

# 対話的マルチモーダル説明とその時間的協調

加藤 恒昭 中野 有紀子 中嶋 秀治 長谷川 隆明

NTT情報通信研究所

〒238-03 神奈川県横須賀市武 1-2356

{kato, yukiko, nakajima, hasegawa}@nttnly.isl.ntt.jp

Tel 0468 (59) 3677

あらまし 装置とその各種設定手続きを対話的に説明するシステムについて述べる。本システムは、2種類の対話的マルチモーダル説明を生成することができる。第一は、音声と指示動作による問い返し質問を受け付け、それに動作を伴った音声言語で回答することができる、構造化テキストと図面による説明である。第二は、図面への指示動作等と同期した音声言語による対話的説明である。本稿では、本システムの知識構造、構造化テキストや説明対話の生成機構について述べる。特に、これら音声言語を利用した説明において問題となる、音声言語と各種動作の同期（時間的協調）の実現について詳述する。

キーワード 自然言語生成, マルチモーダル対話, 音声対話, 同期, プランニング

## Interactive Multimodal Explanations and Their Temporal Coordination

Tsuneaki Kato Yukiko I. Nakano Hideharu Nakajima Takaaki Hasegawa

NTT Information and Communication Systems Laboratories

1-2356 Take, Yokosuka-shi, Kanagawa, 238-03 JAPAN

{kato, yukiko, nakajima, hasegawa}@nttnly.isl.ntt.jp

Tel +81-468-59-3677

**Abstract** A system that explains a machine and its installation is described. This system can generate two types of interactive multimodal explanation. The first, which allows users to make follow-up questions, uses formatted text and pictures. The questions are made by pointing gestures and speech, and the system answers them in speech coordinated with pointing gestures and animation. The second uses spoken explanations coordinated with pointing gestures and other actions. This paper discusses a mechanism for generating these explanation discourses and for interpreting and answering follow-up questions. The mechanism for synchronizing speech and other actions is explained in detail.

**Key Words** natural language generation, multimodal dialogue, spoken dialogue, synchronization, planning

# 1 はじめに

図面と自然言語等、複数のモードを協調させた説明を生成するマルチモーダル説明システムの研究が行われている [André&Rist93] [Feiner&McKeown90] [Maybury93] が、それらの多くは、記述文の列と図面とを生成して単に提示するだけのものであり、そこではユーザとの対話性は考慮されていない。一方、マルチモーダル対話システムと呼ばれているもの [Neal&Shapiro91] [Nagao&Takeuchi94] では、システムの発話は質問への応答が中心で、説明生成についてはあまり配慮がなされていない。本研究では、マルチモーダル説明における対話性を向上させるために、これを2つの方向へ拡張することを試みた。第一は、図面と、箇条書き等の構造を有するテキストとでなされた説明に問い返し質問を許し、それに音声言語と指差し等の動作を組み合わせるものである。第二は、専門家が初心者を前にして行うような、指差し等の動作を同期させた音声言語による対話的説明の実現である。本稿では、構造化テキストと音声言語の両方の生成を可能とする発話生成のメカニズム、前者に問い返し質問を許すメカニズムの実現について述べる。更に、音声言語を用いることにより、各種動作をそれと適切に同期（時間的協調）させることが重要となるが、本稿では特にこの問題について詳述する。

## 2 システムの概要

### 2.1 画面と構成

本システムの画面例を図1に示す。このシステムは留守番機能付電話機について、各種の初期設定や操作の方法等を日本語で説明するものである。画面の左側に電話機の図面が表示されており、その右下の漫画化された説明者が指示棒を動かして、この図面へのポインティング動作を行う（このような擬人化はPPP [Wahlster93]に刺激された）。右側に、指定された電話機の設定や操作を説明する、箇条書き等の構造を有するテキスト（以下、構造化テキストと呼ぶ）が表示されている。この画面に対して、ユーザは図面もしくはテキストの任意の位置をマウスでクリックすることで指示し、それに関する質問を音声によって行うことができる。システムはこのような質問に対して、音声言語と指示動作を含む各種動作とを同期させて回答する。例えば、テキスト中に現れる「応答1ボタン」をクリックして、「どこにありますか」と質問すると、音声言語によるその場所の説明に合わせて、そのボタ



