

擬似日本語によるデータベース検索システム「ヤチマタ」

諸橋正幸・藤崎哲之助・鷹尾洋一・間下浩之・渋谷政昭

(日本アイ・ビー・エム(株) 東京サイエンティフィック・センター)

0. まえがき

ヤチマタは日常語に近い表現でデータベースへの照会を可能にした実験システムで、普段電子計算機と馴染みのない人々にその使用を可能とさせる試みの一つである。

データベースに対する検索言語として日常のことはに近い言語を用いるという実験は数多くみられるが、それと比べるとデータベースの世界が限られており一般性に欠けていた。ヤチマタの大きな特徴の一つは特定のデータベースの世界を想定して開発されたシステムではないので、どのような業務にも比較的簡単な手順で適用できる——すなわち汎用性を持つ点である。ヤチマタの汎用性を実現するためには、オーに適用業務のデータベース構築を日本語の表現法を反映しやすい形にするようなデータ・モデルを導入すること、 α としてデータベース検索に必要な日本語の構文表現を網羅するために自然語の特長である表現の多様性と省略可能性を活かす文法およびその解析手法を得ること、 β 二点が大きな関心事となる。ここではこの二点についてその実現手段を紹介し、最後に実験的に行ったいくつかの適用例を挙げてこれらの業務におけるデータ・モデルの例と設計の際の問題点に触れることにする。

1. 名詞句データ・モデル

ヤチマタを扱うデータ・モデルを我々は名詞句データ・モデルと名付けた。

④ LUNAR (W. A. Woods),
SHRDLU (T. Winograd),
REL (F. B. Thompson),
REQUEST (Plath & Petrik) 等。

このモデルは関係形式⁽¹⁾に基づいたものである。

関係形式データ・モデルではデータベースの設計者が与える基本定義域とそれらの向の関係を表の形であらわす。例えば 倉庫名とそこに収められている部品名とその部品の数量との向の関係を示すと α 1.1図のようになる。この表において各列はすべて対等である。従ってこの表から部品と倉庫を規定して部品の数量を求めるときもできるし、部品とその数量を与えて該当する倉庫を求めることも可能である。

倉庫	部品	数量
1号倉庫	ラジエター	50
1号倉庫	バンパー	120
2号倉庫	ラジエター	10
3号倉庫	カーラジオ	50

α 1.1図(倉庫, 部品, 数量)の関係

この関係形式モデルとそのまま採用したのでは検索文と表との結びつきや各列との対応がとりこにいたのに特別の考慮が払われて2つのタイプの特殊な関係形式——名詞表と動詞表を作り出した。

1.1 名詞表

名詞表は名詞または名詞句と対応させた関係形式の一種で、表に付けられた名前NNは表の名前であると同時に α 1列(これを対象属性と呼ぶ)に現われる要素の集合をあらわす。 α 2列以降(1列しかない名詞表も存在する。その使い方は後述する)を限定属性とよび、この名詞NNに対してどのような修飾が可能であるかを示して

いる。(オ1.2図)。名詞表が関係形式モデルと異なるもう一つの特徴は限定属性に対し格標識と名付けた部分が付加された点である。この役割はUUを修飾する部分にあらわれる名詞が限定属性中のある列に該当するとき、その後につくことば(これを後置詞とよぶ。主に格助詞かくるであろう)がなんであるかを示すことである。なお(格標識, 定義域)の組合せは表中においてユニークでなければならぬ。

NN	C ₂	o o o	C _n
D ₁	D ₂	o o o	D _m
e ₁₁	e ₁₂	o o o	e _{1n}
⋮	o		
o	o		
e _{m1}	e _{m2}	o o o	e _{mn}

NN : 名詞
C₂...C_n : 格標識
D₁...D_m : 定義域名
e_{ij} : 属性値

対象属性 限定属性

オ1.2図 名詞表

在庫量	C5	C8
数量	部品	倉庫
50	ラジオター	1号倉庫
120	バンパー	1号倉庫
10	ラジオター	2号倉庫
50	カーラジオ	3号倉庫

C5 = {の}
C8 = {の, におお, での}

1号倉庫
倉庫
1号倉庫

オ1.3図 名詞表の例

オ1.3図には名詞表の例を示す。左側の表は左端の定義域である数量50, 120, ...の集合に対する呼称である「在庫量」を意味する。この「在庫量」には二つの修飾句, 「(部品)の」と「(倉庫)の」が付き得る。後者については「(倉庫)における」あるいは「(

