

## 表階層モデルに基づく自然語インタフェース処理方式†

絹川 博之††

不特定ユーザを対象とした計算機、各種 OA 機器をより使いやすいものとするため日常会話に近い自然語で、問合せ応答・操作指示できることが、強く要請されている。この要請に応えることを目的に、データベースシステムを対象として、自然語問合せ応答用の新しいインタフェース処理方式を提案している。すなわち、本研究では、(1)対象世界表現のために、関係表表現に、表構成属性項目と、その下位表との階層的連結情報を用いて、複数表間の意味的階層関係を表現できるようにした表階層モデルと、(2)属性項目名と属性項目データ表現語との組合せを基本単位として、問合せ文を構造化した日本語の属性項目単位型表現形式とを、提案し、(3)表階層モデルに基づく対象世界表現情報参照により、ボトムアップ的に、問合せ日本語の意味内容を認定するプログラム構造を設定し、自然語インタフェースを実現している。また、このインタフェースを具体的システムに適用し、評価している。この結果、問合せ日本語として、否定表現、大小関係指示表現、助詞の省略、丁寧な言い回しを含み、疑問文型・依頼文型を問わない自由な表現が可能であること、さらに、システムとして汎用的で、データベースシステムからの独立性を保持していることを確認し、自然語インタフェース実現の見通しを得た。

## 1. ま え が き

不特定ユーザを対象とした計算機、OA 機器を、より使いやすいものとするためには、日常会話に近い自然語で、問合せ応答・操作指示できることが、強く要請されている。このためには、(1)問合せ応答・操作指示表現に自由度があること、(2)応用分野が広く、フロントエンドとしてシステムに組み込みやすいことが必須である。この要請に応えることを目的に、データベースシステムの照会言語の自然語化が各所で研究されている。英語文用としては、70年代に RENDEZVOUS<sup>1)</sup>、REQUEST<sup>2)</sup>、LADDER/INLAND<sup>3)</sup>等が開発され、80年代に NLDBI<sup>4)</sup>、TEAM<sup>5)</sup>等が開発されている。TEAM は、分野依存情報と非依存情報との分離、及び、サービス提供者と窓口ユーザとにユーザ二分、の2点により、(1)、(2)に依っている。日本語文用のデータベース照会システムとしては、格標識の概念の導入により、片仮名疑似日本語照会文を関係代数表現に変換するシステム「ヤチマタ<sup>6)</sup>」や、意味ネットワークに基づく「KID<sup>7)</sup>」が研究開発されている。これらとともに、(2)の要請に応えているが、「ヤチマタ」は、助詞を格標識に利用しているため問合せ文中で助詞を省略できないし、問合せ文型を「--<名詞句>」に限定している。また、「KID」の問合せ文の受容可能範囲は、ユーザがプロダクション

ルールで記述する文法によっている。これに対して、本論文では、(1)、(2)の要請を同時に満たすべく、TEAM と同様ユーザを二分し、日常会話に近い自然な問合せ日本語から問合せ内容を認定する新しい方式を提案し、評価している。すなわち、

(i)対象世界表現のために、関係表表現に、表構成属性項目と、その下位表との階層的連結情報を用いて、複数表間の意味的階層関係を表現できるようにした表階層モデルと、(ii)属性項目名と、属性項目データ表現語との組合せを基本単位に、問合せ日本語を構造化した属性項目単位型表現形式とを、提案し、(iii)表階層モデルに基づく対象世界表現情報参照により、ボトムアップ的に、問合せ日本語の意味内容を認定するプログラム構造を設定し、自然語インタフェースを実現している。また、本方式のインタフェースを具体的システムに適用し、評価している。

以下、本論文では、表階層モデル(第2章)、自然語問合せ文の特性とその構造化表現(第3章)、自然語インタフェース処理系(第4章)、本インタフェースの評価と考察(第5章)について、述べる。

## 2. 表階層モデル

E. F. Codd は、1970年、データベースを関係の集まりとする、データ関係モデル<sup>8)</sup> (Relational Model of Data)を提案した。すなわち、対象世界の意味関係を(2.1)の形式の複数の関係表(以下、表と略す)で表すというのが、関係モデルの考え方である。

関係(属性項目-1, 属性項目-2, …) (2.1)

† A Natural Language Interface Processor Based on the Hierarchical-Tree Structure Model of Relation Tables by HIROSHI KINUKAWA (Systems Development Laboratory, Hitachi, Ltd.).

†† (株)日立製作所システム開発研究所

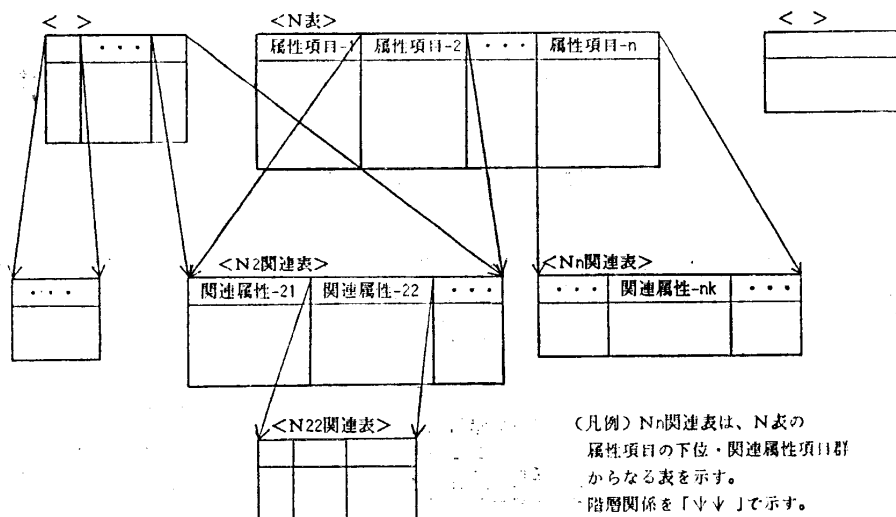


図1 表階層関係の概念図

Fig. 1 The concept of hierarchical-tree structure in relation tables.