図1 システムの画面例

ンが説明者によって指示される。図面中のボタン等を指示して同様に「何ですか」「何に使用しますか」というような質問を行うこともできる。個々の対象の指定は指示動作によって行われるので、認識対象語彙の数と較べて、可能となる質問のバリエーションが多いことがひとつの特徴となっている。

以上に加えて、右下のボタンを押すか、音声を用いて依頼することで、現在画面上のテキストによって説明されている操作を、音声言語によって説明させることができる。この説明でも、説明者による指示動作やカバーを開けるなどの模擬的動作が同期する。この説明は単なるテキストの読み上げではなく、独自にプランニングされ話し言葉として生成されるものである。

本システムの構成を図2に示す。ハードウェアとしては、音声認識装置と音声合成装置が繋げられたWSからなっている。音声認識装置はあらかじめ登録された句を認識するもので、認識結果として対応する識別子を返す。音声合成装置は、テキストを受け取り、それに適切な音韻情報を付与して合成を行い、発声する。発声が完了した時点で、それを呼び出し側に伝え

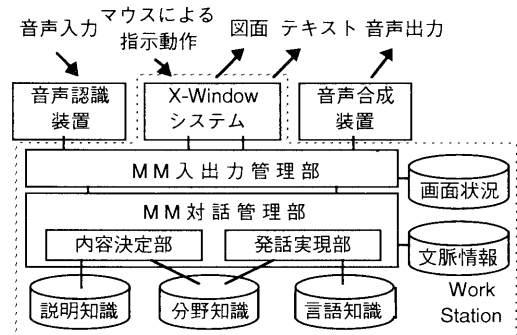


図2 システムの構成

ようになっており、これを利用して発声とその他の出力との同期がとられる。その他の入出力、つまり、マウスクリック等の入力を受け取りと画面への描画は、一般的なX-Window システムによっている。

MM対話管理部がシステムの中心であり、説明や応答のためのプランニングやスキーマ選択を行う内容決定部と、発話や動作の実現を行う発話実現部とから構成されている。MM入出力管理部は、MM対話管理部が行う意味レベルの入出力とX-Window システム等が行う下位の物理レベルの入出力との橋渡しを行う。画面への出力では、図面の仕様を受け取りそれを描画する。この時、画面のどの部分にどの部品が描かれたかを画面状況として蓄積し、これを参照して入力の解釈と出力の実現を行う。例えば、その図面に対してマウスクリックがなされた時、画面上の物理座標ではなく、何という部品へのクリックであったかをMM対話管理部に伝える。指示動作の実現については、ある図面のある部品を指示するようにというMM対話管理部からの依頼を受け取り、これを適切な画面上の座標に変換して、そこを指示する。更に、ある部品が現在どのような大きさで表示されているか、ユーザから問題なく見えているかというようなMM対話管理部からの質問に対しても回答することができる。構造化テキストについても同様で、テキストの内容と構造を受け取って、具体的な位置決めを行うのが、このMM入出力管理部である。テキスト中の要素へのクリックに対しても図面へのそれと同じようにテキスト中の要素の識別子を返却する。このモジュールの存在によって、MM対話管理部は意味レベルだけを意識してプランニング等を行うことが可能になっている。

## 2.2 各種の知識と説明の生成

システムで利用される知識は3つに大別される。第一は、文の生成に用いられる言語知識である。第二は、説明対象となる装置とその各種初期設定手続きに関する知識であり、分野知識としてまとめられている。この分野知識の一部が分野依存の単語辞書となっている。第三は、説明の生成や質問に回答するための、発話内容の決定で用いられるスキーマやプランオペレータ群である。この説明知識については、後の章で説明する。

