

## マルチウィンドウシステムに対する音声操作

谷 幹也                  芝 温子                  市山 俊治

**NEC** 関西 C&C 研究所

{m-tani,shiba,ichiyama}@obp.cl.nec.co.jp

現在行なわれているマルチモーダルインタフェースの研究では、音声とポインティングを同時に入力として解釈することが多くなっている。このような統合解釈では、キーボードからの自然言語入力では文脈上の意味のみを持つ指示語(「これ」「この」「ここ」)などが画面上の指示対象を示すことになり、スコープの曖昧性と意味的曖昧性が問題となる。従来の研究では、このスコープの曖昧性を減少させるため、アプリケーションが表示するウィンドウを単一のものとして、研究を行なっている場合が多い。我々の目的は、アプリケーションが複数のウィンドウを同時に表示するようなシステムにおいて、2つの曖昧性を解消する方法を検討することである。本報告では、表示している各メディアに対する、音声入力とポインティングの状況を分類するとともに、その一部を解決する手法を考案し、マルチモーダルインタフェースを有する文献検索システムに適用した。本システムでは、ウィンドウフォーカス管理情報に表示内容の階層構造情報を導入することで曖昧性の解消を行なっている。

## Voice Operation for Multi Window Systems.

Mikiya TANI and Haruko Shiba and Shunji ICHIYAMA

Kansai C&C Research Laboratory, **NEC** Corp.

NEC Kansai Bld., 1-4-24 Shiromi Chuo-ku Osaka 540 Japan.

In the domain of multi-modal interface research, many researchers handle the synergistic interpretation of voice input and direct pointing. This interpretation has mainly two problems: scope ambiguities and semantic ambiguities on anaphora interpretation, that is "here", "this" etc., that has no means in the only keyboard inputs. They have been researched such an application that display only one window in order to limit the scope ambiguities. In this paper, we aim to classify the situations on multi window systems by the content of those windows and propose the general method to decrease ambiguities on this situation. We also report the method that resolve those ambiguities by introducing the content information to the focus window management system on our developing interface system for Multi modal document retrieval system.

## 1 はじめに

マルチメディア(音声、高精細静止画、動画)を入力できるPC、WSの低価格化、CD-ROM、MO、ハードディスクなど記憶装置の大容量化・低価格化、ネットワークの高速・広帯域化によるマルチメディアデータの転送可能性により、エンドユーザもより簡単にマルチメディアデータの提示、作成が可能になってきている。これに伴い、ヒューマンインタフェースの研究分野において、複数の入力手段をもち、これらを同時あるいは逐次に用いることが可能なマルチモーダルインタフェースの研究が盛んになってきている[17]。しかし、GUIでも問題になっているように、ポインティングジェスチャのみでは、指示対象のスコープを一意に決定することは難しく、この曖昧性を音声入力で絞り込む、マルチモーダルインタフェースでは、画面の各領域(ウィンドウ)に単一の意味を持たせ、スコープの曖昧性を軽減させていることが多かった。

一方、情報の氾濫にともない、ユーザは明確な検索要求のないまま、情報検索や情報提示などを行なうことが多くなってきている。ユーザは検索してきた結果を見ながら、そこから関連の情報を検索したり、表示させるような検索を行なうことになる。関連情報の検索をよりスムーズに行なうためには、検索結果や提示情報を複数のウィンドウによって有効に表示しておく必要が生じる。本報告では、マルチモーダルインタフェースの現状を概観し、同様の情報を持つ複数のウィンドウが存在する場合の状況を分類するとともに、その一部を解決する手法を考案し、マルチモーダルインタフェースを有する文献検索システムに適用した。本システムは、自然言語インタフェース構築ツール: IF-Kit[12]を文献DBに適用したものを日本語検索サーバとし、この文脈解釈部に、ポインティング情報、音声入力を解釈する統合解釈部を追加したものである。また、クライアント側に、ウィンドウ管理情報にウィンドウの表示内容を構造化して持たせることによって、情報検索を行なった結果ウィンドウを複数表示した状態で、ウィンドウコンテキストやスコープ曖昧性の解消を行なうことができるようにした。