以下、〈関係〉を表す語を関係名、関係(2.1)を構成する定義域を表す語を属性項目名、ある属性項目に属するデータ・値を表す語を属性項目データ表現語という。本論文では、これに加えて、(2.1)の複数の表の間に存する全体-部分、所有関係、対象とその属性群という上位-下位関係を(2.2)に示す階層関係で表すことにした。

関係 S 属性項目-n => 関係 T (2.2)

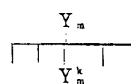
これは、関係 T が別のある関係 S の属性項目-n の下に位置づけられることを示す。なお、下位表中には、上位表の属性項目-n と結びつけるためのキー属性項目が含まれ、表を構成する属性項目は、排他的でなく複数の表を構成しうる。また、(2.2)の関係は、すべての表間に定義されるわけではない。(2.1)、(2.2)の階層関係の概念を図1に示す。

次に、関係名、属性項目名及び各属性項目データ表現語(ここでは、数値も含む)には、種々の同義語、上位・下位語があり、自然語文では、これらが、自由に使用されている。同義語、上位語については、標準的な言葉に置き換えても、意味的情報は、それほど失われないが、意味を木目細かく限定する下位語については、情報を失わないようにする必要がある。このため、対象世界表現において、データレコードを D. K. Hsiao<sup>9)</sup> の属性モデルに基づき、次の〈レコード〉形式(2.3)で表すこととした。

〈レコード〉 := (属性項目-1 キーワード,  
属性項目-2 キーワード, ...)  
〈属性項目-n キーワード〉 := <[IDW, ]DTW>

	属性項目名: $Y_m$		
	< $Y_m^k : y_m^k$ >		

$Y_m$ : 属性項目名  
 $Y_m^k$ : k 番目のレコードの属性項目名  $Y_m$  の下位語  
 $y_m^k$ : k 番目のレコードの属性項目データ表現語  
 なお、 $k \neq 1$  について、 $Y_m^k = Y_m^1$   
 ある 1 について、 $Y_m^1 = Y_m$   
 が成り立ちうる。



注)  $Y_m^k$  は、省略可。  
 既存 DB は、省略表現される。

図2 表に属するレコード構成の概念図

Fig. 2 The concept of table record structure.

<IDW〉 := 属性項目名表現語  
 | 属性項目名下位語  
 <DTW〉 := 属性項目データ表現語  
 | 属性項目データ下位語 (2.3)

本論文では、(2.1)に加えて(2.2)、(2.3)を「表階層モデル」と呼び、このモデルに基づき、対象世界を表すこととした。本論文の「表階層モデル」は、意味データモデルのうち意味ネットワーク形式のノード表現部に、フレーム表現形式と同様に複数の属性関係を定義できるようにしている点と、ある属性項目に対する複数の下位属性関係を、(2.2)により一つにまとめて表現できる点に特徴がある。これらにより、対象データ構造との親和性が増している。図2に、表に属するレコード構成の概念を、図3に、表階層モデルによる番

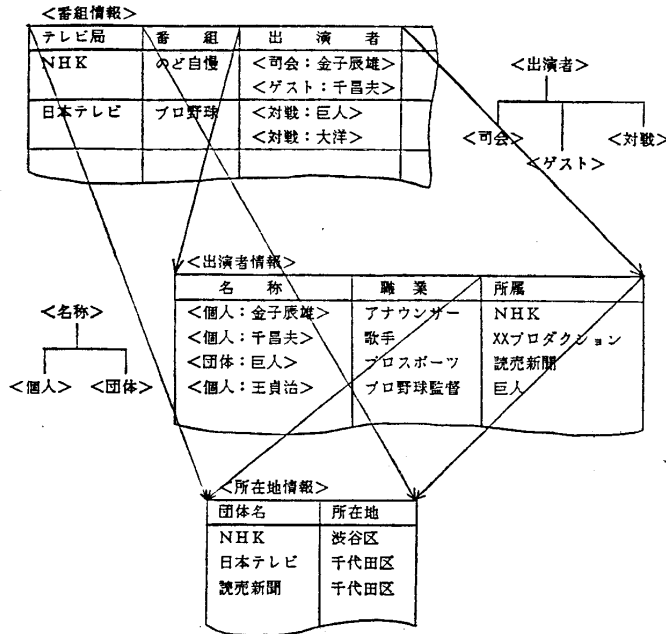


図 3 表階層モデルによる番組案内表現例  
Fig. 3 Expression of T.V. program information based on hierarchical-tree structure model.

組案内表示例を示す。

### 3. 自然語問合せ文の特性とその構造化表現

窓口ユーザの問合せ内容が、問合せ日本文中でどのように表現されているか明らかにする必要がある。このため、問合せ日本文中の問合せの手掛かりとなる表現について、第2章の表階層モデル、特に(2.3)を抛り所に、問合せ対象分野に依存しない特性を抽出し、これに基づいて構造化表現形式<sup>10)</sup>を設定することにした。この構造化表現形式設定の目的は、この表現形式の生成を目指す問合せ日本文認定方式を、分野に依存しないものとする事と、検索系から独立したものとする事の2点である。特性抽出に当たっては、問合せ文の内容の複雑さと表現の自由さを表す次の観点から意識的に種々創出することにより、問合せ文の基本的な例を、200例以上収集した。

- (ア) 問合せ条件, 問合せ対象それぞれについて, 明示的な表現, 曖昧な表現, 間接的な表現, 論理的に複雑な内容を持つ表現
- (イ) 日本文表現として, 同義表現, 文末の言い回し

収集は、筆者の属する研究室の8名によった。

以下に、抽出した問合せ表現の主な特性を示す。抽出特性例では、英大文字で属性項目名を、対応する英

小文字で各属性項目データ表現語を表し、具体例も付す。

(1) 検索条件は、以下に示す特性を有する。

- (a) 属性項目名と属性項目データ表現語とが、等号関係、大小関係指示語で結ばれている。
- (b) 属性項目名は、属性項目データ表現語と隣接して表されるか、省略されている。
- (c) 「等しい」関係表現は、明示されることが多いが、時には「等しい」「同じ」等の言葉が明示される。
- (d) 数値表現の大小関係を表す語として、「から」「まで」の付属語に加えて、「以上」、「以下」、「大きい」、「早い」等の自立語がある。
- (e) 数値表現の大小関係を示す語が、数値を取り得ない属性項目表現に付接することがある。

(例1) aより早いB (8時以前の番組)

          : ウルトラアイより早い番組)

——線: 大小表現部

(f) 等号/大小関係の否定を表す付属語は、属性項目名、属性項目データ表現語のいずれにも付接しうる。以下の(例2)(例3)は、同意である。

(例2) AがaでないBは? (出演者が, レーガンでない番組は?)