倉庫)での」という言い方も許される。この表のオ一行に着目すれば, 「1号倉庫におけるラジオターの在庫量は50である」という事実を示していることになる。これに対して右側の表は修飾句を持たない名詞表でこの表のようにただ一つの事実しかあらわさないようなものは質問文中にあらわれる定義あるいは固有名詞を扱う際に用いられる。

1.2 動詞表

前述の名詞表と比べると動詞表はより関係形式に近い形を持ち, 各列は並べ換えに因りまったく対称である。この表は用語(動詞, 形容詞, 形容動詞)の意味をあらわしていて, オ1.4図の構造を持つ。

VV	C ₁	C ₂	o o o	C _n
D ₁	D ₂	o o o	D _m	
e ₁₁	e ₁₂	o o o	e _{1n}	
o				
o				
e _{m1}	e _{m2}		e _{mn}	

VV: 動詞
C₁...C_n: 格標識
D₁...D_m: 定義域名

限定属性

オ1.4図 動詞表

ヤチマタのデータ・モデルにおいては用言の使い方に制限を与えており, この動詞表は連体修飾句としてのみ現われる。すなわち基本的には<動詞(連体形)><名詞句>なる表現だけが許されている。しかもこの並びに対して<名詞句>の定義域と等しい定義域を持つ限定属性を動詞表から取り出し, それを対象属性に変えた名詞表を作る。このように, 動詞表が実際に意味解析をするにあたり常に名詞表の形で扱われるため, 動詞表はヤチマタにおいては名詞表の補助的役割を果たしている。名詞句データ・モデルの名詞の由来もここからきたものである。オ1.5図に動詞表の例と それが名詞表に変換

される様子を示す。

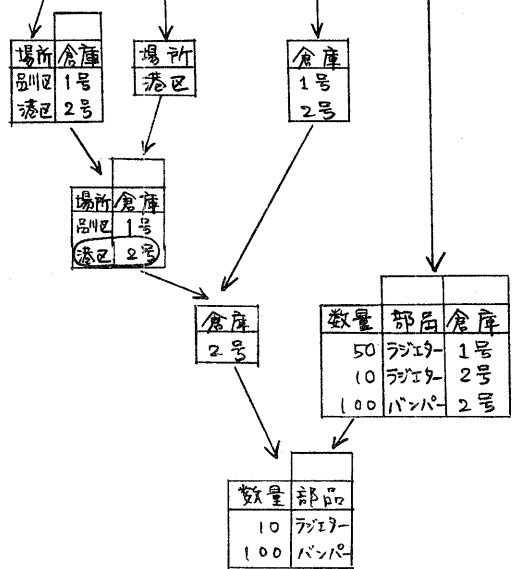
「納品する」

C3	C7	C10	
業者	倉庫	部品	C3={成の}
青葉商会	1号倉庫	ラジエター	C7={に}
。	。	。	C10={を}
。	。	。	
。	。	。	

「納品する(倉庫)」

	C3	C10
名庫	業者	部品
1号倉庫	青葉商会	ラジエター
。	。	。
。	。	。
。	。	。

所在地が港区である倉庫の在庫量は



オ1.5図 動詞表とそれが名詞表に変換された例

従ってマチマチでは表とそれらの向の演算はすべて名詞表の形で取扱われる。

1.3 検索文の評価

一つの名詞表あるいは動詞表に同じで、その構造と意味するところは上記に述べた通りであるが、実際の検索文においてそれがどのように扱われるか、また複数個の表にわたる検索について表向の演算が何如に行なわれるかにつき概観する。

マチマチは名詞句の繰返しによる文型を基にして照会文を解析し、その解析した構造に従って表の向の演算を行なっていく。最終的に<名詞句><は>>の文型が確認されるとそのときまですてに出来あがった表を答として使用者に表示する。この向の経緯を模式的にあらわすとオ1.6図のようになる。図中に何種類かの演算があらわれているが、これらをまとめ、対応する構文と共に列挙したのがオ1.7図である。このうち正順制約を例にとり演算の操作を説明

