言語・画像を利用した行動の解釈（2）†

——粗筋抽出と質疑応答——

安部憲広†† 曽我厳哉††† 辻 三郎††

簡単な画像とこれに対するナレーションを与え，ナレーションで言及された対象物体を画像内で探索すると同時に，画像データを利用して，一意的に意味の定まらない言語表現を解析して，ナレーションの記述と画面に表現されている事象の関連をもとめて，話の筋を抽出する．この場合，まず個々の画面での言語と画像との関係を求め，次に画面相互の関連付けを行うことにより，全体的な話の流れを把握する．それはナレーションによる記述を中心としたものと，画面内での対象物体の位置的記述と，全く注目されず原画の状態のまま残された物という3層に分けて，管理される．そうすることにより，不必要な記述の量が減少するとともに，要求される事象や関係の探索が必要に応じて効率良く行われる．これは画像理解にとって不可欠な点であり，粗筋抽出後のシステムへの質問応答で実際に適用される．最後に，本研究の改善すべき点などについても述べる．

1. まえがき

言語・画像間の相互参照に基づく粗筋の抽出を目的とした実験に必要な，物体の画像からの抽出方法について文献1）で報告した．本稿では，与えられたナレーションと画像とを相互参照することにより，粗筋抽出を行うシステムの詳細について報告する．

システム内では，ナレーション及び対象に関して得られた記述はすべて，Schank の提唱した CD 理論に基づいて記述される．この理由の1つは，表層言語表現を統一した内部構造で表現することにより，その多様性を吸収するという CD 理論本来の目的によるものである．現在さまざまな言語処理システムが提案されているが，非常に複雑な文を扱える反面，文例が極度に制限されたものが少なくない．本研究では，そのような複雑な文を対象とせず，その代り多彩で平明な表現の理解を目的としている．すなわち，我々が英語の初学者と対話するような状況下で行うのと同じ程度の英文をナレーションとして与える．しかし構文的には平易であっても，構文則のみでは多義語や係り受けの解釈は不可能である．我々は気軽に have, take, get 等の語を使うが，それは使用の脈絡が相互に理解されていると信じているから，多用できるのである．句の係り方や文相互の関連も同じである．本システムは，この脈絡に当たる状況を主として画像情報に求めることにより，文相互の関連が把握できることを示す．

CD 表現を用いるもう1つの理由は，行動様式（本研究の場合，正確には行動帰結時の状態）を記述し，検証する際，それを CD 表現要素で記述すれば，言語表現との対応がとりやすく，かつ画像内での対象間の関係の検証が容易になることである．すなわち各 CD 表現要素に対応する状態を検証する画像処理関係を作り，それを組み合せることにより，さまざまな行為の帰結状態を検証できるからである．これはシステムの拡張にとって重要である．システム拡張により，検証すべき行為が増加しても，各行為と，それを表現する言葉の CD 表現の記述を追加すれば良く，既存の処理ルーチンの修正は不要である．しかも表層言語表現を CD 表現に変換するために，後述するように ATN により導出された動詞を中心とする粗い構文則を手がかりとし，次に各語の特性，世界の知識を参照して得た対象の属性，画像から求めた関係を用いて動詞の意味を決めるというように，次第に詳細な事項を調べる方法をとっている．したがって CD 表現への変換が表層レベルの雑事に惑わされることなく，多様な表現と状況に対処できるシステムになっている．

2. 言語処理部

入力されたナレーションを分析し，それを世界の知識および入力画像と対応付けて解釈する方法について

† Interpretation of Act Using both Language and Image Data —Extraction of a Plot and its Question Answering by NORIHIRO ABE (Department of Control Engineering, Faculty of Engineering Science, Osaka University), ITSUYA SOGA (Mitsubishi Co. Ltd.) and SABURO TSUJI (Department of Control Engineering, Faculty of Engineering Science, Osaka University).

†† 大阪大学基礎工学部制御工学科
††† 三菱電機(株)