(例3) aがAしないBは何か? (レーガンが, 出演しない番組は何か?)

——線: 属性項目データ表現語

——線: 属性項目名, ~~~線: 否定表現部

(2) 検索結果の提示法を指示する部分の表され方を以下に示す。

(a) 未知の内容を知るための What 型文は、疑問指示語、または、属性項目名で指示する。後者については、対になる属性項目データが、表されず、ほとんどの場合、文末に近いところに現れる。

(b) 既知の、または、曖昧な事実を確認する Yes-No 型文は、検索条件を表す表現の羅列で示され、検索結果の提示の仕方が、一般には示されていない。

(3) 並列句は、接続詞または、並列助詞で表される。論理和を示す語がない時は、論理積を表すことが多い。

- (1) 〈ユーザ指示文〉 := 〈指示番号〉 : 〈指示動詞〉 : 〈指示条件〉 : 〈指示結果処置情報〉
  - (2) 〈指示動詞〉 : 検索 | 追加 | 削除 | 置換  
以下, 「検索」に限定し, (1) をさらに展開するものとする. 以下, 「指示」を「検索」に読み替えるものとする.
  - (3) 〈検索条件〉 := 〈検索論理式〉
  - (4) 〈検索論理式〉 := 〈検索論理項〉 | 〈検索論理式〉 〈論理演算子〉 〈検索論理項〉
  - (5) 〈検索論理項〉 := 〈条件属性項目キーワード〉 | 〈検索論理項〉 (〈検索論理項〉)\*
  - (6) 〈条件属性項目キーワード〉 := 〈属性項目データ〉 : 〈条件指示子〉  
| 〈属性項目名〉 : 〈同一属性項目データ〉 : 〈条件指示子〉  
| 〈属性項目名〉 : 〈処理条件語〉  
| 〈属性項目名〉 : 〈異条件属性項目キーワード〉 : 〈条件指示子〉
  - (7) 〈論理演算子〉 := AND | OR
  - (8) 〈条件指示子〉 : 属性項目データに対する等号関係, 大小関係を示し, 次の6種をいう.  
1 = : 等しい, 2 ^ = または = ^ : 等しくない, 3 < : 大きい,  
4 < : または < = : 等しいか大きい, 5 > : 小さい, 6 = > または < = : 等しいか小さい
  - (9) 〈検索結果処置情報〉 := 〈結果提示条件〉 〈結果所望件数〉
  - (10) 〈結果提示条件〉 := 〈提示属性項目記述〉 | 〈結果提示条件〉 〈論理演算子〉 〈提示属性項目記述〉
  - (11) 〈提示属性項目記述〉 := 〈属性項目名〉 | 〈疑問指示語〉 | 〈属性項目名〉 : 〈疑問指示語〉
  - (12) 〈属性項目名〉 := 〈関係(2.1)を構成する定義域を表す語〉
  - (13) 〈属性項目データ〉 := {ある属性項目に属するデータ・値を表す語}
  - (14) 〈疑問指示語〉 := {5W1Hの疑問を表す語 (例, いつ, どこ, 何)}
  - (15) 〈処理条件語〉 := {得られた結果の加工の仕方を指示する語}
- 注) \* : 修飾関係における修飾部を示す. なお, 簡約的に, 修飾関係部全体を { } で括り, { } 内の頭部に被修飾語を置き, その後に修飾語を「AND」で羅列表現することがある.

図4 属性項目を基本単位とする問合せ日本語の構造化形式  
Fig. 4 Expression format of Japanese-language queries.

- (4) 数値データに対して, 「最大/最小値」, 「平均値」を, 問合せ結果に対して, 「個数」を求める指示等がある.
- (5) 属性項目名, 属性項目データ表現語は, 名詞とは限らない. (例3)で「レーガンが出ない」という動詞を用いた表現もありうる.
- (6) 曖昧な条件を示す語((例)昼)が, 使用される.
- (7) 文末は, 疑問指示助詞が付接するとは限らず, 願望・依頼を表す言い回しや, 命令表現となることがある.

第2章の表階層モデルと, 以上の特性分析に基づき, 問合せ日本語を構造化した表現形式として, 分野に依存しない形式(図4に示す)を設定した. また, 属性項目を基本構成単位としているので, 以下, 属性項目単位型表現形式と呼ぶものとする.

#### 4. 自然語インタフェース処理系

表階層モデルに基づく, 自然語インタフェース処理系の構成を以下に示す(図5参照).

- (1) 自然語インタフェース処理系<sup>11)</sup>
  - (a) 対話制御 : 以下の(b), (c), 及び, (2)の制御
  - (b) 属性項目抽出型問合せ日本語認定 : 日本語

テーブルウェア, 表階層モデル世界表現情報を参照し, 入力問合せ日本語を解析し, 属性項目単位型表現形式に変換する. 本研究の主要部であり, 本章の以下の節で述べる.

