① MOP が含む少なくとも一つの mem_str_name

が入力文の概念構造の一部と一致する。

- ② oblig_list の変数がすべて代入されている。
- ③ condition_list の条件の否定が証明されない。

8. 多義性への対処

一般に、構文解析、意味解析の過程では多義性が生じるが、省略を含む文の解析においてはその傾向が促進される。よって、多義性を扱うための特別な工夫が必要となる。本研究では、各文の解析によって得られる解釈が一つの信念のもとに行われているものとし、その信念が成立するところの世界を定め、多義性を多世界モデル⁸⁾として扱うことにした。

8.1 多世界モデル

個々の文の解釈はテキスト全体の文脈に依存して決定される。したがってテキスト中のすべての文の解釈が終了した時点で初めて各文の正しい解釈が定まる。それまでは複数の妥当な解釈を保持している必要がある。いまテキスト中に連続する文 (N は文の個数) を $\{s_m\} = s_1 s_2 \dots s_m \dots s_N$ とおいたとき、文 s_m の可能な解釈 (K は解釈数) を $\{I_{m,n}\} = I_{m,1} I_{m,2} \dots I_{m,n} \dots I_{m,K}$ とし、 $\{I_{m,n}\}$ と 1 対 1 に対応する有向グラフの頂点を $\{x_{m,n}\}$ とする。同様に、解釈 $\{I_{m,n}\}$ のもとでの文 s_{m+1} の解釈を $\{I_{m+1,p}\}$ とし、それに対応する頂点を $\{x_{m+1,p}\}$ としたとき、頂点 $\{x_{m+1,p}\}$ から頂点 $\{x_{m,n}\}$ へ有向辺を引くことによって構成されるグラフ W を考える。ただし、頂点 $\{x_{1,n}\}$ は頂点 dummy に有向辺で結ばれる。図 9 にグラフ W を示した。解釈 $\{I_{m+1,p}\}$ は解釈 $\{I_{m,n}\}$ に依存しているので、頂点 $\{x_{m,n}\}$ から頂点 dummy までの有向路上にある頂点集合に対応する解釈の集合は一貫した信念のもとに決定されたものであり、文 s_m を解釈するまでに形成された一つの世界観を表している。文 s_m について解釈 $I_{m,n}$ をとるときは、この世界観の上において行う。

8.2 Prolog 上での多世界モデル

頂点 $x_{m,n}$ に対して Prolog の atom である db_name と正整数 $N(m,n)$ を与える。 $x_{m,n}$ に対応する解釈を形成している概念は次のように Prolog 上の term で表される。

db_name ($N(m,n)$, Concept).

Concept は Prolog の term で表された概念である。次に、 $x_{m,n}$ から出ている有向辺に接続している頂点を $x_{m-1,j}$ とするとき、それに対応する Prolog の atom を db_name', 正整数を $N(m-1,j)$ として次のような term で有向辺を表す。

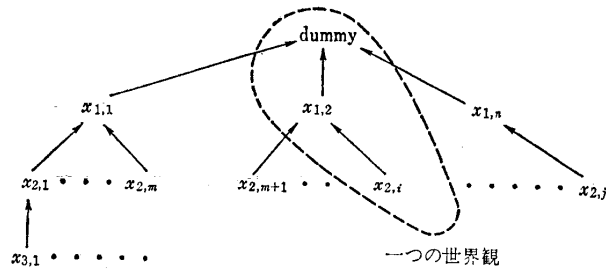


図 9 多世界を表すグラフ W
Fig. 9 Graph W representing many-world.

```
demo(DB, Id, sys(P), CUT) :- P.
demo(DB, Id, true, CUT) :- !, true.
demo(DB, Id, !, CUT).
demo(DB, Id, !, cut).
demo(DB, Id, (P;Q), CUT) :- !,
    (demo(DB, Id, P, CUT);
     demo(DB, Id, Q, CUT)).
demo(DB, Id, (P,Q), CUT) :- !,
    demo(DB, Id, P, P, CUT),
    (P-CUT == cut, CUT = cut ;
     demo(DB, Id, Q, CUT)).
demo(DB, Id, not(P), CUT) :- !,
    not(demo(DB, Id, P, CUT)).
demo(DB, Id, P, CUT) :-
    clause_db(DB, Id, P, Q),
    demo(DB, Id, Q, CUT),
    (CUT == cut, !, fail;
     true).

clause_db(DB, Id, P, Q) :-
    X =.. [DB, Id, P],
    X,
    Q = true.
clause_db(DB, Id, P, Q) :-
    X =.. [DB, Id, (P:-Q)],
    X.
clause_db(DB, Id, P, Q) :-
    db_scope(DB, Id, DB1, Id1),
    clause_db(DB1, Id1, P, Q).
```

図 10 demo 述語
Fig. 10 Demo predicates.

db_scope (db_name, $I(m,n)$, db_name'
 $I(m-1,j)$).