## 2 マルチモーダルシステムの現状

現在の主要な研究の一部を、その使用している modality とアプリケーションの軸から分類したものが、図1[21,14,8,11,10]である。用語は、[3]に基づいている。もちろん、この他にも多くのシステムが存在する。作図システムでは、この分野の草分けである Put-That-there[6]の系列や、情報検索・質疑応答システムとしては、[4]や[5]も有名である。

特に、[5]は、マルチモーダルインタフェースを利用したコンテキストに着目した研究であり、照応参照

(anaphora)とコンテキスト(context)の直接操作による解決として、1)ある質問に対する答を、その質問と同じウィンドウ中に表示する。そしてその答に関する質問はその答のウィンドウの内で行なうことによって、コンテキスト(対話の状況)が視覚的に制限される。2)対話の流れを質問のツリーとして表示することで、現在の対話状況が明らかになってユーザは対話中ではどの問い合わせに対してもフォローできる。現状の研究では、ウィンドウシステム自体に対するマルチモーダル化を行なっている WSS[21,14]、影武者[11]、LIMSI-CNRS[1]を除けば、単一のアプリケーションが画面上の各領域の意味を定めて、曖昧性を減少させた状況で行なっている。

## 3 ウィンドウ内容と音声操作

### 3.1 GUIにおけるスコープ

画面におけるポインティング動作では、指示対象のスコープを表すことができない。そのため、GUIでのインタフェースでは、次のような場合がある。

- 指示点の曖昧性を最初からなくしてしまうもの。
  - 選択できるものをボタン、異なる色など視覚的に他と区別できるようにすることで、現在指示している対象を決定する。
  - 事前に選択できる範囲は示していないが、対象を指示(クリックなど)した瞬間にそのスコープの範囲を反転・枠囲いなどを行なうことで、指示している対象を視覚化する。
- 指示している対象の曖昧性を残すもの。
  - 同じ点でも複数の解釈がある場合、クリックを続けることで、そのスコープの範囲を変化させるようなもの。[19]

例えば、プレゼンテーションシステムやハイパーテキストなどある一連の定まった情報を作成・提示する場合は、指示点の曖昧性を最初からなくす形式の方がナビゲーションとしては優れていると考えられる。

ところが、作図システムや文章作成などの編集操作及び情報検索の検索条件を作成するための条件情報としてポインティング対象の情報を取り入れる場合、指示している点が指し示している対象が実際指し示している点の含むどのスコープの対象かを決定することは難しく、ものを掴む点を変化させることで対応したり、マウスのクリックに合わせて、解釈範囲を広げたり狭めたりすることが多かった[19]。ところが、このような操作は、直感的ではないため、初心者ユーザにとっては、扱うことが難しかった。

アプリケーション	モダリティ	入力				モダリティの統合形態	マルチウィンドウ
		キー	音声	ポインティング	その他		
汎用	Windows3.1 & WSS(NEC)	○	○	○		exclusive	○
	System7 & PlainTalk(APPLE)	○	○	○		exclusive	○
	影武者(沖電気)	○	○	○		concurrent	○
	ウィンドウマネージャ(LIMSI-CNRS)		○	○		synergistic	○
デザイン・描画	インテリアデザインプランニング支援システム(日立)		○	○		synergistic	
	VoicePaint(IMAG)		○	○		synergistic	
	MM-DrawingTool(NEC)		○	○		synergistic	
	TKDraw & S-tgif(早大)	○	○	○		alternate	
情報検索・注文	MM-SIMPLA(NEC)	○	○	○		alternate	
	画像検索システム(ATR)	○	○	○		alternate	
	the MAIA system(IRST)	○		○		exclusive	
	the MERIT system(GMD-IPSI)	○		○		alternate	
	TOSBURG I(東芝)		○		マットセンサ	synergistic	
	MultiksDial(東芝)		○	○	光電センサ	synergistic	
・案内 ・応答	電話番号案内システム(NTT)		○	○		exclusive	
	表情付き音声対話システム(SONY)		○			synergistic	

図 1: マルチモーダルシステムの比較

この問題を音声入力を用いて解決しようとしたものとしては、Put-that-there やインテリアデザイン構築システム、S-tgif、3次元画像における音声操作などが上げられるが、いずれもそれぞれのアプリケーションが出している一つのウィンドウシステムに関して行なうマルチモーダルシステムであり、マルチウィンドウシステムに対する研究は少ない。