- (c) 応答生成 : 検索結果を評価し, 表示する.
- (2) 検索系<sup>12)</sup> : 属性項目単位型表現形式をデータ

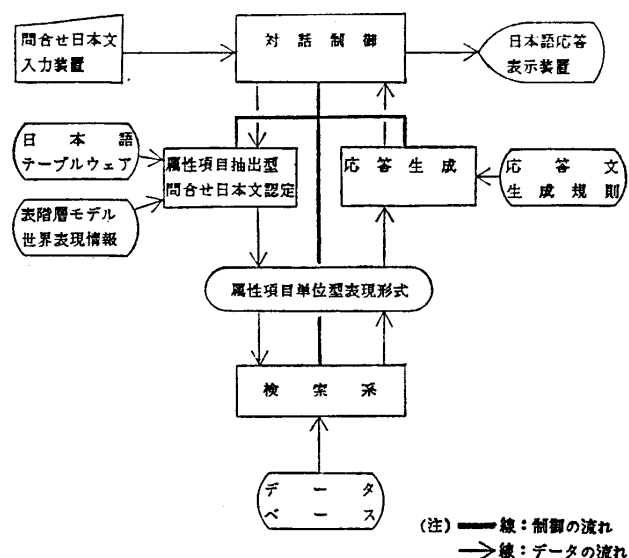


図5 データベース検索に対する自然語インタフェースの構成  
Fig. 5 System configuration of the natural-language interface processor for database access.

仮名見出し	漢字見出し	正規語	品詞	意味分類	種別
テレビキョク	テレビ局	テレビ局	<名詞>	A 1	<項目名>
シュツエンジャ	出演者	出演者	<名詞>	A 3	<項目名>
シュツエン	出演	出演者	<サ変語幹>	A 3	<項目名>
デ(ル)	出(る)	出演者	<一段動詞>	A 3	<項目名>
シカイ	司会	司会	<サ変語幹>	A 3	<項目名下位>
カネコタツオ	金子辰雄	金子辰雄	<名詞>	A 3	<データ>
ショゾク	所属	所属	<サ変語幹>	B 3	<項目名>
NHK	NHK	NHK	<名詞>	A 1	<データ>
NHK	NHK	NHK	<名詞>	B 3	<データ>
ショザイチ	所在地	所在地	<名詞>	C 2	<項目名>
ダレ	誰	誰	<名詞>	A 3	<疑問指示語>
ドコ	どこ	どこ	<名詞>	A 1	<疑問指示語>
ヒル	昼	昼	<名詞>	A 4	<曖昧語>
モットモ	最も	SUP*	<副詞>	XX	<関係指示語>
ハヤ(イ)	早(い)	早(い)	<形容詞>	A 4	<関係指示語>
マタハ	又は	又は	<接続詞>	<OR>	—
カツ	かつ	かつ	<接続詞>	<AND>	—
トコロ	ところ	ところ	<形式名詞>	<単文接続>	—
ソノ	その	その	<連体詞>	<指示代名詞>	—
サテ	さて	さて	<接続詞>	<不要語>	—
サイダイ	最大	MAX**	<名詞>	A 5	<処理条件語>
イゴ	以後	以後	<名詞>	A 4	<関係指示語>

注1) 品詞, 意味分類, 種別の各項目に関し, <N>は, Nを表すIDが, 実際にセットされている。

注2) 意味分類項目のコードは, 次の属性項目例を示しているとする。

A 1: テレビ局 A 3: 出演者 A 4: 開始時刻 A 5: 視聴率  
B 3: 所属 C 2: 所在地 XX: 属性項目いずれも可

注3) \*: SUP: 最上級を意味するものとする。

\*\* : MAX: 最大値を意味するものとする。

注4) 活用語尾は, ( ) で括って示す。

注5) <項目名>: 属性項目名, <データ>: 属性項目データ表現語

注6) 実際には, 辞書照合に適するよう, 仮名見出しに関して, ソートされている。

図 6 用語辞書—自立語収録形式とその例

Fig. 6 Dictionary record format and examples.

ベースシステムの検索言語表現に変換し, 検索処理を実行し, 結果を返す。

システムを構成する上で, 問合せ日本語を仮名・ローマ字で入力する場合は, 文節単位仮名漢字変換方式とした。これは, 条件を表す語の羅列入力の許容と操作結果の目視確認性をよくするためであり, ベタ書き入力方式<sup>13)</sup>とは, しなかった。一方, 漢字仮名まじりて入力する場合は, 字種の変わり目を分かち書き位置とみなせる<sup>14)</sup>ので, 特に分かち書きを前提にしなかった。以下に, 述べる方式は, 用語認定を除いて入力方式に依存しない。

#### 4.1 問合せ文認定に使用するテーブル類

属性項目抽出型問合せ日本語認定に使用するテーブル類の構成を示す。

##### (1) 日本語テーブルウェア

(a) 用語辞書: 問合せ日本語中に現れうる自立語

や, 条件指示付属語 (図4.(8)の条件指示子のこと), 日本語での頻出付属語, 文末判定機能語 (文末を示す語, 例. ますか, 下さい) のうち, 以下の自立語は, 同義語とともに図6の形式に定義し, 用語辞書に収録する。

①属性項目名 ②属性項目データ表現語

③疑問指示語 ④処理条件語

(①②③④は, 図4で定義。例は, 図6参照)

⑤曖昧語 (本項(c)参照)

⑥関係指示語 (本項(b)参照)

⑦指示代名詞 (例. その, この)

⑧接続指示語 (接続詞, 形式名詞)

⑨不要語 (問合せ日本語中でよく使用されるが, 問合せの手掛かりにならない語)

各収録語には, 品詞, 属性項目を示す意味分類情報 (⑦⑧⑨の区別を含む), 上記①-⑥の種

別情報、を付与する。ここで、品詞は、名詞、形容動詞語幹、サ変動詞語幹、動詞（段型・行で細分）、形容詞、副詞、連体詞、接続詞、形式名詞の種別を示す。

なお、上記3種の付属語類は、分野に依存しないので、前置語品詞との接続性により分類

正規語	指示内容	関連語数	関連語1	関連語2
早(い)	>>	1	*より	
早(い)	>>	1	*よりも	
⋮	⋮	⋮	⋮	
以後	=<	1	##	
⋮	⋮	⋮	⋮	
遅(い)	<<	1	*より	
遅(い)	<<	1	*よりも	
遅(い)	MAX	1	SUP(*1)	
⋮	⋮	⋮	⋮	

注) 「\*より」: 付属語「より」付きの語が比較対象語であることを示す。

「##」: 直前の自立語が比較対象語であることを示す。

(\*1): 最上級を示す語の出現により、「最も遅い」ことを示す。

図7 関係指示語定義表形式とその定義例

Fig. 7 Relation defining words and examples.