言語知識は、*pattern-concept pair* [Jacobs85] とほぼ等価なプランオペレータ群によって表現されている。各々のオペレータは例えば、以下のような形式をしている。

### OP-1

```
(:pgoal (request s h (done h (action ?x)))
:fttr (scat = s pol = yes)
:ptrn (s ?y (aux "tekudasai"))
:sgoal ((?y . ((action ?x)
(scats = v2 vform = renyol
coord = yes))))))
```

これらのオペレータは、:pgoal で示された意味内容を:fttr で指定された統語的素性を満たして実現すると、:ptrn で示された構造となることを示している。:ptrn 中の変数には、:sgoal で指定された意味内容と統語的素性を満たして実現された構造が埋め込まれる。例えば、このOP-1は、ある行為?xを依頼するような文を実現するためには、その行為?xを動詞句連用形で実現し、それに依頼を意味する助動詞を付与すればよいことを示している。ここで、:sgoal に示されたものを下位ゴールと考えれば、詳細化に基づくプランナと同じ機構で文の実現が行える。

以下に示すOP-2は、ある行為を動詞句として実現するためのものである。:cndはこのオペレータの適用条件であり、ここに記述することで分野知識、大局的な実現のパラメータ、焦点スタック等の文脈情報を参照できる。この例では、分野知識を参照して、その行為の対象と述語を得て、それらを実現するようなサブゴールを呼んでいる。文型の選択や代名詞化についても、この:cndで文脈情報を参照することで行われている。:ptrnには、先と同様、実現される統語構造が記述されているが、この構造の節点に意味情報を注釈付けることができる。この例では-s?xによって、この節点が行為?xを表現しているのものであると注釈付けられる。

### OP-2

```
(:pgoal (action ?x)
:fttr (scats = v2 vform = ?v tense = ?t
pol = ?p coord = no)
:cnd ((kb (action ?x)
(obj = ?o p-type = ?pr)))
:ptrn (v2 -s ?x ?ol ?pr1)
:sgoal ((?ol . ((object ?o) (scats = p2 p = wo))
(?pr1 . ((pred ?pr)
(scats = v0 vform = ?v
tense = ?t pol = ?p))))))
```

分野知識は個々の動作(行為)、状態、出来事、対象それぞれについてフレーム形式でまとめられている。

行為については、

- ・説明の対象となるある行為がどのような下位行為か

ら構成され、それら下位行為の間にどのような時間関係が成り立っているか。

- ・その行為の実行の結果（効果）として、どのような状態が生じるのか。
- ・その行為の対象が何であり、どのような動作や述語で表現されるものであるのか。
- ・その行為が画面上で模擬可能か否か。

状態については、

- ・その状態の対象が何であり、どのような用言や述語で表現されるものであるのか。
- ・ある状態は知覚可能であるのか、また、その状態であることは何を意味するのか。

対象については、

- ・その対象の位置、ラベル、色、大きさ等の特徴と名称。
- ・その対象が関連する動作や状態。

これらの知識の中で、対象の名称や述語名等が言語知識の:ptm, :ftr, :sgoalに相当する3つ組となっており、これが分野依存の単語辞書として働き、分野知識と言語表現とを対応付けている。

ユーザからの質問や要求を受け取ると、内容決定部は説明知識と分野知識を用いて、説明や回答の内容を決定する。発話実現部は言語知識と分野知識を用いて、詳細化に基づくプランナによって与えられた内容から発話と動作を生成する。内容決定部は、その発話によって達成すべき言語行為を発話実現部に渡すだけでなく、実現に影響を与えるパラメータと焦点や共有信念等の文脈情報の初期値を指定する。パラメータによって、様々なタイプの文や発話、更にはそれと同期した行為が生成される。焦点や共有信念等の文脈情報の初期値を与えられ、それらを実現部で管理することで、複数の文や発話の生成の際の整合性が保たれ、質問への回答等においても適切な焦点を保った文生成が可能となる。