一方、検索すべき情報が世界中に分散している現在では、単一の検索ではなくマルチメディアで表示される検索結果を用いて対話的に関連検索を行なっていく必要があり、より有効に関連検索を行なうため、複数の検索結果を表示しながら検索を行なうマルチウィンドウによる検索が必要となってくる。マルチウィンドウにした場合の問題点は上記のスクープの範囲が、従来より広がる他、他のアプリケーションが出しているウィンドウに対する指示をどう解釈するかという問題が存在する。

### 3.2 音声指示操作における状況

(1) 音声操作における指示語、省略の現象 音声操作における現象は大きく3つに分類できる。

#### 1. ポインティングに関する指示語

A) 「このX」「このX」のようにXが存在する指示詞

- Xが構成要素を特定させる情報の場合
  - 構成要素名を直接指示する場合  
(例) 「この段落」、「この語」、「この部分」 etc.
  - 構成要素が示す対象の上位概念で指定する場合  
(例) 「この場所」、「この人」 etc.
- Xが属性を指定する場合  
(例) 「この下線部」、「この網掛け」、「この赤い部分」 etc.
- Xが表示されている文字列の場合  
冗長ではあるが、このような入力もかなり存在する  
(例) 「この音声入力部が...」 etc.

B) 「これ」「これら」のように対象が不明確な場合

- 入力の操作用語から特定できる場合  
(例) 「これを印刷」 etc.
- 入力の他の部分から特定できる場合  
(例) 「これと同じ作者の本は」 etc.
- 文脈から特定できる場合  
(例) 「今年のこれは」 etc.

C) 意味分類を含む指示語(「ここ」「こっち」「こちら」)の場合

- 入力の操作用語から特定できる場合  
(例) 「これをここへ」 etc. .... 移動->移動先を表す
- ポインティング指示対象の意味分類から特定できる場合  
(例) 「ここに関する本は?」 etc.

D) 範囲を示す指示語(「これ位」)の場合

- ポインティングしている構成要素のいずれかが程度を表している場合  
(例) 「ここにこれ位に」.
- ポインティングジェスチャ自体が範囲を示している場合  
(例) 「ここにこれ位に」.

#### 2. 文脈に対する指示語

フォーカスの当たっているウィンドウによって指示対象が変化する。次の場合が存在する。

- 1) 範囲限定 「その中で」、「この中で」、...
- 2) 時間限定 「さっきの」、「すぐ前の」、...
- 3) 順序限定 「最初の」、「その3番目」、...

### 3. 省略

マルチモーダルインタフェースでは、自然言語入力インタフェースにおける省略 [20] の他に、次のような省略が起こる。画面上での指示点が明確な場合、ポインティング動作や指示対象を示す指示語「これ」「この」が省略されることが多い。また、1) 到着点を指し示す言葉（「ここへ」、「ディスクへ」、「プリンタへ」など）や、2) 状態を表す言葉（「これくらいに」、「赤に」など）の場合は、音声操作を表す用言は省略されることが多い。特に、マルチウィンドウにおいては、ウィンドウのフォーカス履歴を考慮に入れる必要がある。

(2) 音声操作の種類 ポインティングに対する音声操作は、次の3つに分類できる。各操作を表す音声は、通常用言表現として現れる。

1. 編集操作 削除、複写、新規入力、移動、リサイズ、属性の変更、交換、前へ、後ろへ
2. 物理操作 印字、再描画、保存、破棄
3. 情報取得 検索条件としての利用など、指示対象から富語的条件を獲得するもの

### 3.3 ウィンドウ提示情報

ウィンドウに提示される情報を内容から考えた場合、表1のように、8つのモダリティとその融合したものに分割される。どのモダリティにおいても、ウィンドウ全体で表示している対象は、いくつかの下位構成要素からなり、表示ウィンドウをトップとする階層構造をなす。また、プリミティブが融合したものは、階層構造が再帰的になっていると考えればよい。以下、各メディアに関するポインティングの曖昧性に関して述べる。

テキスト 表示テキストの一部をポインティングした場合、その文字、単語、文、パラグラフ、文章全体、表示ウィンドウ全体のどれを指しているかが明確でない。そのため、GUIのみでの構築では、指定する範囲をクリックによって広げたり、狭めたりする操作方法 [19] や、ボタンを押しながらマウスを移動させる（ドラッグ）操作によって範囲を特定している。音声による情報を用いることで、この指示対象範囲を設定することが可能な場合がある。指示している範囲を反転表示する場合には、視覚的に確認できてしまうが、ポインティング操作によるスコープの拡大、縮小は直感的