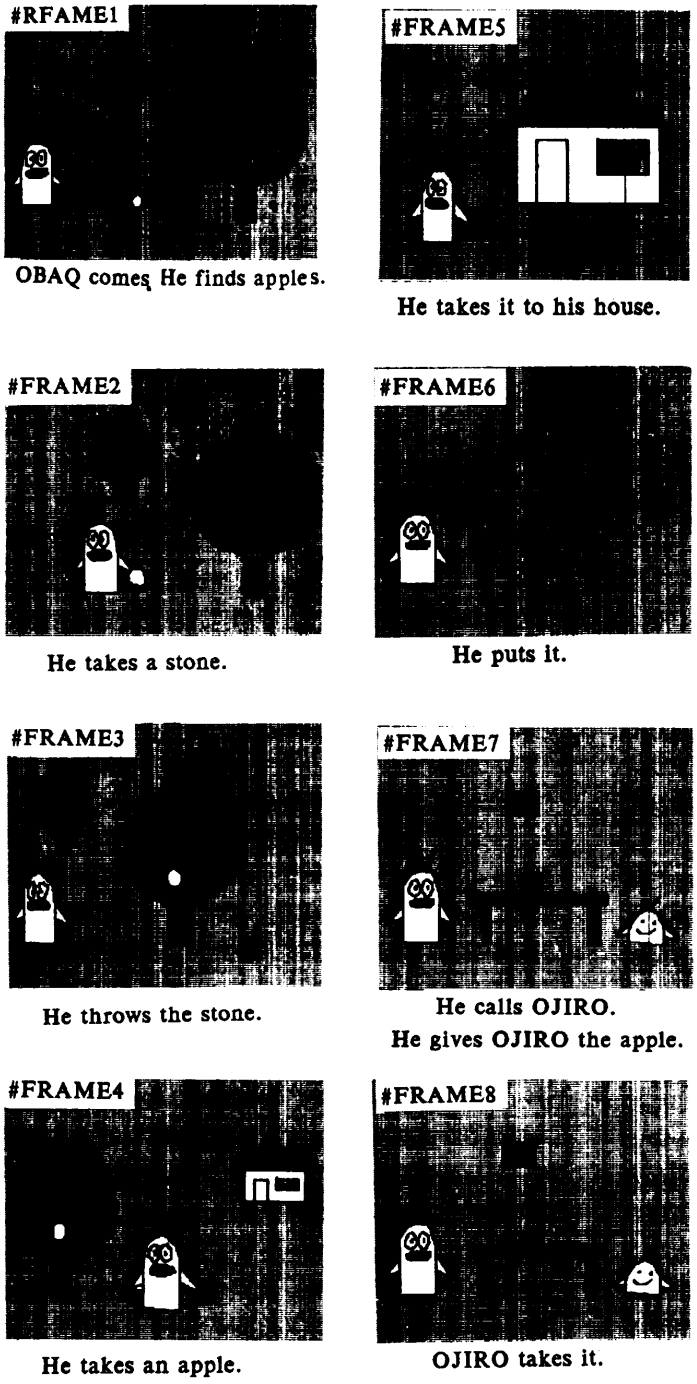


図 1 入力画像とナレーションの例
Fig. 1 An example of input figures and narration.

論じる．**図 1** に本システムの実験対象として用いた入力画像とナレーションを示し，**図 2** に全体的な処理の流れを図示する．

言語処理部は，入力された文を，既報の ATN[2] の一部を使って，名詞句，動詞句などの構文的な基本単位に分割し，必要な統語処理を行う．その結果を**図 3** に示す．詳細は文献 2) に譲る．この結果から，主語，動詞，目的語のような構文上の最高レベルの要素のリスト（以後，中間表現と呼ぶ）を作成する．この時同時に，名詞句を修飾する形容詞をとり出して，その名詞に対応する対象物体の OF (object frame) の属性リストに，それを記入する．そして，文の意味を確定させるために，以下の 2 つを試みる．

（1） 多義解釈を持つ語の分析
（2） 係り受けの決定

これらの解析が自然言語処理に不可欠なことは論をまたぬが，本研究では特に画像データ参照に的を絞って考察しており，他の要素の援用には特に力を注いではいない点をあらかじめ断わっておく．

本研究では，（1），（2）の処理を経た表現（内部表現と呼ぶ）は CD 表現 (Conceptual Dependency Representation) として表わされる．CD 表現を使用する理由は，ⓐ概念レベルの表現を使うことで，質問応答に要求される事物の照合が容易になり，ⓑ因果関係の記述が簡潔になることである．多様な発話のどれとどれが同意な内容を表現しているかという問題は，きわめて難解な問題であるが，CD 表現を使う限り，通常の疑問文に出現する EAT と TAKE の照合などは容易となる．

2.1 CD 辞書

中間表現を概念依存表現に変換するために，個々の動詞に対して，その動詞の意味を CD 表現に変換するためのルールが準備されている．各ルールには，その前提条件と結論となる CD 表現のパターン，およびその時に設定すべきデーモン (demon) やこのルールに関連して行うべき副作用を実現するための手続きが記述されている．

デーモンは，ある行為がなされた結果として生起するであろうと予測される事柄を監視し，発見次第，それに対してしかるべき処置を施したり，行為の因果関係を求めるために，使用されている．**図 5** にルールの一例をわかりやすく書き直して示す．現在 CD 辞書は約 50 語の動詞に対して準備されている．

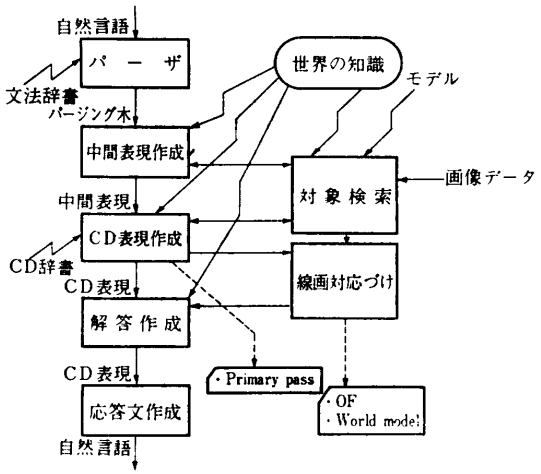
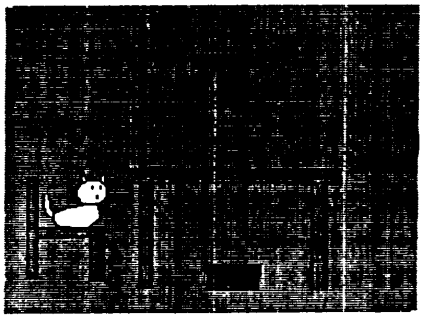


図 2 解析の流れ

Fig. 2 A flow of analysis of the system.