実現のパラメータとしては、以下のものがある。

モード 音声言語の生成か、テキストの生成かを指定する。対象の名称に含まれる記号や略号など特殊な読みを持つもの（字面と音の対応が一般的でないもの）への対処である。

動作の同期 言語のみを生成するか、それに伴う指示動作や模擬動作を生成するかを指定する。

対象の同定 一般には、対象の名称をユーザが既に知っているかは文脈情報によって管理され、既知の場合はその名称を用いて同定が行われ、未知の場合は、場所、色など同定に必要な情報を伝達し、名称を伝える。特殊な場合として、常に名前だけで対象

の同定を行うことを指定する。

代名詞化の可否 一般には、焦点の情報は文脈情報として管理され、焦点となっている対象、動作、状態については、代名詞化が行われる。特殊な場合として、これを禁止し、代名詞化を行わないことを指定する。

### 3 構造化テキストと問い返し質問

#### 3.1 テキストの生成と質問の解釈

第一のタイプの対話的マルチモーダル説明は、問い返し質問を受け付ける構造化テキストである。[Hovy93]でも指摘されているように、箇条書きやタイトル等、テキストを構造化する道具立ては、それぞれ意味を担っている。ここでは、説明における役割に応じて、各部分で適当な道具立てを利用することを考えた。

以下のスキーマに基づいて、指定された操作の説明が、構造化テキストを表現する木構造として生成される。

- ・ section の title として、操作を名詞句として生成する。例 「メッセージの消去」
- ・ その操作でどんな対象を利用するかを述べる文を生成する。例 「消去ボタンと応答1ボタンを使います」
- ・ その操作を達成するための下位行為を itemize して並べる。例 「・カセットカバーを開けます」
- ・ 下位行為の具体例等、補助的な情報は、quotation して挿入する。
- ・ その操作の結果状態で、認知可能なものとその意味するところを文として述べる。例 「ピッと音がして、メッセージが消去されます」

このスキーマによって得られる構造を図3に示す。ここで、title, itemize, quotation は、ほぼ LaTeX のそれと同じレイアウトを生成し、この構造から、例えば、図1の右側のようなレイアウトを持った構造化テキストが実現される。文の実現においては、全ての対象の同定を名前で先行し、代名詞化は行わない。この指定により簡潔で、マニュアルとして相応しい説明文が生成される。

前章のOP-2で示したように、文生成の過程で、文章中に現れる行為の表現や対象の表現にはすべてその識別子が注釈付けられる。例えば、図3に楕円で示したものが注釈で、カセットカバーを示す単語にその識別子である c-cover が、カセットカバーを開けるという動

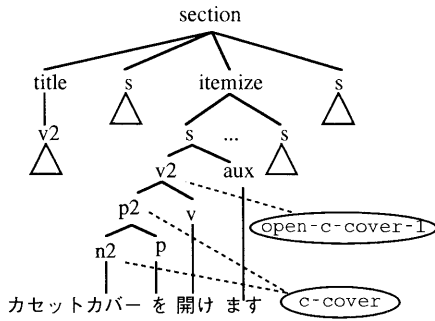


図3 構造化テキストを表現する構造

作を表現する動詞句の根にその行為の識別子 open-c-cover-1 が振られている。これを用いることで問い返し質問が理解される。例えば、図のテキストの「カセットカバー」の部分をクリックして、「これはどこにありますか」等と質問することができ、MM対話管理部はテキスト構造の注釈から「これ」がカセットカバー c-cover であることを理解する。このテキストの構造を維持していることが重要で、これにより、同じ場所を指示して、「どうなりますか」と質問すると、これが動作についての質問であることから「これ」は c-cover ではなく、構造を参照して、それを含んだカセットカバーを開くという動作 open-c-cover-1 であると解釈され、その行為の結果が音声言語で回答される。