に分かりにくい。音声操作による指示範囲の曖昧性解消ができる場合はその方が良いと考えられる。

表 表示する表の一部をポインティングした場合、その文字、単語、文、セル、カラム全体、ロー全体、表全体の場合が存在する。カラムあるいはローに関する名前づけ情報があれば、その名前が入っている検索入力において、指示語の同定を行なうことができる。例えば、図書カードのような情報に対して、各項目名を確保しておけば、「この作者が書いた本は？」などは、図書カード上のどの点をポインティングしていても「こ」にあたるものは、そのカードが示す「本」であると理解できる。

アイコン icon は画面に表示されている階層構造情報で最下位の情報である。つまり、この構造より下部の提示情報はないと考えられる。マルチウィンドウシステムに関して習熟したユーザであれば、それを直感的に把握できるが、そうでない場合は、このアイコンの中の一部に対して、指示を行なうユーザが行なう場合もある。逆に、icon が意味する情報をアイコンにリンクすることで、情報取得操作が可能になる。

グラフ 大量のデータの関係を表すために用いられるグラフは、データの概要情報を表すためにも非常に有用であり [16]、特に情報検索系のアプリケーションで非常によく使用される。グラフの場合、棒グラフ、円グラフ、線グラフ、3Dグラフなど様々な場合があり、構成要素もそれぞれ、棒、扇型、線が対応することになる。グラフ表現の場合、属性による指定もかなり多く見られるが、特に、グラフの特性に関する入力がある場合も多い。例えば、折れ線グラフにおける「この山とこの山の間」のような入力である。

オブジェクト 作図、デザインなど CAD/CAM のアプリケーションにおいては、指示対象となるものが、複数の部品から組み上げられたグループ化オブジェクトであることが多い。GUI上では、指示ポイントをクリックする際に、反転、網掛け、矩形領域表示などを用いてスコープを明らかにしていることが多い。音声操作を伴う場合、提示している各パーツに関する名前情報が入っていれば、同じ指示点であっても指示対象を容易に決定することができる。CGの場合も、各パーツの表示がシステムによって制御されている場合は、オブジェクトの場合と同じである。

静止画 オブジェクトに分割できない静止画の場合、指示対象を表示された静止画全体とする他はないが、

表 1: 表示メディアと構成要素

表示メディア	構成要素
テキスト	語, 単語, 文節, 句, 段落, 節, 章, ページ, 文全体, 表示ウィンドウ
リスト	項目の階層, リスト全体, 表示ウィンドウ
表	セル, 行, 列, 表全体, 表示ウィンドウ
グラフ	グラフ構成要素 (線, 棒, 扇型 etc), グラフの特徴点, 表示ウィンドウ
オブジェクト	素構成要素, グループ化された構成要素, 全体, 表示ウィンドウ
ICON, CG	素構成要素, グループ化された構成要素, 領域, 全体, 表示ウィンドウ
静止画	領域, 全体, 表示ウィンドウ
動画	領域, 全体, 表示ウィンドウ

領域に対する階層構造情報が存在する状態 [18] では、音声操作によって決定することが可能となる。

**動画** 動画も静止画と同様の問題が存在し、特に時間軸方向に対象となる領域が変化してしまうため、時間感覚をおいたフレームに領域を指定し、そこから間の時間領域を補完することで、時間的な領域を指定する方法 [13] などが知られている。

### 3.4 マルチウィンドウに特有な問題点

**A) ウィンドウコンテキスト** 情報検索において、検索結果を表示するウィンドウを複数提示するようなシステムにおいては、対話の中で以前の結果を持っているウィンドウの情報を利用して検索対話を続行するようなコンテキストが存在する。

このような場合、照応関係や省略の処理の際に、現在着目しているウィンドウの情報とともに、より以前のウィンドウが持っている情報をそのフォーカスの履歴情報として持つ必要がある。