図5を用いてルールを簡単に説明する．TAKE を例にとる．IP 1, IP 2 はこのルールの第1条件部であり，入力文で TAKE が使われ，かつ ATN の生成する中間表現が IP 1 とマッチするなら，次に条件 R1 が調べられる．今 R1 は default ゆえ，無条件で 3, 11 に印された CD 表現が生成される．もし IP 1 に失敗すれば，IP 2 が調べられ，成功すると R2 が調べられる．R2 が失敗なら R3 を調べる．この時 R2, R3 に世界の知識や画像を調べる命令や述語を記すことができる．つまり IPi で構文則を，Ri で文脈条件等をチェックしている．そして，たとえば R2, R3 が失敗すると自動的に R4 が実行され，2の CD 表現が生成され，同時にデーモン D1 が起動されて，この CD 表現に関連する事物間との因果関係が求められる．



A Cat sees the Clock above a box above a desk on the chair.

⇒

```
(CAT1 SEE CLOCK1)  … 中間表現

((*POS) CAT1 ON CHAIR1)
((*POS) CLOCK1 ABOVE BOX1)   } 内部表現
((*POS) CLOCK1 ABOVE DESK1)
```

図 4 画像とナレーションの関係

Fig. 4 The relation between the image and its narration.

2.2 係り受けの処理

中間表現が作成されると，次に前置詞句が，副詞的であるか，形容詞的かを調べ，もし形容詞句として働くとすれば，名詞句内での係り受けはどうなっているのかが，画像情報を用いて推論される．次のような恣意的な例を考えてみよう．

A cat sees the clock *above a box* (PREPG 1) *above a desk* (PREPG 2) *on the chair.* (PREPG 3)

言語レベルの解析だけで，前置詞句 PREPGi (i=1, 2, 3) の係り受けは定まらない．しかし図4のよう

```
HE TAKES IT TO HIS HOUSE.
((**NP** NIL (NOUN (HE (NOUN PRONOUN))))
  (**VP** PRESENT NIL NIL (VERB (TAKE (VERB *TRANS *TRANS2) (MES TAKES S))))
  (**NP** NIL (NOUN (IT (NOUN PRONOUN)))) (TO (PREP))
  (**NP** NIL (ADJ (HIS (ADJ S-ADJ NOUN PRONOUN) (R-TYPE HE)))
  (NOUN (HOUSE (NOUN)))))


WHEN OBAQ CAME , WHAT DID HE FIND ?
((WHEN (CJ CONJ)) (**NP** NIL (NOUN (OBAQ (C-NOUN NOUN))))
  (**VP** PRESENT NIL NIL (VERB (CAME (VERB *INTRANS) (R-TYPE COME)))))
((WHAT (CJ CJN)) (**NP** NIL (NOUN (HE (NOUN PRONOUN))))
  (**VP** PRESENT NIL NIL (SVERB (DID (VERB *DO *TRANS) (R-TYPE DO)))
  (VERB (FIND (VERB *TRANS *TRANS22)))))


WHAT HAPPENED AFTER HE THREW THE STONE ?
((AFTER (CJ)) (**NP** NIL (NOUN (HE (NOUN PRONOUN))))
  (**VP** PRESENT NIL NIL (VERB (THREW (VERB *TRANS) (R-TYPE THROW))))
  (**NP** S (DET (THE (DET))) (NOUN (STONE (NOUN)))))
((WHAT (CJ CJN)) (**VP** PRESENT NIL NIL (VERB (HAPPEN (VERB *TRANS)
  (MES HAPPENED ED))))))
```

図 3 構文解析後の結果

Fig. 3 Results obtained through the syntactic analysis.

な線画が与えられていれば，係り受けは容易に決定される．この例の場合，前論文(1)で述べたように，まず猫，時計，箱，机とイスの5物体が画像中に存在するとプログラムは確信する．その内でも，机とイスは床上にあると考え，画面の下辺で左方からイスを発見する．そして次に on the chair からイスの上に接した物体が残る4物体のいずれであるかを検証する．これは on が接触した上方を意味する事実を用いている．しかし，above は単に上方というだけなので，on ほど束縛は強くないため，on とは異なり，次のような網羅的手法で係り受けを調べている．

係りの決まった前置詞句を除いた結果

NP 0 Prep 1 NP 1 Prep 2 …NP k

という句の並びが残ったとする．その時，文尾から順に，

Prep$i$ NPi が NPi-j $(1 \leqq j \leqq i)$

に係るとした時，その関係が画像中で検証されるか否かを調べてゆき，矛盾のない結果* を得れば，処理を終る．

ところで，以上の操作は，句の係り受けが一意に定まらない場合の処理であり，逆に句の係り受けが明確ならば，それを利用して物体の存在範囲が限定される点に注意しなければならない．A cat on a chair sees a clock. という文からは，猫またはイスの一方が発見されれば，他方はただちに発見される．

ところで，句の係り受けに，対象の位置関係を利用した例として，Winograd のシステム[4]があるが，これは実際の画像を対象としておらず，本研究の方がより実際的側面を有しており，また本システムが係り受けと対象探索との関連を考慮に入れている点が，大きく異なる点である．

2.3 語義の決定

中間表現の CD 表現への変換手法について述べる．例として TAKE, THROW を挙げる．TAKE の語義は文献5) で紹介されているように，非常に多くのものが考え得るが，本研究では，画像上で表現され得る行動に関する語義に的を絞って考察を行っている．さきほど例示に用いた図5に基づいて説明する．