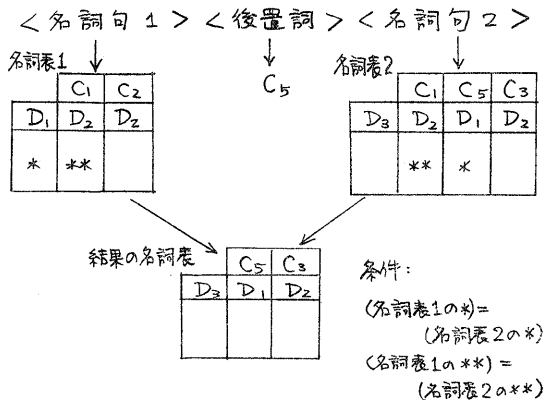
オ1.6図 照会文とその評価

基本検索演算	対応する構文
正順制約	<名詞句><後置詞><名詞句>
逆順制約	<名詞句>が<名詞句>である<名詞句>
論理積	<名詞句>であり<名詞句>であり(3)
論理和	<名詞句>または<名詞句>
二項算術演算 (+, -, *, /)	<名詞句><演算子><名詞句>
集合関数 (平均合計)	<名詞句>の<関数名>

オ1.7図 代表的検索演算と対応する構文

する。図にある通り正順制約は<名詞句1><後置詞><名詞句2>の句構造に対応しており<名詞句1>の対象属性と同じ定義域でかつ<後置詞>を含む格標識である限定属性を<名詞句2>から求め、<名詞句1>の対象属性と<名詞句2>の該当する限定属性との向の等号条件により二つの名詞表を結合する。更に統合可能な限定属性(両方の表において定義域、格標識と同一一致する限定属性)があればそれらの向にも等号条件を与える。ただ

し制約の対象となつた列は結果の表に残るが、限定属性同氏が統合の対象となる場合には結果の名詞表にその列はあらわれない。結果の名詞表において制約の対象となつた限定属性は再び正順制約の対象とならばいように細工がなされる。これは「バンパーのバンパーの在庫量は？」という類いの文を許さないためである。



※1.8図 正順制約における演算操作

なお、その他の演算に関しては参考文献(2)に詳しく紹介されている。

1.4 照会結果の保持と参照

ヤチマタは対話型を照会・回答を行なうので、以前の照会結果を参照しながら次の照会を行なうという場面が多々見られる。こうした状況を考えて、ヤチマタは結果の保持と参照の機能を備えている。保持される結果(これはまた名詞表である)は、対象属性の定義域毎に最新のものだけである。これらを参照するには<その><名詞表>等の限定指示による表現が用いられる。また単に<それ>等の指示代名詞のみで参照することも許す。この場合には直前の照会結果を参照する。結果参照の例を以下に示す。

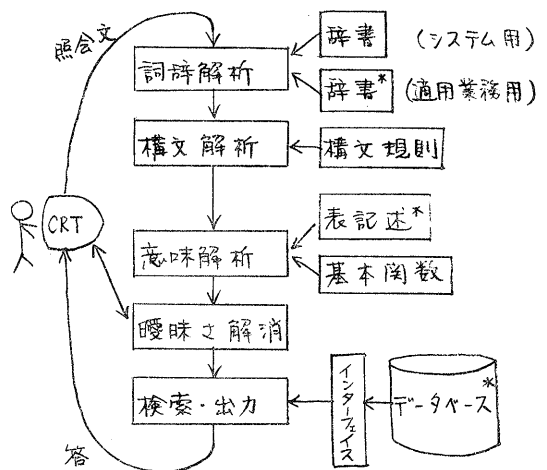
「ラジエターのある倉庫は？」

「その倉庫の所在地は？」

2. 照会文の解析

構文解析で使われる方法は、REL⁽³⁾

で用いられたKayアルゴリズムで、上昇型インタックス・ディレクティブ型である。構文規則は文脈自由文法と同じ記法によって書けるが左辺にも複数個の項を許す。この手法を使うことにより文法上合法的な複数個の解釈を全て得られるために、検索文に曖昧さがある場合に威力を発揮する。しかし逆に解析に失敗した場合にその原因を探るのが難しい。ここでは質問文の解析の手法を述べるとともに曖昧さを解消するための手続き、解析に失敗した場合にその原因を究明する手続き等についても触れる。