全く同様のことが、図1に示された左側の図面でも行われる。この図面も描かれる際に画面のどの部分がどの部品に対応するかが記録され、部品間の階層構造が保持されている。そのため、個別に知識を記述していない対象、例えば、ダイヤルボタンの1を指示して、その役割等をユーザが質問した場合、階層を上昇的に辿って、知識が記述されているダイヤルボタン全体について回答することができる。もちろん、ここで辿られているのは図面における空間的な包含関係に基づく階層関係であるが、これは機能的な関係と殆ど常に対応しているので、実際上の問題はない。

なお、このような指示動作の理解についての関連研究としては、表への指示を解釈するシステム XTRA [Wahlster91] があり、同様に表の階層構造を利用して指示の曖昧性に対処することが提案されている。また、[Moore&Swartout90] では、テキストの修辞構造を保持することが、問い返し質問の理解において必要であると指摘されている。本システムで利用しているのは、これらより細かい、ひとつ発話内の構造と図面の構造である。

### 3.2 回答における時間的協調

問い返し質問として、テキスト中の対象が指示され、その場所が質問されると、以下のスキーマにしたがって、回答が生成される。ここでは、このスキーマのレベルで音声言語と指示動作との同期がとられる。

- ・説明すべき対象が現在隠れていたなら、それが見える状態になるような模擬動作を行うと同時にその動作を行うことを宣言する文を発声する。  
例 「カセットカバーを開けます」
- ・対象が、図面の隅にある場合は、図面を全体を動かして (pan して)、画面の中央付近の一定区域に入るようにする。
- ・対象がしきい値より小さい場合は、拡大する (zoom する)。
- ・対象の場所を述べる文を、その対象を焦点として生成する。  
例 「それはカセットカバーの中の左端にあります」
- ・上記の文が発声を始めると同時に、説明者はその部品に向かって動きだす (最初は歩かせようとしたが、現状では単に移動しているだけである)。
- ・対象のその他の特徴を述べる文をその対象を焦点として生成する。  
例 「それは黄色い小さなボタンです」
- ・説明者がその対象を指示して停止した後、「ここで」 「よろしいですか」という確認を生成する。

この最後の状態の例を図4に示す。カセットカバーが開けられ、その中の消去ボタンが指示されている。

## 4 音声言語と動作の協調

第二のタイプの対話的マルチモーダル説明は、音声言語と動作とを協調させた対話的説明である。前章で

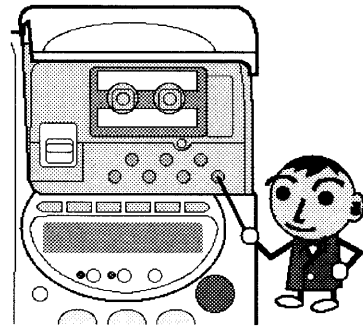


図4 説明者による指示動作 (1)

は説明自体は図面付き構造化テキストとして生成され、それへの問い返し質問と回答が対話的に行われたが、こちらでは、設定手続きの説明それ自体が対話的に行われる。

#### 4.1 説明内容の決定

この対話的説明は、専門家が初心者者に操作を教えると同時に、それをフォローさせているような状況で専門家が行うようなものである。実際、説明生成の基準はこのような状況に関する実験研究 [Nakano&Kato95] から得られている。内容決定部は、これまでの説明の状況を考慮して、説明において、ある設定手続きを達成するために必要な下位行為の依頼以外に、どのような内容を伝達すべきであるのか、例えば、行為の結果生じる状態の説明やある行為の例の提示を説明に含めるべきか、を決定する。また、ひとつの発話中にどの程度の数の行為の依頼やその他の内容を含めるべきかを決定する。

現在の実装では、最初の説明の後、ユーザに操作がうまく行われたか否かを尋ね、その結果を説明の状況として用いて、成功した場合は、初回よりも簡潔になるような復習の説明を、失敗した場合は、初回よりもやや簡単になりながらも、初回では述べなかった情報を積極的に付け加えるという説明を生成する。この説明内容の決定は単なるスキーマの選択ではなく、RST [Mann&Thompson87] に基づくプランオペレータを用いた詳細化に基づくプランナによって行われており、オペレータの適用条件を用いて、このような説明の状況に対する考慮が実現されている。