・ OBAQ takes an apples on a tree. を処理して (OBAQ TAKE APPLE 3) なる中間表現を得たとしよう (on a tree の係りは別に取り扱うとする**)．これは TAKE の第1適用条件部：IP 2 にマッチするので，

```
      >TAKE
  IP1   (%OBJ1 VP %OBJ2 TO %PL)
⇒
  R1      default
  →[3  (%OBJ1 =P= PTRANS -O- %OBJ2 -D- (%%% %PL) -I- 11)
    11 (%OBJ1 =P= PTRANS -O- %OBJ1 -D- (%%% %PL))
  IP2   (%OBJ1 VP %OBJ2)
⇒
  R2  %OBJ2 が病気
  →[4  (%OBJ2 =P= DO -R- 12)
    12 (%OBJ1 === ((PHYS*ST 10) (PHYS*ST -3)))
  R3  %OBJ2 が食べられ
        かつ画面から消滅
  →[1  (%OBJ1 =P= INGEST -O- %OBJ2 -D- (MOUTH STOMACH))
  R4    default
  →[2  (%OBJ1 =P= ATRANS -O- %OBJ2 -R- (%%% %OBJ1))
    %OBJ1 が%OBJ2 に為した行為が入力されれば、因果鎖で結合.  } D1

      >THROW
      <OR (%OBJ1 VP %OBJ2 TO %PL)
          (%OBJ1 VP %OBJ2)  >
⇒       default
  →[1  (%OBJ1 =P= SMOVE -O- %OBJ2 -D- (%%% %PL1))
    前フレームと比較して、%OBJ2 の飛んでいった所に存在していた対象OBJに変化があれば、  } D2
      その変化をアサートし、因果鎖で結合.
      かつ以後、OBJに非常な興味を持つ.
```

図 5 CD 辞書の例

Fig. 5 An example of CD-dictionary.

IP 2 に付加されているルール群 R2, R3, R4 が順次，その適用を検証される．R2 に関しては，% OBJ 2 (=APPLE 3) には病気という属性がないので，棄却される．次に R3 の前提条件が調べられる．% OBJ 2 は食べられない物であり，しかも画像データを見ることによって，% OBJ 2 が消滅していることから，食べられたと判断する（なぜなら，画像は行為の完了時点を表わしているから)．そこで，

(OBAQ =P= INGEST -O- APPLE 3 -D- (MOUTH STOMACH))

が内部表現として得られる．また % OBJ 2 が画像中に存在していれば，R3 は棄却されて，R4 を採択する．そして，

(OBAQ =P= ATRANS -O- APPLE 3 -R- (%%% OBAQ))

が得られ，同時に因果鎖結合用のデーモン D1 が生成される．R4 でデーモンを作り，R3 で作らない理由は明らかであろう．すなわち食べられた物は，以後存在しないが，「取った」のなら，何か意図があってのことであろうから，それを監視する必要がある．な

---

* 係りの非交叉関係" などがあるが，本研究では，位置関係成立を中心に考えている．

** (an apple on a tree) の検証は，以前の画面を使って行われる．

お，%%% は，はっきりしない場所を表現するのに用いている．

3. 世界モデルの作成

入力されたナレーションは，画像情報を手がかりとして分析され，結果は図1の Primary path に記入される．さらに行為間の因果関係は，デーモンによって因果鎖が作られている．

一方，個々の画面は，文献1) で述べた方法によって，モデルとの照合がとられ，各画面内での位置が抽出されている．しかし，画面相互の関連性を把握しなければ，異なる画面に登場した家や木が，同一の物体か否かは明らかにならない．"the tree" という表現により，この tree が既出の木であることを言語処理部が知ったとしても，複数個の木があれば，そのどれを指すかは即断できない．

本研究の対象画像は，繰り返して言うが，連続的な情景を描写したものでなく，別々の画面に同一の対象が登場しても，厳密な意味での対象図形の照合性や，位置関係の一致は期待できない．しかし，直前の画面に出現していた物体に類似した構造的性質を持つ物体が，次の画面にも登場し，かつ映画館から音楽ホールへというように，場合の転換が生じていないとすれば，それらの2つの物体を同一視するのは自然である（そうではないなら，何らかの方法*が明示的にそれを告げるはずである）．

3.1 線画間の対応付け

前述したように，画像データは各画面ごとに独立な座標系で記述されており，画面相互の位置的関連を調べ，個々の画面（n 枚目の画面の座標系を CO_n とする）が，統合的座標系 COA でどのような位置を占めるかを知る必要がある．

また，同一属性を持つ物体の標準的な大きさを知識として，対象のモデルに与えることにより，多少の奥行き方向の広がりを表現するようにしている．画面の対応付けは，以下のようにして行う．

現在，n 枚目の画像を処理しようとしているとする．まず，CO_m と CO_n $(n>m)$ の両者に記述されている対象**で，動かない対象を捜し出す．

（1） 発見できれば，CO_m, CO_n それぞれにおける対象の座標から，COA 上での CO_n の原点の座標を求める．

（2） 発見できないなら，CO_n に記述されている動かない物体を，m 番目の画面で捜し（m 番目の画面で発話指示されず，かつ発話指示された物体と関連のない対象は，m 番目の画面では無視されていることに注意），発見できれば（1）へ戻る．そうでなければ，CO_l $(l>n)$ に対する位置付けにおいて，COA 上での位置が判明するまで，処理をサスペンドする．

（3） 場面の展開（たとえば，「通りから家に入いる」というような場合），および，前後との関連がとれずに，したがって，それまでの統合的座標系 COA との関係が算出不能な時，新たな統合座標系 COA 1 を作る．なお，座標としての関係は算出不能でも，「右へ走った」といったナレーションがあれば，右方への移動が，COA 1 に記される．

最後に，統合座標系が複数個ある時，各系に記述されている対象で，共通な物が存在しているか否かを調べ，もしあれば，統合座標系の再統合を行う．

本来1つの座標系で表現可能なものが，上の処理をしても表現不能な場合がある．それは，2つの統合座標系 COA 1 と COA 2 の両者に共通な対象が，発話指示されぬことにより無視され，かつ他には統合に役立つ手がかりのない時である．これが，本手法の限界でもある．