*通用業務に利用可能な部分

※2.1図 ヤチマタの照会・回答手続き

2.1 照会・回答の処理手続き

照会文が使用者から与えられ、その回答が返されるまでの処理手続きを※2.1図に示す。照会文の解析は三段階ある。詞辞解析は単語を認識するための処理であり、主に辞書との照会を行なう。構文解析は照会文の構造を構文規則を使って知り、それを樹構造の形で表現する。意味解析は構文解析で得られた文の構造から導びかれる名詞表、動詞表の演算手順を示し、構文解析の時点においては、やはり不構造で示されている(これは構文解析の本とはま

なく別のものであり、以後意味の本と呼ぶ)。この意味の本を評価することによりデータベース検索のためのコマンドが生成される。曖昧さの処理は、この時点までに二つ以上の解釈が成された場合——すなわち、一つの照会文に対して二つ以上の検索コマンド(形式的言語で書かれたもの)が生成された場合に行なわれる。それぞれの解釈の仕方をCRTに出カして使用者に選択を促すことにより解釈を一つに絞る。最後の検索・出力の部分は、形式的言語に変換された照会文から結果を得る部分である。ヤチマタではデータベースとしてPRTV(Peterlee Relational Test Vehicle)⁽⁴⁾を用いている。これは関係形式データベースを扱い、テーク言語は代数型言語である。PRTVは物理データに対応する関係形式の他にサブスキーマとして仮想関係形式(他の関係形式を用いて定義されたものである)の実際の表としては作られないものを持つことが出来る。従って名詞表、動詞表と同一の関係形式をPRTVのサブスキーマで定義しておいても、その冗長性はシステムの効率に影響を与えない。この事は、ヤチマタのテーク・モデルとして内容の重複する関係形式を気軽に定義し、照会文をより流暢にすることを可能にする。

以下で各手続きにつき、より詳しい記述を与える。

2.2 詞辞解析

ヤチマタは適用業務に対して汎用であるという特徴を持つ。そのため照会文の解析に使用される詞辞は、適用業務に依存したものと全業務に共通に使われるものとの二種類がある。前者は適用業務が決定した時点で作られるものであり、後者はヤチマタ自身があらかじめ持っている詞辞である。

前者の例としては、

ITRAIN: 名詞表、動詞表内の要素名。「東京」、「1号倉庫」等。

CTRAIN: 名詞表名(普通名詞に対応)。「在庫量」、「所在地」等。

VTRAIN: 動詞表名(用言に対応)。「納品ある」、「飛ぶ」等。

Ptrain: 後置詞(格助詞等に対応)。「の」、「に」等。

DTRAIN: 拡張機能で追加される表現。単位(「cm」、「個」)等。がある。これらは適用業務用の辞書に格納されそれぞれ必要な「特徴」(構文のための情報)が記入されている。

後者の例としては、

PGA: 主格の格助詞(<名詞句>が<名詞句>である)。 「が」、「の」。

JTRAIN: 接続詞(接続詞、接続助詞に対応)。「と」、「および」等。

CTXPX: 文脈を指す表現。「あの」、「その」等。このことは、照会文中の語句を指すのではなく、以前の照会文に対する結果を指すのに使われる。

RCOM: ヤチマタ・コマンド。「LOGIN」(セッション開始)「LOGOUT」(セッション終了)「GRAPHQ」(木構造の表示。

デバッグ用)等。

VAUX: 助詞、助動詞、補助動詞。「た」、「ある」、「する」、「た」等。

PWA: 終端記号。「は?」、「は」。

COMOPR: 範囲を示す名詞。「以上」、「以下」、「以外」等。

等がシステム用辞書に格納され、それぞれ必要な「特徴」が与えられている。この他に、数字は規則として定義されている。

照会文はこれらの辞書や規則の適用により詞辞が割付けられる。辞書中の詞辞を照会文中より見つけたすには、欧米語のように空白を単語の区切りと考える訳にはいかない。各文字についてそれを先頭文字とする綴りかはいかにかどうか調べる。また辞書中の詞辞に空白の可能な位置を明記することにより、逆に妙な部分に空白のはいった

詞辞を合法と認めることを避けている。例えば、「セイト」(生徒)という単語が辞書に登録されており、向に空白が入る事を禁じているならば、次の文においてこの単語と照合される部分はない。