従来、複数のコンテキストを明示的に切替えるための質問系列とコンテキストを視覚化した研究は MAIA [9], パの関係にあり、あらかじめ登録されたユーザは同じ MERIT [7], XTRA [2] のように多数存在するが、ウィンドウが情報を提示したままコンテキストを保持しているようなシステムはない。

**B) 指示対象のスコープ曖昧性の増大** マルチモーダルインタフェースの構築の場合、3次元画像の検索 [15] などを除いて、指示点の曖昧性を減少させるため、画面を分割したウィンドウに意味を持たせ、同様の意味 (例えば、検索結果のウィンドウ) を複数提示するようにはなっていない。複数ウィンドウに表示された検索結果において対象指示を行なう場合、それぞれのウィンドウの選択をも含めた対象指示が必要となる。

## 4 マルチモーダル文献検索システム

本システムは、検索結果をマルチウィンドウを利用して提示する情報検索のインタフェースとして、マルチモーダルインタフェースの評価を行なうために試作したシステムである。本システムは、キーボード、音声入力、タッチパネルによるポインティング情報の入力が可能であり、キーボードから入力された自然言語入力によって検索された検索結果から、音声入力、タッチパネルを用いて関連検索を行なうことが可能である。

### 4.1 システムの概要と構成

(1) システムの全体構成 図2に本システムの構成を示す。本システムでは、ユーザインタフェース用 WS (EWS-4800/350) と音声認識用のパソコン (PC-9801Ap) と統合解釈、DB 用 WS (EWS-4800/360) の3台を使用している。音声認識用のマイクは音声認識用パソコンに接続されている。また、指によるポインティング情報を取り込むために、ユーザインタフェース用 WS のディスプレイにタッチパネルを取り付けた。ユーザインタフェースと統合解釈部はクライアント-サーバシステムを利用することが可能である。

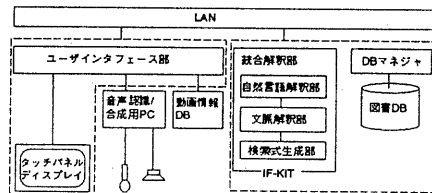


図 2: システムの全体構成

(2) ユーザインタフェース ユーザインタフェース部(以降 UI 部と呼ぶ)は、自然言語入力、システムからの問い合わせ、検索結果表示、マルチメディア対象表示の各ウィンドウを管理し、フォーカス管理と表示情報の管理を行なっている。

ユーザは自分の検索したい入力要求を自然言語(日本語)で入力することが可能であり、検索された結果情報からタッチパネルと音声入力に対話的に検索を続けることができる。

音声による入力は、Windows 上で利用可能な音声認識システム WSS[14] を利用して不特定話者連続音声認識を行なう。操作系入力でない場合、自然言語入力と同様に自然言語処理部へおくる。ユーザへの操作促進メッセージや操作指示は、画面に表示すると共に WSS[21] を用いて音声合成する。

(3) 自然言語処理部 UI 部で、キーボードあるいは音声入力された自然言語による検索文は、自然言語インタフェース構築キット: IF-Kit[12] の解析部を用いて中間構造に変換する。ポインティング情報を含めた解釈は、指示語、省略の解釈を行なう文脈解釈中に行なう。統合解釈結果は、IF-Kit によって対象データベースに適應した拡張意味ネットワーク用いて SQL 式へ変換し、データベースサーバの検索を行なって、結果をユーザインタフェース部へ送出する。UI 部は、フォーカス管理、ウィンドウ情報管理に登録した後、結果表示する。

#### 4.2 階層構造情報による解決手法

(1) 関連検索 図書検索などの場合、欲しい図書が明確である場合でも一回の検索で検索できることは少なく、また欲しい図書自体が不明確な場合が多い。こういった場合、検索してきた結果を元に、関連情報の検索を行なうことが多い。こういった検索では、画面上に表示された情報をもとに検索式を立てることになるが、本システムでは画面上に表示された情報に対するポインティングと指示語を含む音声入力によって、関連検索の操作をより自然に行なえる様にしている。

図 4 では、「掛川に関する本は」という検索入力文から「掛川市」という本を選択し、その結果を用いて、「これと同じ作者の本は」という関連検索を実行した例である。

(2) フォーカス管理 各検索結果ウィンドウの情報及び書誌事項ウィンドウに提示している情報は、ウィンドウのフォーカス履歴情報と共に図 3 で示す管理機構で管理し、文脈処理部からの要請に応じて UI 部が統合解釈部へポインティング情報を返す。例えば、図 4 のように、ポインティングしている指示点が図書カー