4. 質疑応答

以上に述べた処理によって得た内部表現，OF，世界のモデルなどの情報，および未抽出のデータを残した画像データを利用して，線画に記された話の流れに関する質問応答を行う．

4.1 Query の作成

入力疑問文を平叙文に直し，未知部分に変数を挿入することにより，ナレーションの処理とほぼ同一の手順で，query が作成される．

4.2 内部表現の照合

ここで用いるマッチャは，次の点で通常のパターンマッチャとは異なる．

（1） 質問の指す時点を知る必要がある．

（2） 位置関係の記述は，Primary path には格納されていない場合が多いので，座標または原画に戻って計算する能力が必要．

（3） 状態の記述は，たとえ1度しか言明されていなくても，その発生時点から，状態を変化せしむる行為のなされるまでの期間，その状況が保存されている点を考慮する必要がある．たとえば，「家の内」では

---

* 場面の転換を示す表現や，"another tree" のような明示的な修飾等による発話的指示がこれに当たる．

** 一度でも発話指示された対象は探索されることに注意．

```
ATRANS
 [
      *PTN   (=S #SUBJ =P= ATRANS -O- #OBJ1 -D- (#OBJ2 #OBJ3))        ┐ 基本パタン
      *COND  [COND ((EQ #SUBJ #OBJ3)                                  │
                    (TRULE2 (#SUBJ VP1 #OBJ1)))                       │
                   ((EQ #SUBJ #OBJ2)                                  ├ 条件判断部
                    (TRULE2 (#SUBJ VP2 #OBJ3 #OBJ1)))                 │
                   (T                                                 │
                    (TRULE2 (#SUBJ VP3 #OBJ2 GIVE #OBJ3 #OBJ1)))]     ┘
      VP1    (GET TAKE)                                               ┐
      VP2    (GIVE)                                                   ├ 動詞リスト
      VP3    (LET MAKE)                                               ┘
                 )

MTRANS
 [
      *PTN   (=S #SUBJ =P= MTRANS -O- #OBJ1 -R- (#OBJ2 #OBJ3))
      *COND  [COND ((EQ #SUBJ #OBJ3)
                    (TRULE2 (#SUBJ VP1 #OBJ1)))
                   ((EQ #SUBJ #OBJ2)
                    (TRULE2 (#SUBJ VP2 #OBJ3 #OBJ1)))]
      VP1    (FIND SEE)
      VP2    (TELL TEACH)
```

図 6 CD 表現から遷移表現を得るための規則

Fig. 6 Rules to get an intermediate description from the CD-notation

「木」や「石」がどうなっているかは当然書かれていないが，これらは以前の状態を保持していると仮定して，時間をさかのぼって探索する．

4.3 疑問文の種類

（1） Why：因果鎖をたどって得られる表現を答えとする．ここで問題となるのは，因果鎖をどこまでたどるかという点である．1レベルしかたどらないと，質問者は何度も why を繰り返さざるを得なくなる．また，可能な限り因果鎖をたどると，答えが冗長になる．もちろん，これはデーモンの因果鎖の形成と密接に関連していることだが，現段階では，話の筋が簡単なため，たどれるところまでをすべて答えとしている．

（2） How：記述のI-リンク (Instrument：行為・道具格に相当する記述) を答える．I-リンクは，たとえば，ATRANS を含む表現には MOVE を含む表現を I-リンクするというような Schank 等の考え方と，事件の発生原因を探索してリンクする考え方により生成される．

（3） Yes/No：探索できれば Yes，さもなくば No.

（4） Occurrence：指定された時点で言明されている表現を Primary path より抽出して答える．

（5） Component：what, where, when などで始まる文の疑問詞に変数を代入して探索する．

（6） その他：How many…, How long…, などは，それぞれに対応する関数を作って，表現を探索する．

4.4 内部表現の遷移表現化

得た内部表現は，入力処理とは逆のステップを経て，遷移表現に変換する．そのための変換規則を**図 6** に示す．図は ATRANS と MTRANS を含む CD 表現を遷移表現化する規則の例である．＊PTN が CD 表現にマッチさせられて，＃SUBJ, ＃OBJ 1 などを求める．その値を条件部 ＊COND で調べ，それにより TRULE 2 の後の表現を選び出し，その中の Vi を動詞リストの値で置換して，遷移表現を作る．より具体的な例を ATRANS で説明する．まず，基本パターン中の変数（＃あるいは VP で始まる）と内部表現中の対象の対応をとる．次に動詞リストを参照しつつ，条件部を評価する．＃SUBJ に対応する対象が ＃OBJ3 に対応する物に等しいならば，“＃SUBJ get ＃OBJ 1” という文に相等し，遷移表現 (＃SUBJ (GET TAKE) ＃OBJ 1) を得る．たとえば図 7 の 7）の表現 (OBAQ =P= ATRANS -O- STONE 1 -D- (%%% OBAQ) =R= (12)) の場合は，＃SUBJ が OBAQ, ＃OBJ 1 が STONE 1, ＃OBJ 2 が %%%, ＃OBJ 3 が OBAQ であり ＃SUBJ=＃OBJ 3 故 (OBAQ (GET TAKE) STONE 1) が作られる．もし，＃SUBJ と ＃OBJ 2 の指す対象が同じであれば，“＃SUBJ give ＃OBJ 3 ＃OBJ 1” という文に相当し，(＃SUBJ (GIVE) ＃OBJ 3 ＃OBJ 1) を得る．たとえば，図 7 の 31) の表現 (OBAQ =P= ATRANSE -O- APPLE 3 -D- (OBAQ OJIRO) =E= (12 18)) に対しては ＃SUBJ が OBAQ, ＃OBJ 1 が APPLE, ＃OBJ 2 が OBAQ, ＃OBJ 3 が OJIRO なので，(OBAQ (GIVE) OJIRO APPLE 3) が生成される．