ショウガクセイ ト チウガクセイ

2.3 構文解析

構文規則は詞辞解析により得られる範疇(項の名前)を末端記号として文脈自由文法の記法で書かれる。ただし、その左右両辺に複数個の項が置ける。右辺の項には、その範疇が持つ特徴の制約条件を与えることができる。この制約条件が合わなければ、たとえ範疇の並びが同じであってもその規則は適用されない。左辺の項には構文特徴を与えることが出来る。その値は定数であってもよく、また右辺の任意の項のものを与えてもよい。書換規則以上に複雑な処理を要する場合にはそれを行なうサブルーチンを左辺の項の中で呼出すことができる。このサブルーチンの引数は、右辺のすべての項の全情報である。従って構文規則で書けずあらゆる処理がこのサブルーチン呼出しにより行なえる。オ2.2図に構文規則の例を示した。この規則は2.2節で詞辞解析を行なった後、辞書より得られた動詞<VTRAIN>を動詞句<VP>と見做す規則である。RF, LEXSW, CSB, DMBは範疇<VP>が持つ特徴であり定数「1」あるいは<VTRAIN>の持つ値(RFの場合)を与えている。VRBSETはこの名前を持つサブルーチンの呼出しを指示し、このサブルーチンでは<VTRAIN>

<VP: RF=RF(1), LEXSW=1, CSB=1,
DMB=1: IPE-VRBSET:FPE-QJVERB>
← <VTRAIN>;

オ2.2図 構文規則の例

の持つ特徴をもとにして<VP>の特徴に値を与える仕事をしている(ここでの値の設定は構文規則では書けない処理なのでプログラムで行っている)。QJVERBもまたプログラムであるが、これは意味の木で使われる関数呼出しである。

構文規則は以上の表記法により記述されている。ヤチマ夕で使われている構文規則を概観すると以下のようになる。ただし特徴の制約条件や値の設定については必要な場合以外は省いている。

辞書中のエントリーから句へ

詞辞解析により得られた末端記号の範疇から句を認識する。

<NP> ← <NUMBER>;
<NP> ← <ITRAIN>;
<VP> ← <VTRAIN>;
<VP> ← <VP><VAUX>;
<VP1> ← <VAUX>;
<VP1> ← <VP1><VAUX>;
<PP> ← <PTRAIN>;

ここで NP: 名詞句
VP: 動詞句
VP1: 断定の助動詞から成る動詞句
PP: 後置詞

動詞句から名詞句へ

動詞句は後ろに続く名詞句を修飾する形で使われる。

<VP> ← <NP><PP><VP>;
<NP> ← <VP><NP>;

名詞句からより大きな名詞句へ

いくつかの文型が存在する。

- (1) <NP: DMB=DMB(2)> ←
<NP: DMB=DMB(1)><NP>;
DMBは定義域を示し「部品ラジエター」などのように両者が同一定義域となる時にひとまとめにする。
- (2) <NP: DMB=DMB(2)> ←
<NP><VP1><NP: DMB=DMB(1)>;
「マネージャーである人」などがその例。

- (3) <NP> ← <NP> <PP> <NP>;
 (4) <NP:DMB=DMB(5)> ←
 <NP> <PGA> <NP> <VP1> <NP>;

(3), (4) はオ1.7図にある正順制約、逆順制約に対応する構文である。

- (5) <NP> ← <NP> <JTRAIN> <NP>;
 この構文規則は単純に当てはめると曖昧さを増す原因となる。例えば、
 「AとBかC」という句を
 (AとB)かC、
 Aと(BかC)

のどちらに解釈するかはシステムとして一意に決定できない曖昧な句となる。ところが「AとBとC」のように論理和と論理積が混在しない場合には解析は二通りにどまても結果は一つとなることがわかっているからどちらか一方の方向だけでやればよい。ヤチマタではこの両方の場合を特徴に対する制約条件を用いて区別している。