各頂点から db_scope の情報を使ってたどれるすべての頂点 (自身も含む) の表す解釈の集合は、Prolog 上で定義された、他とは異なる独立な一つの internal data base を構成している。他の data base とは独立なその internal data base 上での demo 述語⁹⁾を図 10 に示す。この demo 述語は一つの世界観の上だけで述語の証明機構を提供する。各 demo 節は、システム述語の評価、'true', カットオペレータ、選言、連言、否定、ルールおよびファクトの処理を行う。また clause_db 述語は、internal data base DB から述語 P を頭部にもつ節を探す。このようにして文の解釈は、それが解釈される世界観の上での証明可能性に還元される。

8.3 世界観の可能性に対する評価

テキストの解析処理が進むにつれて世界観の個数は一般に増大していく。そのため、世界観の可能性のより高いものを選び出し低いものは凍結することによってその個数を絞り込む必要がある。凍結された世界観はテキストの解釈が進んだ先で大きく意味が転換する

表 1 視点情報を与える動詞
Table 1 Verbs controlling a focus.

動 詞	視点の位置
ク レ ル	与格目的語
ヤ ル	主 語
…テクレル	非 主 語
…テヤル	主 語
モ ラ ウ	主 語
ヨ コ ス	与格目的語

ような場合に解除されるものとする。このような目的のために世界観の可能性を判定する4種類の評価基準を設けた。世界観の評価は、それらの評価から総合評価値を出すことによって実現される。総合評価値があらかじめ設定した規定値を下回るような場合、その世界観は凍結され、かわりにもっとも最近凍結された世界観の下で総合評価値を計算し、その結果が規定値を上回れば凍結が解除される。なお、規定値は実験的に決定する必要がある。以下に4種類の評価基準を示す。

8.3.1 オブジェクトに対するメッセージの許容度

メッセージを受理する許容度を設けることによってその解析の可能性の程度を示す。オブジェクトは、解析の「もっともらしさ」を示す値を有し、文全体の信頼度は各オブジェクトが提示する値の積で表される。この信頼度とその時点での世界観の評価値とが比べられ、世界観の評価値を低下させるような解析は中止する。

8.3.2 省略の順序に関する制約

省略の順序に関して次の制約が働く。

省略は、より古い(重要度のより低い)情報を表す要素から、より新しい(より重要な)情報を表す要素へと順に行う¹⁰⁾。

この制約によって、古い情報を表す要素が省略されたとする解釈をもつ世界観の方により高い可能性を与える。

8.3.3 視点情報に関する制約

視点情報に関する世界観の評価では、テキスト全体を通じて視点関係の無矛盾性が保たれた世界観ほど高い評価を与える。視点情報は表1に示されるような授与動詞等から得る¹⁰⁾。

8.3.4 場面のつながりによる制約

予測文脈に基づく復元過程に関連した評価基準を次に示す。

ある文を理解するとき、すでに想起されている場

太郎はレストランへ行った。
太郎はカレーライスだ。
(a) 入力テキスト

```

sentence 1
concept (go (0), act, go).
tense (go (0), pas t).
dependency (go (0), actor, human (0)).
dependency (go (0), d_to, res taurant (0)).
concept (res taurant (0), pp, res taurant).
concept (human (0), pp, human).
name (human (0), tarou).
sex (human (0), male).
processing_pointer (go (0), go_to_restaurant, (actor=human (0), food=345,
location=res taurant (0), scene (1, lead_to (2)), memory_pointer (go,
(0), actor=human (0), d_to=res taurant (0), NIL)).

sentence 2
concept (curry_rice (0), pp, curry_rice).
concept (human (1), pp, human).
name (human (1), tarou).
sex (human (1), male).
identity (human (1), human (0)).
concept (res taurant (3), pp, res taurant).
identity (res taurant (3), res taurant (0)).
concept (eat (0), act, eat).
dependency (eat (0), loc, res taurant (3)).
dependency (eat (0), object, curry_rice (0)).
dependency (eat (0), actor, human (1)).
processing_pointer (eat (0), go_to_restaurant, (actor=human (1),
food=curry_rice (0), location=res taurant (3), scene (2, lead_to (0),
memory_pointer (eat, (0), actor=human (1), object=curry_rice (0),
loc=res taurant (3)), NIL)).
    
```

(b) 復元結果

図 11 処理例

Fig. 11 Example of processing.