4.5 応答文の作成

遷移表現に対して，動詞の選択，時刻，態の決定，対象に対する名詞句の作成を行えばよい．この部分には，すでに作成した自然語文生成プログラム[6] を用いた．

ここで問題となるのは，名詞句を生成する場合，たとえば **4.4** の STONE 1 や APPLE 3 などにどの程度

```
** REPRESENTATIONS **
  1) (OBAQ =P= PTRANS -O- OBAQ -D- (%PL %HERE))
  2) (OBAQ =P= MTRANS -O- APPLE51 -R- (%%% OBAQ))
  3) ((*POS) APPLE1 ON TREE1 +.723529E+00 +.264331E+00)
  4) ((*POS) APPLE2 ON TREE1 +.258823E+00 +.238853E+00)
  5) ((*POS) APPLE3 ON TREE1 +.435294E+00 +.563694E+00)
  6) ((*POS) APPLE51 ON TREE1)
  7) (OBAQ =P= ATRANS -O- STONE1 -D- (%%% OBAQ) =R= (12))
 12) (OBAQ =P= SMOVE -O- STONE1 -D- (%%% %PL1) =E= (7) =R1= 16 =R= (18 23 26 31))
 13) ((*POS) APPLE1 ON TREE1 +.715189E+00 +.271242E+00)
 14) ((*POS) APPLE2 ON TREE1 +.265823E+00 +.245098E+00)
 15) ((*POS) APPLE51 ON TREE1)
 16) (OBAQ =P= PROPEL -O- APPLE3 -D- ((+.177382E+03 +.137341E+03 +.977839E+00)
(+.196673E+03 +.173271E+03 +.820605E+00)) -I- 12)
 17) ((*POS) APPLE3 CONTACT OBAQ +.109574E+01 +.743750E+00)
 18) (OBAQ =P= ATRANS -O- APPLE3 -D- (%%% OBAQ) =E= (12) =R= (23 26 31))
 19) ((*POS) APPLE1 ON TREE1 +.743055E+00 +.265100E+00)
 20) ((*POS) APPLE2 ON TREE1 +.312500E+00 +.244966E+00)
 21) ((*POS) APPLE51 ON TREE1)
 22) ((*PSS) HOUSE1 OBAQ)
 23) (OBAQ =P= PTRANS -O- APPLE3 -D- (%%% HOUSE1) -I- 24 =E= (12 18))
 24) (OBAQ =P= PTRANS -O- OBAQ -D- (%%% HOUSE1))
 25) (OBAQ =P= PTRANS -O- OBAQ -D- (%%% (INSIDE HOUSE1)))
 26) (OBAQ =P= PTRANS -O- APPLE3 -D- (%%% %PL) =E= (12 18))
 27) ((*POS) APPLE3 ON TABLE1 +.456522E+00 -.215909E+00)
 28) (OJIRO =P= PTRANS -O- OJIRO -D- (%%% OBAQ) -I- 29)
 29) (OBAQ =P= PROPEL -O- VOICE -D- (OBAQ OJIRO))
 30) ((*POS) APPLE3 ON TABLE1)
 31) (OBAQ =P= ATRANS -O- APPLE3 -D- (OBAQ OJIRO) =E= (12 18))
 32) (OJIRO =P= INGEST -O- APPLE3 -D- (MOUTH STOMACH))
```

図 7 プライマリ・パスの一部

Fig. 7 A Portion of the primary pass.

```
#FR1
 > OBAQ        == (*ACT (1 2) *PLIST ((BODY . 9) (HAND1 . 11) (HAND2 . 10)
  (HAIR . BR7) (((H3 H2 H1) 32 31 30)) (EYE1 . 12) (BEYE1 . 13) (EYE2 . 14)
  (BEYE2 . 15) (MOUTH . 16) (MTH . BR8) (((MTH1) 34))) *WHERE (20 128)
  *SIZE (41 74) *3D (+.205142E+02 +.131291E+03 +.102571E+01))
 > APPLE51     == (*ACT (2) *MEM (APPLE1 APPLE2 APPLE3) *POS (6))
 > APPLE1      == (*GNAM APPLE51 *POS (3) *PLIST ((APP . 3) (HT1 . BR2) (((A1)
  8)) (HT2 . BR1) (((A2) 7))) *WHERE (207 92) *SIZE (13 21)
  *3D (+.203967E+03 +.906520E+02 +.985348E+00))
 > APPLE2      == (*GNAM APPLE51 *POS (4) *PLIST ((APP . 4) (HT1 . BR4) (((A1)
  10)) (HT2 . BR3) (((A2) 9))) *WHERE (168 87) *SIZE (12 23)
  *3D (+.164348E+03 +.851086E+02 +.978261E+00))
 > APPLE3      == (*GNAM APPLE51 *POS (5) *PLIST ((APP . 5) (HT1 . BR6) (((A1)
  12)) (HT2 . BR5) (((A2) 11))) *WHERE (183 130) *SIZE (12 23)
  *3D (+.179022E+03 +.135000E+03 +.978261E+00))
 > TREE1       == (*POS (3 4 5 6) *PLIST ((LEAF . 2) (TRUNK . 1))
  *WHERE (152 61) *SIZE (85 157) *3D (+.134009E+03 +.538121E+02 +.882166E+00))
```

図 8 #FRAME 1 に出現する対象の OF

Fig. 8 The sample object frames for objects appearing in FRAME-1.