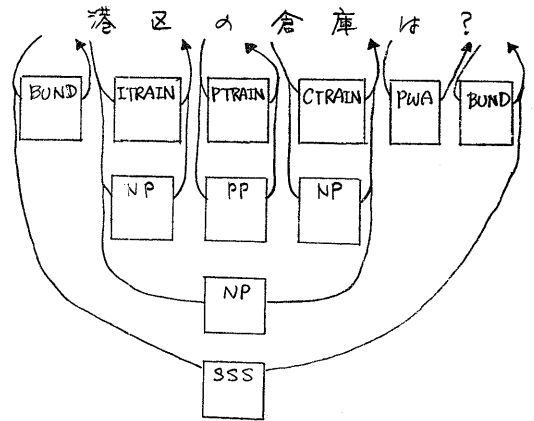
- (6) <NP> ← <CTXPX1> <NP>;
 <NP> ← <CTXPX2>;
 ここでCTXPX1は「その」、「あの」等。CTXPX2は「それ」、「あれ」等の指示代名詞を示す。

名詞句から文へ

- <SSS> ← <BUND> <NP> <PWA> <BUND>;
 <SSS> ← <BUND> <RCOM> <BUND>;

名詞句の後ろに「は?」あるいは「は」が付いたものが文と見做される。<BUND>は境界記号で入力文の最初または最後を示す。二番目の規則は特殊なコマンド(例えば「デバッグ」の指示等)に対するものである。

以上に述べたような構文規則を用いて構文解析が行われ結果は文全体に渡る木構造として有向グラフの形で示される(オ2.3図)。解析が失敗した場合には文全体に渡る木構造が存在せず部分的な木が点存在するだけとなる。この時には、部分木がまったく付いていない部分があればそこに「??...?」を付けて使用者に認識できなかった部分を



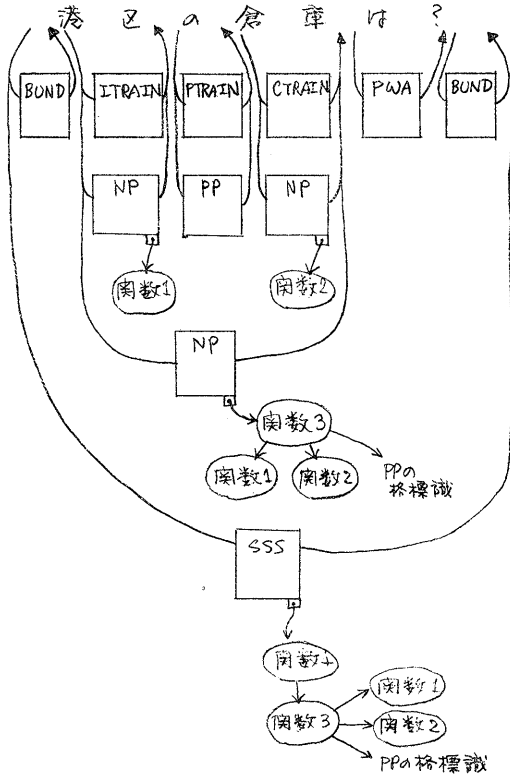
オ2.3図 構文解析による木構造

知らせる。構文上の理由から解析が失敗した場合には、意味の木(後述)を評価するプログラムがエラーの原因を表示するか、あるいは単に解析できない旨を表示するかのどちらかである。逆に文全体に渡る木構造が複数存在する場合は曖昧さが生じた時である。この場合にはどちらの木が正しいか(真向者の意図した照会文か)を選択させる。

2.4 意味解析

ヤチマタにおける意味解析とは照会文からデータベース検索のための形式的言語を作出す作業である。前述のようにこの形式的言語はPRTLVによって処理される代数型の複算式である。意味解析は構文規則中に書かれたFPE(Function Processor Extension)により呼出されるプログラムによって行われる。意味の木は構文解析が行われると同時に形成され、文全体に対する構文の木構造ができあがった時点で評価される。オ2.4図に意味の木の構造を示す。丸印で囲んだ部分がFPEによって呼出されるべきプログラム(関数)である。そこから、伸びた足(例えば関数を3に繋がる関数1, 関数2, PPの格認識)は関数への引数である。引数が関数である場合には、それを評価した結果が実際の引数となる。図中で範

時くSSS>に付けられた意味の木を評価すると求める複算式(PRTVへの入力)が得られる。なお蛇足ながら各関数の機能を述べておくと、関数1、関数2は名詞表の名前(PRTVが認識する内部コード)を作出す関数、関数3は正順制約を行なう関数、関数4は結果の表につきその列の順序を標準化する関数である。関数4により、複数個の<SSS>が生じた場合(曖昧さが生じた場合)にそれらが同一の答えを返したか否かの判定が楽になる。同一の答が得られたならば、使用者に複数個の解釈が成立した事を知らせる必要はない。