正規語	連続	表現語	条件指示子
昼	1	11時	=<
	0	14時	=>

(注) 「連続」の値は、残り条件数を示し、「0」が、末尾を意味する。この例では、「昼: 11時から14時まで」と定義している。

図8 曖昧語定義表形式とその定義例

Fig. 8 Record format of ambiguous words and examples.

(1) 表構成情報

表構成属性項目	表ID
ID 意味・表示情報	
A1 テレビ局	PGM
A2 番組名	PGM
A3 出演者	PGM
A4 開始時刻	PGM
A5 視聴率	PGM
⋮	⋮
B1 名称	PIF
B2 職業	PIF
B3 所属	PIF
⋮	⋮
C1 団体名	ADR
C2 所在地	ADR

(2) 表階層レベル情報

表ID	意味・表示情報	レベル
PGM	番組情報	1
PIF	出演者情報	2
ADR	所在地情報	3,2

(3) 表階層関係情報

下位表ID	上位表構成属性項目ID
PIF	A3
ADR	B3
ADR	A1

図9 表階層モデル世界表現情報の定義形式とその定義例  
Fig. 9 Format of world expressing information and examples.

し、あらかじめシステムに組み込むものとした。

(b) 関係指示語定義表: 等号関係、数値の大小関係と最上級指定を表す自立語の意味を、関連語  $n(n>0)$  との依存関係で定義するものであり、図7に定義形式とその例を示す。なお、関連語は、関係指示語の直前か、関係指示語に対応する属性項目名語の直前に現れる。

(c) 曖昧語定義表: 数値の範囲や抽象的な語の意味を定義する表であり、定義形式とその例を図8に示す。

(2) 表階層モデル世界表現情報: 次の3種の情報からなり、定義形式とその例を図9に示す。

(a) 表構成情報: (2.1)を定義する。

(b) 表階層レベル情報: 当該表が、階層木の何段目に当たるかを示し、下位表か否かの判別を使う。

(c) 表階層関係情報: (2.2)を定義する。

サービス提供者は、(1)(2)の各情報をインタフ

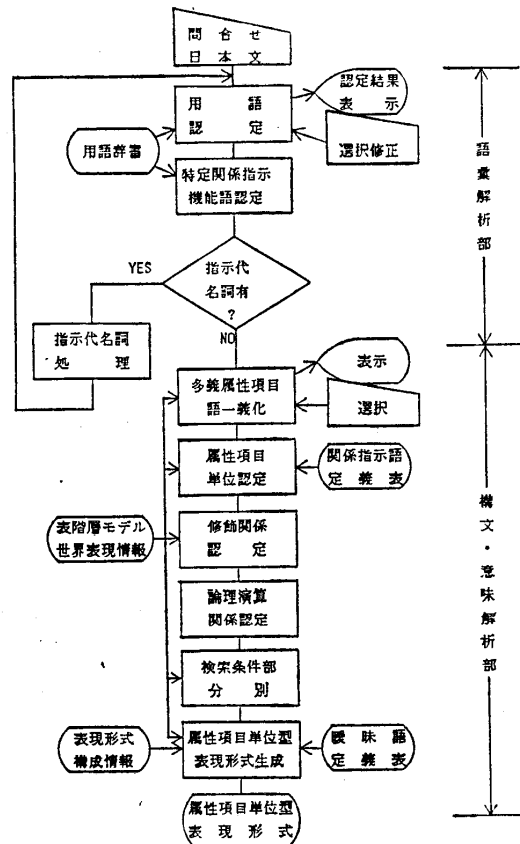


図10 属性項目抽出型問合せ日本文認定方式  
Fig. 10 Processing modules of meaning recognition for Japanese-language queries.

ューズ生成時に定義・登録する必要がある。また、窓口ユーザは、(1)(a)を適宜追加できる。

#### 4.2 属性項目抽出型問合せ日本文認定

属性項目抽出型問合せ日本文認定処理は、まず語彙解析部において、用語辞書照合により、入力文字列から語を認定し、次に構文意味解析部において、世界表現情報、関係指示語定義表、曖昧語定義表を参照し、語句の修飾関係、並列関係を順に認定し、属性項目単位型表現形式を変換生成する。以下、付録を適宜用いて、その処理方式を述べる(図10参照)。

(1) 語彙解析部：以下の、手順で処理する。

(a) 用語認定：最長一致照合による語認定。

付録の認定確認部の〈〉付き部が、属性項目キーワード構成語(例、〈歌手〉)である。

①同音異義語は、表示選択(例、エンゲイ：園芸、演芸)。ただし、以下の多義属性項目語の選択は、後述。

① 複数属性項目が付与される語

(例) 図6の「NHK」

② 付与される属性項目の上位属性項目が複数あるもの(例) 図3の「所在地」に属する語  
なお、疑問指示語も多義属性項目語となりうる。

②辞書非一致時は、頻出付属語を文字列末尾から除き、残りを未知語とし、不要語とするか否かを表示選択する。

(b) 特定関係指示機能語認定：(a)の認定文字列に対し、以下の機能語の付接の有無を認定する。

①等号/不等号関係を示す付属語の付接の有無

②並列関係を示す付属語、接続詞の付接の有無

③数値表現と認定された語に対して、大小関係を示す付属語、機能語の付接の有無

(例) 付録の第2文例の「〈歌番組〉でない」の「でない」を認定する。

(c) 指示代名詞処理：文頭に指示代名詞を検出した時、前文の〈検索条件〉部を補う。

(例) 付録の第4文例。

(2) 構文意味解析部：以下の、手順で処理する。

(a) 多義属性項目語一義化：多義属性項目語を以下の方式で一義化し、入力文中の認定語列を同一表構成属性項目群単位に分割する。

(i) 同一表構成の一義決定属性項目語で前後を括られている時、当該表構成の属性項目とする。

(ii) (i)でない時、多義属性項目語の直前、または直後に出現する一義決定属性項目語の属性

項目と同一表構成の属性項目を残す。

(iii) (ii)で、残すとされなかった語の属性項目が下位表を構成するものである時、後方に当該表の上位表構成属性項目語が、出現していれば、その属性項目を残し、他は、捨てる。