ド上のキーワード(“掛川城”)であった場合、指示対象は、次の 3 つの場合が存在する。

1. キーワード「掛川城」
2. 図書カードが表す本「掛川市」
3. 図書カードウィンドウ

フォーカス管理部は、図 3 の (ID:02) の場合のように、最も小さい構成要素から順に、情報の名前、上位階層情報上の属性、及び下位階層構造を持つ。3 つの解釈がそれぞれ選択される場合を、例を上げて説明する。

1. 「これに関する本は？」

入力から、「これ」は「本」の一属性を表すことになり、「本」の一属性(KW)の値を格納している最も小さい構成要素「掛川城」を「これ」と判断できる。したがって、「掛川城」に関する本は？」と解釈する。

2. 「これと同じ作者の本は？」

「これ」は“作者”という属性を持つ対象(「本」と判断でき、図書カードが表す本「掛川市」が対象として、「作者が“掛川市”の本は？」と解釈する。

3. 「これを印字」

物理操作として印字操作を指定し、「表紙」「内容」に関してのポインティングをしていないため、「これ」は“図書カードウィンドウ”全体と判断できるので、「“図書カードのウィンドウを印字せよ」と解釈する。

(3) ウィンドウコンテキスト 図 4 では、「これと同じ作者の本は？」の入力のため、一番手前の検索結果ウィンドウにフォーカスが当たることになる。ここで、ユーザが「掛川に関する本は？」に関する検索結果ウィンドウをポインティングしながら、「その中で最近の本は？」のような入力をした場合、絞り込む範囲を (ID:01) のフォーカス管理情報が保持している結果において行なう必要がある。このように、検索結果のようにマルチウィンドウを表示している場合、ウィンドウにフォーカスが当たった履歴を保持する必要がある。

## 5 考察

### 5.1 文脈処理

検索結果など同様の情報を同時に提示しているマルチウィンドウシステムでは、ウィンドウの選択によって文脈を変更することが可能であり、「その」などの指示語や省略の同定を行なう際に、ウィンドウコンテキストを利用する必要がある。複数のウィンドウを出

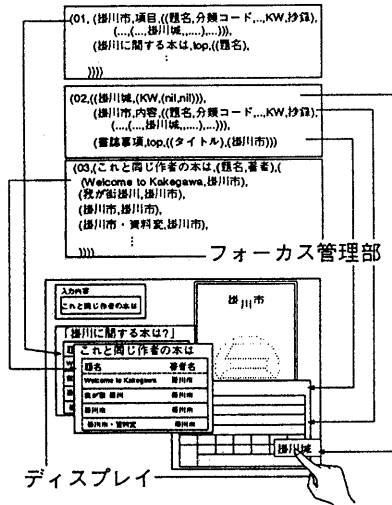


図 3: フォーカス管理機構

さない場合には、対話履歴自体をツリー上に表示してユーザにコンテキストをスイッチさせる [5] 方法もあるが、ウィンドウ管理情報として、フォーカス情報や位置情報の他に、指示対象を持つ全ての階層構造情報をリンクすることで、マルチウィンドウでのコンテキストの切替が可能になる。これは、全てのモダリティで表 1 の階層構造が仮定できるから、画面全体をトップとして、指示している対象点が含まれる全ての階層構造を明示することで実現できる。

### 5.2 スコープの曖昧性

上記のような、階層構造を導入し、各階層構造が自分の言語的情報と上位階層構造上の属性情報、下位階層構造の構造情報を持つことで、音声入力による絞り込みが可能になる場合が多い。ポインティングによる指示対象のスコープを直感的に決定するためには、この階層構造を明確にさだめる必要がある。

### 5.3 特徴データの抽出の必要があるメディア

検索を行なう際に、画面上での指示の対象とマッピングがとれない音声入力が行なわれることが多い。例えば、人が写っている静止画像で、画像の各領域に対する言語情報が張られていない場合、その人を指示しながら、「この人に関する...」という質問を発しても、システムは解釈することができない。この問題を解決するため、静止画などに関してはその領域分割を自動的に行ない、階層的な情報の作成支援を行なう手法 [18] などが考案されている。今後、数値的情報からグラフ