修飾句（節）を付加するか，それとも代名詞としておくかという問題である．ある対象物に付属する属性がいくつあっても，人間がその対象物を指していると判断するのに，それほど多くの限定詞を加える必要はないと思われる．逆に多くの修飾限定は，煩わしさを伴う場合が多い．特に，共通の脈絡があればそうである．現時点では，扱う話が単純なため，原則として修飾はしていない．

動詞の選択は，疑問文に使用された動詞が，候補動詞リスト中にあれば，それを用いることにした．

5. 実 行 例

5.1 入 力 処 理

図 1 に示した入力画像およびナレーションの例に対して，**図 7** に Primary pass に記述された表現を，**図 8** に対象フレーム OF の一部を，また**図 9** には，話の前半部を表わす画面の位置データと，位置的世界モデルを図示したものを示す．以下，処理過程を概説する．

#FRAME 1 では，ナレーションに登場するオバ Q とリンゴ（複数）を探索する．この時，発話指示されていない石は，同定されていない．しかしながら，木はリンゴとの関連から先に捜され，次にリンゴが発見されている．また図 7 の 3)～5) の表現の後部の数は APPLE 51（発話指示の apples に対応）というグループに属している個々のリンゴの位置を表わしている．

#FRAME 2 では，“take” に対する変換規則が用いられた結果，表現 7) を得ている．そして同時に **2.3** で解説したデーモン D1 が生成され，D1 によって後の画面から，表現 7) に =R= (12) が加えられる．これは表現 12) は，7) の記述があったため得られたこ

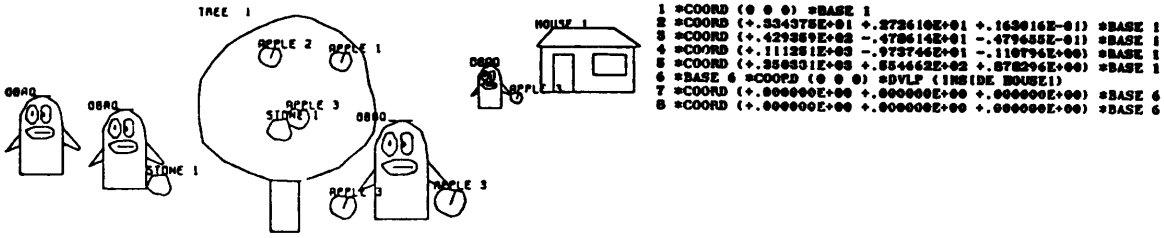


図 9 物語の前半部に対する位置的世界のモデル

Fig. 9 The positional world model in the first half of the story.

```
〈1〉 ::: WHEN OBAQ TOOK AN APPLE, HOW MANY APPLES WERE THERE ON THE TREE?
     Two.
〈2〉 ::: HOW DID HE TAKE THE APPLE?
     He threw the stone to let the apple fall.
〈3〉 ::: AT THAT TIME, WHERE WAS HE?
     He was near a tree.
〈4〉 ::: WHY DID HE TAKE THE APPLE?
     To take it to his house and to put it and to give it to Ojiro.
〈5〉 ::: WHERE DID HE PUT THE APPLE?
     He put it on a table.
〈6〉 ::: DID OBAQ GIVE OJIRO THE APPLE?
     Yes, he did.
〈7〉 :::DID OJIRO EAT THE APPLE?
     Yes, he did.
〈8〉 ::: WHEN DID HE GET A STONE?
     At #frame2.
〈9〉 ::: WHAT COLOR IS IT?
     It is white.
〈10〉 :::WAS THERE A CLOCK?
     Yes, there was a clock in the house.
〈11〉 ::: WHAT HAPPENED AT #FLAME6?
     Obaq entered the house.
     Obaq put the apple.
〈12〉 ::: HOW MANY APPLES WERE THERE ON THE TABLE?
     One.
〈13〉 ::: WHAT HAPPENED AFTER HE THREW THE STONE?
     Obaq let the apple  down.
     Obaq got the apple.
     Obaq took the apple to the house.
     Obaq put the apple.
     Obaq gave Ojiro the apple.
```

図 10 質問応答の例

Fig. 10 Some results on question answering.

とを意味している．

#FRAME 3 では，“throw” の変換規則が適用され，表現 12) とデーモン D2 が作られる．D2 は，#FRAME 2 から #FRAME 3 の間に APPLE 3 が落下したことを察知して，ナレーションで入力されていない「石がリンゴを落とした」という表現 16) を推測かつ検証し，かつ以後の画面で，APPLE 3 に対して，主人公がどんな行為を及ぼすかに注目させるデーモンを生成する．

#FRAME 5 では，直前の #FRAME 4 との間に共通な動かない対象が見つからない (#FRAME 4 では，家は発話指示に無関係なので，認識されていない点に注意) ので，#FRAME 5 で初めて認識される HOUSE 1 を #FRAME 4 に戻って探索している．そしてこの操作によって #FRAME 5 が統合座標 COA と関連付けられる．

#FRAME 6 では，「家に入る」ことによって，場面の転換が生じており，統合座標系を切り換えている．これは図 9 に記されている．以後の処理過程も，ほぼ同様の手順で行われるので，詳細は略す．