(iv) (ii)(iii)の後での多義は、表示選択。

(例) 「NHK」は、「NHKの金子辰雄」中では、〈所属〉と認定され、「NHKの番組」中では、〈所属〉の意味は捨てられる。

(b) 属性項目単位認定：世界表現情報と関係指示語定義表を参照し、図4の(6)(11)を認定する。

(c) 修飾関係認定：以下の条件及び修飾関係の非交叉条件により、図4の(5)を認定する。

(i) 以下のいずれかの条件を満たす下位表構成属性項目語は、後出の対応する上位表構成属性項目の体言との間の修飾関係を認定する。

①体言のみで、付接する付属語がない。

②体言に、「の」、「である」等の連体修飾の付属語が、付接している。

③連体形用言、連用形用言であっても、後方に連体形用言が現れる時は、含める。

(ii) 関係指示語に対し、関連語が結合された時、その関連語を含む分割語列((a)参照)中の(i)①②③を満たす語は、関連語を修飾するとみなす。

(iii) 単文接続指示語の直前の分割語列((a)参照)中の(i)①②③を満たす語は、その接続指示語の後方の初出の体言を修飾するとみなす。

(例) 付録第3文例の「歌手かつ男優である」は、「人」を修飾すると認定し、第1文例の「一である」は、本項(b)により、既に最上位レベルの属性項目キーワードとして括られているので修飾関係と認定されない。

(d) 論理演算関係認定：図4の(4)及び(10)を認定。論理和指示語の解釈を以下に示す。なお、明示のない時は、論理積とみなす。

A, Bi ( $i > 0$ ), C, D: 異属性項目語。ただし、Bi ( $i > 0$ ) は、同一属性項目語

OR: 論理和指示語とする。

(i) 「A・B1・B2・OR・B3・C」は、

「A and (B1 or B2 or B3) and C」

と解釈する。(and: 論理積, or: 論理和)

(ii) 「A・B・OR・C・D」は、

「(A and B) or (C and D)」

と解釈する。

(e) 検索条件部分別: 本節(2)(a)-(d)項の認定結果, 及び, 第3章(1)(2)項から, 入力問合せ文中の〈検索条件〉部と〈検索結果処置情報〉部とを分別する。この時, 不要語は無視する。また, 〈検索条件〉がないか, 問合せ文解析不能の時は, その旨表示し, 問合せ文変更入力を要求する。

(f) 属性項目単位型表現形式生成: 本節(1)項及び(2)(a)-(e)項の結果から, 属性項目単位型表現形式を生成する。曖昧語については, この段階で曖昧語定義表に定義された意味に置き換える。

以上述べたことにより, 認定処理方式は, 分野に依存しない方式となっている。

4.3 検索系の機能

検索系の主な機能は, 属性項目単位型表現形式から検索対象データの蓄積されているデータベースシステムの検索言語に変換し, 検索実行することである。

- (1) 検索条件, 検索結果処置情報に指示されている属性項目から, アクセス対象のデータベースの選択
- (2) データベースアクセス順序の設定: 他を修飾する属性項目, 表階層レベルの下位に位置する表の属性項目を先にアクセスする。異なる複数の表を, ある属性項目を介して結びつけ, 逐次探索するため, 検索系は, 各属性項目の定義域とその共通度を知っている必要がある。
- (3) 検索結果の提示情報の取得: 表階層レベルの上位に位置する表階層の属性項目を先に取得する。

4.4 実験システムの構成

4.2 節の認定方式の検証のため, 仮名入力のプログ

表 1 実験プログラム容量  
Table 1 Sizes of programs in the estimating system.

構成プログラム名称		容 量
日本語 インタフェース系	対話制御部 (含: 応答生成)	1.0 キロステップ
	日本語問合せ文認定部	3.3 キロステップ
検 索 系		3.0 キロステップ
辞書類保守系		1.4 キロステップ

- HITAC Mシリーズ PL/I
- 辞書: VSAM ファイル
- データ: 日立関係データベースシステム RDB1

表 2 実験に使用したテーブルウェア容量  
Table 2 Dictionary & tables sizes in the estimating system.

日本語テーブルウェア名		語 数
用 語 辞 書	番組案内用	1,413 語
	行事案内用	674 語
条件指示機能語		20 語
未知語推定・文末判別用機能語		88 語

ラム開発を行い, テレビ番組案内(問合せ例を付録に示す。)と, 行事案内に適用した。実験プログラム容量を表1に, 日本語テーブルウェア容量を表2に示す。

5. 本インタフェースの評価と考察

5.1 問合せ日本文の自由度

本問合せ日本文認定方式は, 問合せ内容の手掛かりとなる表現を抽出し, 表階層モデルに基づいて構造化した属性項目単位型表現形式に変換することを基本としており, 問合せ内容に関連のない表現は, 認定の対象にしていない。この基本処理方式により, 本インタフェースの受け付け可能な問合せ日本文は, 以下の(1)-(4)の自由度を持ちうる。

- (1) 助詞〈テ, ニ, ヲ, ハ〉の抜け, 使用誤りを含んでいてもよい。これは, 条件指示付属語以外は, 認定の対象にしないことによる。
- (2) 能動表現に限らず, 受身表現も可能。(1)と同じ理由による。
- (3) 疑問文型, 依頼文型いずれも可。
  - (a) 疑問文型: What 型, Yes-No 型いずれも可。これは, 図4(11)に定義されているため。
  - (b) 依頼文型: 命令調, 願望調, 丁寧調いずれも可。(1)と同じ理由による。
- (4) 問合せ表現において, 窓口ユーザは,
  - (a) データベース構造を意識する必要がない。これは, サービス提供者があらかじめ定義するため。
  - (b) 検索条件, 結果の提示法に対する指示の仕方を意識する必要がない。これは, 図4の表現形式, 4.2 節の処理方式, 及び, 用語辞書が同義語を含んでいることによる。
  - (c) 検索条件の論理和/論理積は, 接続詞, 並列助詞を用いて表現可能。
  - (d) 前問合せ文を受ける指示語表現が可能。



(e) 問合せ文中で否定表現が可能.