の構造的特徴を把握し、階層構造へ導入する方法を検討する必要がある。

### 5.4 階層構造情報の一貫性

上記で述べたディスプレイ全体をトップとする階層構造モデルは、全てが一つのアプリケーションからなるなら、すぐに構築することが可能である。しかし、様々なウィンドウシステムにおいて、ユーザの好みに応じたアプリケーションによって情報検索などが行なわれている。表示画面全体に関して階層構造情報を入れていくには、各ウィンドウに対して透明ウィンドウをかぶせることで指定を可能にする影武者 [11] のような方式をとることもできるが、ここで検討した階層構造情報は、さまざま検索場面で利用できることから標準化を目指すことが必要だと考える。

### 6 おわりに

マルチウィンドウシステムにおいて、ポインティング操作と音声入力によるマルチモーダル入力を行なう際の問題となる指示詞、省略の現象をまとめ、情報検索アプリケーションにおける指示詞の解消を行なう方法の一つとして階層構造情報の利用方法を提案した。また、文献検索システムのマルチモーダルインタフェースのウィンドウフォーカス管理情報に、表示内容の階層構造を導入することで、ウィンドウコンテキストへの対応と指示対象のスコープ曖昧性の解消を行なった。今後は、各モダリティに関する階層構造情報を規定し、一般化すると共に、様々なモダリティ情報を提示するマルチモーダル文献検索システムにおいて、ポインティング、領域指示操作と音声入力との統合による実用的なマルチモーダルインタフェースの研究を行なっていく予定である。

### 参考文献

- [1] Yacine Bellik and Daniel Teil. A multimodal dialogue controller for multimodal user interface management system application: A multimodal window manager. In *INTERCHI'93 Adjunct Proceedings*, pp. 93-94, 1993.
- [2] J.Allgayer, R.Jansen-Winkel, C.Reddig, and N.Reithinger. Bidirectional use of knowledge in the multi-modal nl access system xtra. In *IJ-CAI'89*, pp. 1491-1497, 1989.
- [3] Laurence Nigay and Joelle Coutaz. A design space for multimodal systems: Concurrent processing and data fusion. In *INTERCHI'93 Conference Proceedings*, pp. 172-178, 1993.
- [4] P.J.Hayes. Using a knowledge base to driven an expert system interface with a natural language component. In *Expert Systems: The User Interface*, pp. 153-182, 1988.
- [5] P.R.Cohen and et al J.W.Sullivan. Synergistic use of direct manipulation and natural language. In *CHI'89 Proceedings*, pp. 227-233, 1989.
- [6] R.A.Bolt. Put-that-there: Voice and gesture at the graphics interface. In *Computer Graphics*, volume 14, 1980.