面と重なるような解釈は、そうでない解釈よりも可能性が高い。

この制約は解釈の妥当性を決定する本質的なものであるので、他の評価基準よりも大きな重みを与えた。

9. インプリメンテーション

復元システムは、VAX-11/750 (DEC 社) を用いて、C-Prolog ver. 1.2 a (OS UNIX 4.1 BSD) 上に作成した。システムの clause 数は核の部分が約 600 である。図 11 に実行例を示した。実行時間は、図 11 の例の場合で約 9 sec cpu time であった。

10. 考 察

日常的な内容の会話を対象にした日本語テキストについて省略されている情報の復元を試みた。その結果、本研究において制限を加えた範囲内での省略現象についてはいずれもうまく解析が行われ、その解釈も妥当なものだった。また、以下に示すような点が明らかとなった。

- 1) 前文とのアナロジーに基づく復元は、かなり複雑な構文も扱える。またその復元の曖昧さも省略の順序による制約で解決可能なものが多い。
- 2) MOPs の呼び出しに非常に時間を費やすためその想起法に工夫を要する。
- 3) MOP 中の二つ以上の場面が一つの文で表現されている場合には対処できず、その結果単文しか復元

できない。

4) 理解の深さは MOPs の記述の詳しさに依存するところが大きい。MOPs の作成には多大なコストがかかる。MOPs のもつ学習能力⁷⁾を積極的に利用していく研究が必要である。

5) 理解の広さは MOPs の量に依存するが、実装量が多くなればそれに伴って呼び起こされる MOPs や場面も多くなる。その量を適当なものに制限する方法が開発されねばならない。

6) 本論文で述べた手法では、概念関係のアナロジーを利用していないために、以下に示すような例が復元できない。

『リスプ マッカーシー。
パスカル ヴィルト。』

第1文において「リスプ is-designed-by マッカーシー」という概念関係が復元され、その関係が第2文に適用されて初めて妥当な解釈が得られると考えられる。

本研究は次のような応用が考えられる。

- a) 音声認識における曖昧性の解消
- b) 知識ベースに対する柔軟なインタフェースの設計
- c) 人間の言語認知プロセスのシミュレーション

11. む す び

省略表現の多い日本語テキストがいかんして理解されるのかという課題に基づき、省略された情報を復元するシステムを Prolog 上にインプリメントした。対象とする省略表現は、(1)助詞の省略、(2)前文とのアナロジーに基づく省略、(3)予測文脈に基づく省略に制限し、それらに適合した構文・意味解析手法を確立した。また、解釈の多義性の扱いは多世界モデル上での証明可能性として実現した。今後は知識ベース・

インタフェースとして実装し、実際の対話プロトコルの分析を通じて本手法を評価していく予定である。

謝辞 最後に、討論していただいた研究室の諸氏に感謝いたします。また、本研究の一部は文部省科学研究費「特定研究」による。

参 考 文 献

- 1) 唐沢, 松永, 小川, 田村: 省略表現を含む不完全テキストを知識を用いて復元する英文補完システム, 情報処理学会論文誌, Vol. 26, No. 2, pp. 374-381 (1985).
- 2) 唐沢, 小川, 田村: 英文補完システムの補完能力評価, 情報処理学会自然言語処理研究会資料, 43-4 (1984).
- 3) チャールズ, J., フィルモア: 格文法の原理, p. 373, 三省堂, 東京 (1975).
- 4) 唐沢, 堂坂, 小川, 田村: 省略を含む日本語テキストの復元に関するコンピュータ・モデル, 日本認知科学会第1回大会発表論文集, B-2, pp. 32-33 (1984).
- 5) Schank, R. C.: Reminding and Memory Organization: An Introduction to MOPs, Yale Univ. Research Report #170 (1979).
- 6) Schank, R. C. and Abelson, R. P.: *Scripts Plans Goals and Understanding*, Lawrence Erlbaum Assoc., Inc., New Jersey (1977).
- 7) Schank, R. C.: Language and Memory, *Cognitive Science*, Vol. 4, pp. 243-284 (1980).
- 8) 国藤, 麻生, 竹内, 宮地, 北上, 横田, 安川, 古川: Prolog における対象知識とメタ知識の融合とその応用, 情報処理学会知識工学と人工知能研究会資料, 30-1 (1983).
- 9) 北上, 麻生, 国藤, 宮地, 古川: (2) 知識同化機構の一実現法, 情報処理学会知識工学と人工知能研究会資料, 30-2 (1983).
- 10) 久野: 談話の文法, p. 332, 大修館書店, 東京 (1978).

(昭和59年12月17日受付)

(昭和60年7月18日採録)