#### 4.2 発話実現における動作との同期

前章における音声言語と動作との同期は、粒度の荒いもので、発話実現部の外側で同期を取れば充分であった。一方、実際に専門家の説明を観察すると、ひとつの発話の中の細かい部分に同期して、対象への指示動作や模擬動作が行われている。これを実現するためには、発話実現部の内部での同期が必要となる。このような同期を、2章で述べた言語知識を言語だけでなく動作をプランニングできるように拡張することで実現した。以下に示すオペレータはその例である。

##### OP-3

```
(:pgoal (action ?x)
:ftr (scat = v2 vform = ?v tense = ?t
      pol = ?p coord = yes)
:cnd ((par (sync = yes))
      (kb (action ?x)))
```

```
(simul = t decomp = nil)))
:ptrn (sync ?x1 (action ?x))
:sgoal ((?x1 . ((action ?x)
                (scat = v2 vform = ?v tense = ?t
                  pol = ?p coord = no))))))
```

ここでは、coord = yes という属性をもって行為を表現するとき、更に、大域パラメータとして、同期が指定されており (sync = yes)、その行為が模擬可能なものであったら、sync というノードを生成し、その長女を行為の言語による表現とし、次女をその行為の実行とすればよいということを指定している。長女で指定された行為の言語表現は2章で示した OP-2 を用いて実現される。ここで、coord という属性は、その動作の表現が持っている文 (発話) の中での役割によって決定される。例えば、2章の OP-1 は、ある行為を聞き手に実施させるための命令文を生成するためのものであるが、この場合のサブゴールである行為の表現は模擬動作を同期させるよう指示される、つまり、coord = yes である。これにより、例えば、「ハンドセットを取ってください」という文には、ハンドセットを取る動作が同期する。一方、文の条件節の中に動作の表現がある場合、例えば、「ハンドセットを取るとランプが点滅します」においては、一般に動作を同期させるべきはない (ただし、操作の教示という強い文脈の下では例外的な事象も観察されている)。このような同期の指定を coord 属性の値を伝播させることによって行っている。

以下に示す OP-4 は対象への指差し動作を生成するためのものである。

##### OP-4

```
(:pgoal (object ?x)
:ftr (scat = p2 p = ?p)
:cnd ((par (sync = yes))
      (kb (object ?x) (isa = ?c))
      (not (eql '?c 'abstract))))
:ptrn (sync (action point ?x)
        (p2 ?x1 ?p1)
        (action pop-pointing))
:sgoal ((?x1 . ((object ?x) (scat = n2))
        (?p1 . ((postposition ?p) ())))))
```

このオペレータは、対象を意味とするような後置詞句を生成する際に、大域パラメータとして同期が指定されており、その対象が抽象物でなければ、まず、その対象への指差しを始め、次にその対象の言語表現を実現し、最後にその対象への指差しを終えるということを指定している。模擬動作と異なり、文中の役割にかかわらず指差し動作が同期する。ここでの指差し動

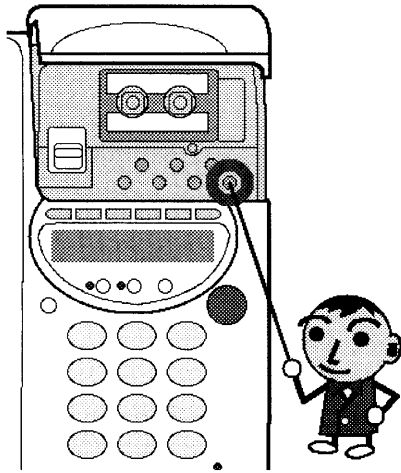


図5 説明者による指示動作(2)

作は図5に示すような指示棒と腕の移動変形によるもので、前章での説明者の移動と異なり、瞬時に開始終了することができ、同期を容易にしている。また、対象を表現する名詞句でなく、後置詞句の前後で指示動作の開始終了を行っているのは、音声合成における韻律の自然さを重視したためである。