※2.4図 意味の木

2.5 曖昧さの処理

照会文における曖昧さはいろいろな原因で生ずる。それらは以下のように分類できるであろう。

- (1) 係り受け、接続詞、否定語の有効範囲等、人間にとっても不明確なもの。
- (2) テータ・モデルの設計に起因するもの。例えば一つの名詞(動詞)に複数個の意味をあてる場合。
- (3) 仮名入力による同音異義語。
- (4) 分かち書きを自由にした事によるもの。例えば「の、に、と」等はどこに現われても助詞と見做される可能性がある。

これらの原因別にヤチマウほどのような対応をしているかについて述べる。

(1)については、ある状況ではシステムが一意に解釈し、またある状況では使用者に選択を迫る。一意に解釈する例としては構文解析の接続詞の規則で述べたように同じ性質の詞群が並ぶため、どう解釈しようか結果が同じになる事がわかっている場合である。その他にも同じような表現が並列的に並ぶような句、「買った部品と売った部品」などは「(買った部品)と(売った部品)」と解釈する。(2)は設計された名詞表(動詞表)に左右される問題であるが、多重定義された名詞(動詞)が照会文中にある時に常に曖昧さが生ずる訳ではなく、構文規則により一意の解釈だけが残る場合も充分あり得る。(※3.1図において「××会社の収入は？」という照会文には曖昧さはない。) (3)はほとんどの場合に構文解析あるいは意味解析の段階で解決する問題である。もしこれが原因となり部分的には余分な木構造が作られるが文全体の木構造に紛込むことはほとんどない。

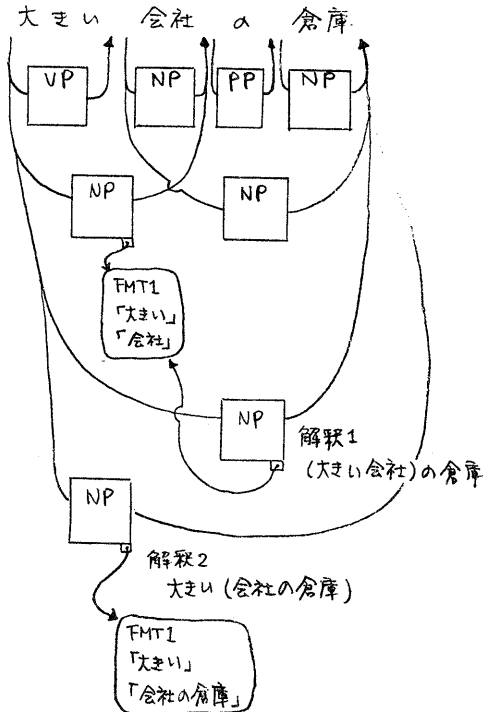
次に、不幸にして曖昧さが構文解析、意味解析を終えても解消されなかった場合の使用者に対する選択枝の表示方法に触れておく。これは曖昧さを生じそうな句を取扱う構文規則の左辺に表示すべき選択枝とその文章中で置換すべき文字列とを記憶させるよう指示する機能を用いる。※2.5図に例示した通りAMBは表示すべき文章(その中

で①, ②はいずれ置換えられる部分である事を示す)と置換えるべき文字列を取出す範囲を指定する。

構文規則:

<NP: AMB-FMT1, 1, 2> ← <VP> <NP>
<NP> ← <NP> <PP> <NP>;

構文解析:



表示文章一覽表:

FMT1: 「①」は「②」にかかる
FMT2: 〇〇〇
 〇
 〇

表示文:

正しい解釈の番号を入れて下さい。
1. 「大まい」は「会社」にかかる
2. 「大まい」は「会社の倉庫」にかかる

※2.5図 曖昧さ解消のための対話

3. 名詞句モデルの設計における問題

ヤチマタをある業務のデータベース検索に用いようとする場合、設計者は名詞句テータ・モデルに則って名詞表、動詞表を決定する必要がある。我々は現在までに三つの相異なる業務、
①自動車部品の在庫状況
②国内航空便の発着時刻など
③都道府県別地域統計