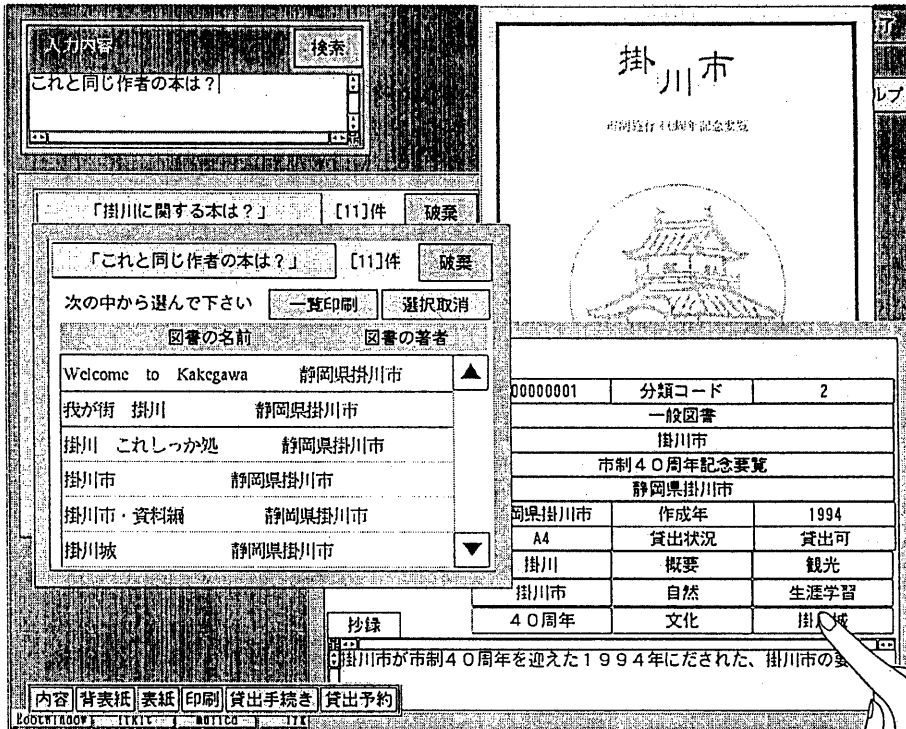


図 4: 関連検索の画面例

- [7] Adelheit Stein and Ulrich Thiel. A conversational model of multimodal interaction in information systems. In *Proc. of the 11th National Conference on Artificial Intelligence*, pp. 283-288, 1993.
- [8] Tom Thompson and Ben Smith. マルチメディアへの新しい回答 apple と sgi が提案する av ビジネス・マシン. 日経バイト, No. 118, pp. 146-157, 1993. PlainTalk.
- [9] Massimo Zancanaro, Oliviero Stock, and Carlo Strapparava. Dialogue cohesion sharing and adjusting in an enhanced multimodal environment. In *Proc. of the 13th International Joint Conference on Artificial Intelligence*, pp. 1230-1236, 1993.
- [10] 安藤 ハル, 北原 義典, 畑岡 信夫. インテリアデザイン支援システムを対象としたマルチモーダルインタフェースの評価. 電子情報通信学会論文誌, volume J77-D-2, pp. 1457-1474, 8 1994.
- [11] 奥村晃弘, 田川忠道, 宮崎敏彦. 入力マルチモーダル化ツール - 影武者 -. 第 48 回情報処理学会全国大会予稿集, number 4J-5, 1994.
- [12] 谷 幹也, 飯野 香, 山口 智治, 市山 俊治. 自然言語インタフェース構築キット:IF-Kit. 技術研究報告 NLC91-62, 電子情報通信学会, 1991.
- [13] 平田 恭二, 川崎 成人, 原 良兼. 動画ハイパーメディアシステム「雅」による電子魚図鑑システムの構築. 第 49 回全国大会予稿集, pp. 4.195-196(1W-7). 情報処理学会, 1994.
- [14] 磯 健一, 高木 啓三郎, 篠田 浩一, 山田 栄子, 服部 浩明, Farzad Ehsani, 渡口 淳, 古賀 真二, 畑崎 香一郎, 渡辺 隆夫. パソコン向けソフトウェア音声認識. 日本音響学会講演論文集, number 2-Q-21, pp. 66-72, 10 1993.
- [15] 望月 研二, 岸野 文郎. 対象物知識を用いた強靭作業空間における 3 次元対象物操作. In *Human Interface 8*, 10 1992.
- [16] 丸山 宏, 諸橋 正幸, 野美山 浩. 電子図書館 III-Information Outlining 触ってわかる情報の輪郭 - 第 49 回・情処全大会予稿集, Vol. 4, No. 2W-8, pp. 213-214, 10 1994.
- [17] 末永 康仁, 間瀬 健二, 福本 雅朗, 渡部保日郎. Human reader: 人物像と音声による知的インタフェース. 電子情報通信学会論文誌, volume J75-D-II, pp. 190-202, 1992.
- [18] 長尾 真. 電子図書館, chapter 4, pp. 73-79. Number 15 in 岩波科学ライブラリー. 岩波書店, 1 edition, 9 1994.
- [19] 野村 直之, 村木 一至. 言語の構造単位を保持した文書執筆支援システム. 情報処理学会研究報告, Vol. 91-NL-82, No. 10,, 3 1991.
- [20] 宮部 隆夫. ヒューマンマシン対話における照応解析. 第 44 回全国大会予稿集, pp. 3.201-202(5P-5). 情報処理学会, 1992.
- [21] 岩田 和彦, 高橋 一裕, 三留 幸男, 永野 敬子. パソコン向けソフトウェア日本語テキスト音声合成. 日本音響学会講演論文集, number 2-8-13, pp. 66-72, 10 1993.