この二つのオペレータを用いて生成される、動作が同期した音声言語の例を図6に示す。この発話はカセットカバーを開けることを依頼するもので、動作の対象であるカセットカバーを表現する後置詞句が発声されている間、そこへの指差し動作が行われ、行為の表現が終わると同時に模擬動作が実施される。

指差し動作は同時に一箇所に対してしか行わないようになっており、指示対象はスタックで管理され、埋

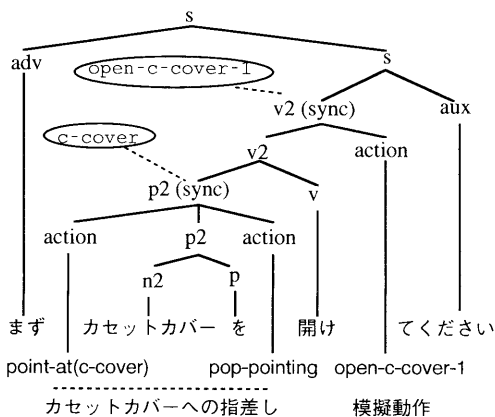


図6 音声言語と各種動作の同期

め込まれる場合がある。このような場合、現在の指示動作の終了というのは、埋め込まれていた指示対象への指示を再開することである。この仕組みは図7で示すような場合に有効である。この例では、応答1ボタンを同定させるためにその位置が述べられているが、その位置が液晶ディスプレイとの相対位置によって指定されている。指示対象をスタックを用いて管理することで、このような場合も、液晶ディスプレイに言及している間は液晶ディスプレイを指示し、その後、応答1ボタンを指示するという自然な同期が実現される。ただし、このような処理は言語依存である可能性がある。少なくとも日本語においては常に主辞が最右に置かれるため、このような指差し動作は全く自然なものである。

## 5 今後の課題

今後の課題として、第一に考えなければいけないのは、システム全体としての文脈の管理である。現在は、音声言語による対話的説明の中で、説明とその結果を文脈として、それを反映した説明の生成を行っているが、それ以上の広い範囲に渡る文脈やユーザのモデルは管理していない。例えば、音声による説明を行なった後でも、表示されるテキストは以前と同じであるし、そこへの問い返し質問の応答も変化しない。テキストや図面や音声言語というような様々なモードで情報が提供されるシステムで、全体的な文脈情報をどのように管理し、その時々で、どのような説明対話がユーザにとって適当かを今後考えていかなければならない。この検討は、同時に、現在ユーザからみて関連の弱い2種類の対話的説明を統合することにも繋がると思われる。

更に、様々な動作間の協調という問題がある。ここまでは明示的に述べなかったが、本システムでは3種類の動作が利用されている。第一は、対象への指差し動作、第二は、カセットカバーを開ける等の模擬的動作、第三は、図面のズームやパンなどの修辭的動作である。音声言語による手続き説明では前者2つのみが利用されている。問い返し質問への回答では、一応3種類の動作が協調しているが、そこでの利用はまだまだアドホックである。また、修辭的動作については、[FeinerLitman&etal93]でもある程度の検討がなされているが、体系的なものではない。これら、3種類の動作を適切に組み合わせ、時間的に協調した説明を生成するための規範を明らかにしていく必要がある。