(c)(d)(e)は、4.2節の(1)(b)(c)及び4.2節の(2)(d)による.

本インタフェースは、付録に示す例を〔(歌手が出る) (歌番組でない) 番組〕と解釈している. 〔(歌手が出る歌番組) でない番組〕については、(d)項の指示語を使用し、「歌手が出る歌番組は? それ以外の番組は?」と表すことにしている.

(f) 曖昧な条件を表す言葉の使用が可能. これは、4.2節の(2)(f)による.

(5) 「すべて」「ある」という限量表現の認定は、今後の課題として、残っている.

(6) 辞書未収録語は、属性項目名を表す語と接続すれば、その語の属する属性項目を推定できる. しかし、適用分野の特性上、辞書未収録語の属する属性項目を特定できない場合もあるため、システムとしては、未収録語検出時に、その未収録語を対話的に補追できる機能を、具備している必要がある.

(7) 適用分野に関するだけでなく、システムの能力、会話の状態に関することも問合せする. 後者については、問合せが想定される項目を表階層モデルによる世界表現に含めるとともに、応答生成機能に状態追跡を行わせることにより、対処できると考えるが、今後、さらに検討を要する.

## 5.2 方式の独立性・汎用性

(1) 本インタフェースの問合せ日本語認定方式は、第4章に述べたように、分野に依存する情報を日本語テーブルウェア、世界表現情報にデータとして収録することにより、手続き的には分野に依存しないものとなっている. また、本インタフェース適用時の、

- ①検索対象データの作成・登録 ②日本語テーブルウェアへの語収録 ③表階層モデル世界表現情報の定義

という分野に依存する作業は、サービス提供者の仕事である. このため、窓口ユーザは、何を問合せできるかを知っていればよく、対象世界の階層構造がどう定義されているか知っている必要はない. この意味で本インタフェースは、適用分野から独立性を保持している. また、システムの熟知者の指導下で二つの実験検証システムを、いずれも、1人月程度の作業で実現できた点と、属性項目単位型表現形式に変更が発生しなかった点から、本方式では、独立性に加えて汎用性・可搬性を有していると考えられる.

(2) データベースシステムとの関連について

(a) 関係モデルに意味的な階層関係を定義できるようにした表階層モデルに基づいて対象世界表現しているため、関係データベースシステムを使用する場合は、整合が、取りやすい.

(b) 問合せ日本語を属性項目単位型表現形式に構造化し、データベースシステムの検索言語に変換する方式としているため、検索対象情報の蓄積されているデータベースシステムからの独立性を確保している.

(3) 表階層モデルの(2.2)により、全体-部分、所有関係、対象とその属性群という上位-下位階層関係を、(2.3)により、属性名と属性値関係、上位・下位概念関係を、表現できるが、このモデルの精緻化は、今後に残っている.

(4) 属性項目単位型表現形式は、5.1節の(1)-(4)に述べた自由な問合せ日本語を表現できるが、この形式の「検索」以外への拡張は、今後の課題である.

## 6. む す び

不特定ユーザを対象とした計算機、OA 機器をより使いやすいものとするため、日常会話に近い自然語で問合せ応答できるようにすべく、(1)対象世界表現のために、関係表表現に、表構成属性項目と、その下位表との階層的連結情報を用いて、複数表間の意味的階層関係を表現できるようにした表階層モデルと、(2)属性項目名と、属性項目データ表現語との組合せを基本単位として、問合せ日本語を構造化した属性項目単位型表現形式とを提案し、(3)表階層モデルに基づく対象世界表現情報参照により、ボトムアップ的に、問合せ日本語の意味内容を認定するプログラム構造を設定し、自然語インタフェースを実現した. このインタフェース方式では、用語辞書参照による語認定、対象世界表現情報参照による入力語列からの同一表構成語群の認定、及び、これらに基づく構造化表現形式の生成が、基本技術である. また、本インタフェースを具体的システムに適用し、評価した. この結果、問合せ文として、否定表現、大小関係指示表現、助詞の省略表現、丁寧な言い回しを可とし、疑問文型、依頼文型を問わない自由な表現が可能であること、さらに、システムとしての汎用性と、データベースシステムからの独立性を兼ね備えていることを確認し、上記の目的を満たす自然語インタフェース実現の見通しを得た.

謝辞 終りに、本研究の機会を与えていただいた(株)日立製作所システム開発研究所所長 川崎博博士、本研究を進めるに当り、ご討論と、実験検証プログラム・辞書作成に、ご協力いただいた、筆者の所属する研究室の主任研究員 河崎善司郎博士、近藤秀文主任研究員をはじめとする室員諸氏に、深く感謝します。