について実験を行って、その経験をもとに、ヤチマタをある適用業務を稼動させようとする場合の問題点及びそれに対する考慮につき述べる。

(1) 名詞句モデルに由来する限界

名詞句モデルにおいては定義域の異なる名詞を接続詞で繋ぐ事ができない。例えば、

「東京の人口と人口密度は？」

がそうである。これを完全に許すためには名詞表を演算しそのまま複数個並べられるような機能を与えるなどのモデルの根本的変更が必要である。

(2) 一つの名詞(動詞)に二つ以上の表を対応させる事による曖昧さの発生

例えば※3.1図の例では「1977年の収入は？」という照会文に対して曖昧さが生ずる。この状況に打ち合わせるため、ヤチマタは各表に説明のための文字列を付加してあり曖昧さ解消のために次のような文章を出力する。

「検索文中の「収入」は、1. 「人の収入」、2. 「会社の収入」のいずれを意味しますか？」

	"人の収入"			"会社の収入"	
収入			収入		
金額	人	年度	金額	人	年度

※3.1図 名詞「収入」の二重定義

②限定属性に同一定義域が複数個存在

この場合にも曖昧さが生ずる事がある。表3.2図の表で「JL005 の空席が30以上の空港は？」のような照会文で曖昧さが生ずる。この場合も格標識にその意味するところを文字列として与えれば、ヤチマタから選択のための文が表示される。

		C ₁₀	C ₁₁	C ₁₂
	数	フライト	空港	空港
	30	JL005	大阪	福岡
	⋮	⋮	⋮	⋮
	⋮	⋮	⋮	⋮

↑ "出発地" ↑ "到着地"

表3.2図 限定属性で曖昧さを起す例

(4) 検索結果の限定属性

結果に含まれている限定属性は設計者の意図したものが残るかこれは必ずしも検索者の要求に合っているとは限らない。必要な列のみを表示するように使用者に選択させる機能をヤチマタは持つが、各照会毎に選択を要求してくるためにかえって煩わしさが増すのであまり利用されてはいない。

(5) 名詞、動詞、後置詞の与え方

これらは新しい業種のデータベースをヤチマタに適用させる際に必要となるがこの詞彙の名称は便宜的なものであり、文法的な厳密さを要求している訳ではない。設計者はこれらを定義するにあたり、その名称に拘らず名詞句モデルとしてうまく動く語の単位を辞書に登録すればよい。我々の実験においても名称と合わない語が現われている。例えば、名詞として「労働者の平均月収」、「医者数」等があり、後置詞として「に従事している」等がみられる。

4. あとがき

ヤチマタで行われる照会には自然言語に近い文を用いているが、照会文に現われる文型に限られているために日

常見固まりする日本語の文型すべてを網羅している訳ではない。また照会として現われる文型のすべてという訳でもない。疑問文の代表的文型を挙げてみると、

- ① 何誰型 <名詞句> <は> <何(誰)か>
- ② 肯定・否定型

<名詞句1> <は> <名詞句2> <ですか>

にわけられる。しかしこれらは末尾の形を少し変形するだけの充分ヤチマタの文型<名詞句><は?>で表現できる。①では単に<何(誰)か>の部分省略するだけでヤチマタ照会文の文型に一致するし、②の文型は

<名詞句1'> <か> <名詞句2'> <であるのは>

という照会文の回答が空か否かを調べるのと同じである。従って照会文の能力を知る上で<名詞句><は?>の文型のみを用意すれば充分である。

また照会文の流暢さを増すためには単に辞書に登録するべき同義語の数を増やすだけでは充分ではなく、名詞(動詞)表の設計に重点を置く必要がある。この設計が悪いと解釈されない文が出たり、曖昧さが生じたりする。

参考文献

- (1) E. F. Codd, "A relational model of data for large shared data banks", CACM vol.13 No.6, 1970
- (2) 鷹尾他「ヤチマタにおける名詞句データベース・モジュール」情報DB研究会資料DB32-1, 1977
- (3) F. B. Thompson, "REL - a rapidly extensible language system", ACM National Conf., 24, 1969
- (4) S. Todd, "PRTV: Peterlee relational test vehicle", proc. of international conf. on VLDB, 1975