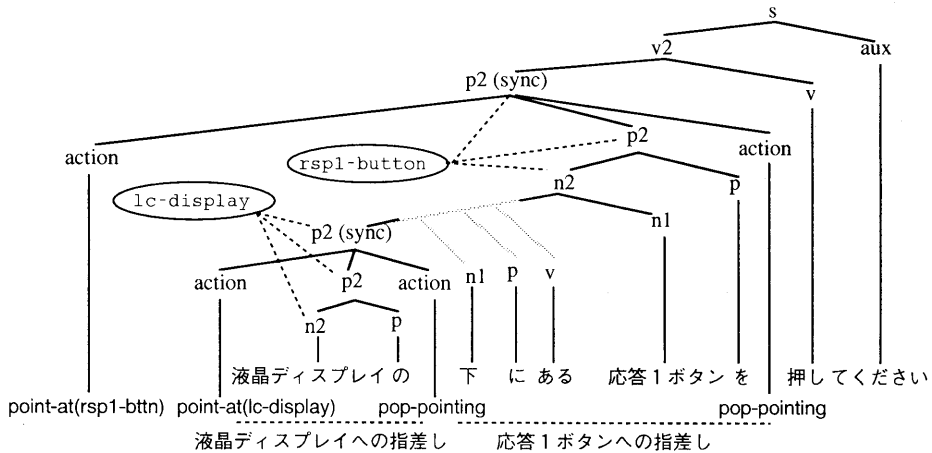


図7 複数の指示動作の同期

参考文献

[André&Rist93] André, E. and Rist, T. "The Design of Illustrated Documents as a Planning Task" in "Intelligent Multi Media Interfaces" eds. Maybury, M.T. The AAAI Presss / The MIT Press 1993 pp. 94 - 116

[Feiner&McKeown90] Feiner, S. K. and McKeown, K. R. "Coordinating Text and Graphics in Explanation Generation" in Procs. of AAAI -90 pp. 442 - 449 1990

[FeinerLitman&etal93] Feiner, S.K., Litman, D.J., McKeown, K.R. and Passonneau, R.J. "Towards Coordinated Temporal Multimedia Presentations" in "Intelligent Multi Media Interfaces" eds. Maybury, M.T. The AAAI Presss / The MIT Press 1993 pp. 139 - 147

[Hovy93] Hovy, E. H. "Automated Discourse Generation Using Discourse Structure Relations" in Artificial Intelligence 63 (1993) pp. 341 - 385 also in "Natural Language Processing" eds. Pereira, F.C.N.and Grosz, B.J. 1994 MIT Press/ Elsevier pp. 341 - 385

[Jacobs85] Jacobs, P.S. "PHRED: A Generator for Natural Language Interfaces" in "Natural Language Generation Systems" eds. McDonald, D.D., Bolc, L. ch. 7 Springer-Verlag 1988

[Mann&Thompson87] Mann, W.C. & Thompson, S.A. "Rhetorical Structure Theory: Description and Construction of Text Structure" in "Natural Language

Generation: New Results in Artificial Intelligence, Psychology and Linguistics" Kempen, G. eds. 1987

[Maybury93] Maybury, M.T. "Planning Multimedia Explanations Using Communicative Acts" in "Intelligent Multi Media Interfaces" eds. Maybury, M.T. The AAAI Presss / The MIT Press 1993 pp. 60 - 74

[Moore&Swartout90] Moore, J. D. and Swartout, W. R. "Pointing: A Way Toward Explanation Dialogue" in Proc. of AAAI-90 pp. 457 - 464 1990

[Nakano&Kato95] Nakano, Y.I. and Kato, T. "Task Context Dependency of Explanation Strategy in Instruction Dialog" in AAAI-95 Fall Symposium Series Embodied Language and Action pp. 94 -100 1995

[Nagao&Takeuchi94] Nagao, K. and Takeuchi, A. "Speech Dialogue with Facial Displays: Multimodal Human-Computer Conversation" in Procs. of 32nd Annual Meeting of the ACL pp. 102 - 109 1994

[Neal&Shapiro91] Neal, J.G. and Shapiro, S.C. "Intelligent Multi-Media Technology" in "Intelligent User Interfaces" eds. Sullivan, J.W. and Tyler, S.W. ACM Press 1991 pp. 11 - 43

[Wahlster91] Wahlster, W. "User and Discourse Models for Multimodal Communication" in "Intelligent User Interfaces" eds. Sullivan, J.W. and Tyler, S.W. ACM Press 1991 pp. 46 - 67

[Wahlster93] Wahlster, W. "Planning Multimodal Discourse" Invited Talk in ACL '93 1993