### 参考文献

- 1) Codd, E. F. : Seven Steps to RENDEZVOUS with the Casual User, *Proceedings of the IFIP Working Conference on Data Base Management*, pp. 179-200, North-Holland, Amsterdam (1974).
- 2) Plath, W. J. : REQUEST: A Natural Language Question-Answering System, *IBM J. Res. Dev.*, Vol. 20, No. 4, pp. 326-335 (1976).
- 3) Hendrix, G. G. et al. : Developing a Natural Language Interface to Complex Data, *ACM Trans. Database Syst.*, Vol. 3, No. 2, pp. 105-147 (June 1978).
- 4) Ginsparg, J. M. : A Robust Portable Natural Language Data Base Interface, *Proceedings of the Conference on Applied Natural Language Processing*, pp. 25-30 (Feb. 1983).
- 5) Grosz, B. J. : TEAM: A Transportable Natural-Language Interface System, *Proceedings of the Conference on Applied Natural Language Processing*, pp. 39-45 (Feb. 1983).
- 6) 藤崎哲之助ほか: データベース照会システム「ヤチマタ」と名詞句データ模型, 情報処理学会論文誌, Vol. 20, No. 1, pp. 77-84 (1979).
- 7) 牧之内顕文ほか: 知識に基づいた自然言語インタフェース KID の開発, ほか4編, 情報処理学会第30回全国大会講演論文集(II), pp. 1421-1430 (Mar. 1985).
- 8) Codd, E. F. : A Relational Model of Data for Large Shared Data Banks, *Comm. ACM*, Vol. 13, No. 6, pp. 377-387 (1970).
- 9) Hsiao, D. K. et al. : *Data Base Computers Advances in Computers*, Vol. 19, Academic Press Inc., New York (1980).
- 10) 絹川博之: 日本語インタフェースにおける自然語問合せ文解析の考察: 情報処理学会第29回全国大会講演論文集, 1J-2, pp. 1395-1396 (1984. 9).
- 11) 絹川博之: 日本語インタフェース処理方式, 情報処理学会第30回全国大会講演論文集(II), 2G-7, pp. 1625-1626 (1985. 3).
- 12) 近藤秀文ほか: 知識ベース管理システムの基本構想, ほか1編, 情報処理学会第30回全国大会講演論文集(II), pp. 1383-1386 (1985. 3).
- 13) 絹川博之ほか: 速記反訳システム, 情報処理, Vol. 16, No. 6, pp. 484-491 (1975. 6).

- 14) 絹川博之ほか: 日本語文構造解析における文節構成語の認定, 情報処理学会第18回全国大会講演論文集, 97, pp. 189-190 (1977. 9).

### 付録 テレビ番組案内に対する検索例

- ・昭和60年4月14日分のテレビ番組である。
- ・表示画面のハードコピーで示す。
- ・属性項目コードは、図6と異なるものを使用している。KW: キーワード, ST: 開始時刻, AC: 出演者, TI: 番組名, CH: テレビ局

#### (第1文例)

```
S: 質問を入力して下さい。
U: ショウジョウカ カルト タンゴチアブル アソクマエ

S: 質問の内容を確認します。
   正しければハイを入力
   誤まっていたら修正を行なって下さい。
   <歌手>が<歌手>と<俳優>である<番組>は
U: 片

          属性項目単位型表現形式
          (AC=歌手).AND.(AC=俳優)
          検索条件      TI
          検索対象

S: 2件が照合一致しました。
   番組NO      X 1 0 0 2 1
   番組名      あの歌・この人・60年
   番組NO      X 1 0 0 3 6
   番組名      春の波濤

   他に質問がありますか
U:
```

#### (第2文例)

```
S: 質問を入力して下さい。
U: カラオケ ショウジョウカ カルト タンゴチアブル アソクマエ

S: 質問の内容を確認します。
   正しければハイを入力
   誤まっていたら修正を行なって下さい。
   <歌手>が<歌手>する<歌番組>でない<番組>をしりたい
U: 片

          属性項目単位型表現形式
          (AC=歌手).AND.(KW=歌番組)
          検索条件      TI
          検索対象

S: 1件が照合一致しました。
   番組NO      X 1 0 0 3 6
   番組名      春の波濤

   他に質問がありますか
U:
```

#### (第3文例)

```
S: 質問を入力して下さい。
U: カラオケ カルト タンゴチアブル カルト ショウジョウカ アソクマエ

S: 質問の内容を確認します。
   正しければハイを入力
   誤まっていたら修正を行なって下さい。
   <歌手>かつ<俳優>である<人>が<出演>する<番組>は
U: 片

          属性項目単位型表現形式
          (AC={人}.AND.AC=歌手.AND.AC=俳優)
          検索条件      TI
          検索対象

S: 1件が照合一致しました。
   番組NO      X 1 0 0 3 6
   番組名      春の波濤

   他に質問がありますか
U:
```

(第4文例)

S : 質問を入力して下さい。  
 U : ツノヲシマシマシ  
 S : 質問の内容を確認します。  
 正しければハイを入力して下さい。  
 誤まっていれば修正を行なって下さい。  
 <歌手>は <松坂慶子>か <中村羅俊>か  
 <出演者>は <春の波濤>か <NHKの最もはやい番組の開始時刻>は <何時>ですか  
 U : はい

属性項目単位型表現形式  
 検索条件 SAME  
 検索対象 T I , A C

S : 1件が照合一致しました。  
 番組NO X 1 0 0 3 6  
 番組名 春の波濤  
 出演者 松坂慶子  
 出演者 中村羅俊

他に質問がありますか

(第5文例)

S : 質問を入力して下さい。  
 U : NHKノモトモナイカンクニノイサコクノカンパニ  
 S : 質問の内容を確認します。  
 正しければハイを入力して下さい。  
 誤まっていれば修正を行なって下さい。  
 <NHKの最もはやい番組の開始時刻>は <何時>ですか  
 U : はい

属性項目単位型表現形式  
 検索条件 (CH=NHK).AND.(ST=最もはやい)  
 検索対象 T I , S T

S : 1件が照合一致しました。  
 番組NO X 1 0 0 0 1  
 番組名 ニュース天気予報  
 開始時刻 0 6 : 0 0

他に質問がありますか

U :

(昭和 60 年 6 月 18 日受付)  
 (昭和 61 年 2 月 20 日採録)



絹川 博之 (正会員)

昭和 22 年生。昭和 45 年東京  
 大学理学部数学科卒業。同年  
 (株)日立製作所入社。以来、漢  
 字・日本語情報処理システム、  
 仮名漢字変換、自動インデクシ  
 ング、日本語文書処理、自然言語処理などの研究  
 開発に従事。現在、同社システム開発研究所第三  
 部主任研究員。電子通信学会、計量国語学会、  
 ACL 各会員。