5.2 質問応答

図 10 に質問応答例を示し，その一部に付いてのみ，解答抽出の過程を簡単に解説する．例の引用は，文頭の番号で行うものとする．

〈1〉: “How many…” で始まる疑問文に対しては，query にマッチする対象の数を答える．例では “when OBAQ took the apple” から質問は #FRAME 4 であると判断し，query ((*POS) ?-APPLE ON TREE 1) にマッチする表現を #FRAME 4 で探索し，表現 19), 20) を得て two と答える．

〈2〉: 表現 18) の =E= (その行為を可能ならしめた行為を表わす関係) を見て，表現 12) が答えであることを知る．

〈3〉: OBAQ の位置は，Primary pass にはない．OBAQ の OF の #FRAME 4 の記述と，そこで記述されている対象で OBAQ の近くにあり，かつある

程度の大きさ以上の物を引用して答える．

〈4〉：表現 18) の因果鎖 =R= を見て，そこに記されている表現 23), 26), 31) を答えている．**4.3** の(1)で述べたように，冗長な解になっている．今後検討を要する点である．

〈5〉："table" 上に置いた点は述べられなかったが，"put" という語と，リンゴの位置から，支持台が推測・検証されているので，答えられる．

〈7〉：CO 表現を用いているので，eat と，take との照合は簡単である．

〈9〉：石の色は石の OF より求めている．

〈10〉：時計は全く話題になっていないので，全く認識されておらず，原画のまま保存されている．「CLOCK は通常家の内にある」という知識を用いて，「家の内」に対応する状況 #FRAME 6 を取り出し，その原画から CLOCK を探索することにより，時計が発見される．このように最初は注目されず認識されていない物体も，その存在を問うというように，その物体の存在を示唆する表現を利用することによって，適宜発見可能である．ただし，現在，#FRAME 7 以降で時計がどこにあるかは調べていない．したがって **3.2** で記したように，時計が座標系統合に決定的役割を果す場合でも，依然として座標整理は行われぬままになっており，改善が必要である．

6. 検　討

視覚情報と言語情報を相互参照することによって，以下のような効果を得た．

1）発話指示された対象を中心として注目物体を限定し，かつその存在範囲を限定した探索が可能となった．

2）対象の位置関係を座標，原画として残すことにより，記述がコンパクトになった．

3）言語解析のみでは，処理が定まらない前置詞の係り受け，多義語の解釈が，画像を利用することによって容易となった．

しかし，改善すべき点も少なくない．まず対象の同定に関してだが，人間の認識によく合った振舞をする反面，形状的特徴を重視していない点，モデルの回転を考慮していない点が問題である．「手を曲げる」というような行為は前者の理由により，「寝る」「飛び込む」などは後者の理由により，同定できない．前者は，より正確な形状情報をモデルに与え，可変テンプレートマッチ[7]を行うことにより，また後者は対象とモデル照合時の外接長方形の設定プログラムを改善して，長方形の傾きから回転角を求めて，対象またはモデルを回転角に応じて変換する必要がある．

本研究で対象とした物語は簡単なものであったので，これをより長い，複雑な筋に適用して，不足している点を補う必要がある．文献 1) に記したように，アスペクトと行為の時系列の関係をモデル化することも挙げられよう．また，現在は筋が簡単なため，本研究ではいわゆる script のような典型的状況に関する知識は用いていないが，より充実した推論を行わせるためには，こうした従来の成果[8]を組み込む必要もあろう．

また，モデルを完全に 3 次元化して，体のねじれなどの回転[7]を考慮した行為の推測または同定も，問題ははるかに困難になるが，行う価値があろう．

参 考 文 献

1) 安部，曽我，辻：言語・画像を利用した行動の解釈(1)—発話指示による対象の同定—，情報処理学会論文誌，Vol. 23, No. 2. pp. 124-132 (1982).
2) 滝，安部，辻：知識ベースを利用した構文解析と拡張されたプロダクション・システムを持つ自然言語理解システム，信学研資 AL 79-119 (1980).
3) 長尾，辻井，田中：意味および文脈情報を用いた日本語文の解析—名詞句・単文の処理，情報処理，Vol. 17, No. 1, pp. 10-18 (1976).
4) Winograd, T.: Understanding Natural Language, Academic Press (1972).
5) Winston, P. H.: Artificial Intelligence, Addison-Wesley Publishing Company (1977).
6) 曽我：ルールに基づく推論システムのための言語処理プログラム，大阪大学特別研究 (1979).
7) Tsuji, S., Osada, M. and Yachida, M.: Three Dimensional Movement Analysis of Dynamic Images, IJCAI 6, pp. 896-901 (1979).
8) 小西：スクリプトを用いた談話理解，大阪大学修士論文 (1978).
9) 岡田，田町：図形の意味解釈とその自然語記述—意味分析—，信学論，Vol. J 59-D, pp. 331-338 (1976).
10) 雨宮，島津：日本語の意味解釈過程—図形世界を対象例として，信学論，Vol. J 60-D, pp. 633-640 (1977).
11) Tsuji, S., Kuroda, S. and Morizono, A.: Understanding of Simple Cartoon Film, IJCAI 5, pp. 609-610 (1977).
12) Schank, R.: Conceptual Information Processing, North-Holland publishing Company.
13) Waltz, D., Boggess, L.: Visual Analog Representations for Natural Languages, IJCAI 6, pp. 926-934 (1979).
14) 曽我，安部，辻：言語・画像間相互参照によるプロットの把握，信学研資，AL 80-85 (1981).

(昭和 56 年 5 月 15 日受付)
(昭和 56 年 9 月 7 日採録)