

## マルチメディアプレゼンテーションの自動生成に向けて —自然言語生成からマルチメディア生成へ—

Towards Automated Generation of Multimedia Presentations by Tsuneaki KATO (NTT Information and Communication Systems Labs.).

加藤 恒 昭<sup>1</sup>

<sup>1</sup> NTT 情報通信研究所

### 1. はじめに

WWWのホームページに代表されるような、テキストとグラフィクス、音声とアニメーションを多様に組み合わせた、いわゆるマルチメディアプレゼンテーションが世間に溢れている。これらの作成を支援するさまざまなツール類の開発も盛んである。このような中で、作成の支援ということから一歩進んで、作成そのものを自動的に行うという試みがある。

達成したい目的、伝達すべき意図があるとき、たとえばある内容を読み手や聞き手に伝え、それを理解してもらいたいとき、達成すべき目的と言語表現との関係に関する一般的な知識を利用して、そのための文章を自動的に生成しようという試みが自然言語処理研究における文章生成の研究である<sup>1)</sup>。グラフィクスにしろ、アニメーションにしろ、その背後になんからの達成したい目的、伝達すべき意図があるのであれば、同じ枠組みでそれらの自動生成が可能ではないか、そしてそこに共通の枠組みが存在するのであれば、それらと言語とが有機的に組み合わせさせたマルチメディアプレゼンテーションを自動的に生成することが可能ではないかというのが着眼である。

本稿では、言語(テキストや音声)とそのほかのメディア<sup>2)</sup>を組み合わせたマルチメディアプレゼンテーションの自動生成について、最近の研究動向を報告する。まず、自動生成が有効であるのは

どのような場面であるかを考察する。次に自動生成技術の鍵となるメディア間の協調(coordination)についてその必要性を説明し、その実現のための代表的な方式を概説する。その後、現在提案されているシステムを紹介する。最後に今後を展望してまとめる。

### 2. 自動生成の利点と適用領域

文章の目的がさまざまであるのと同様、マルチメディアプレゼンテーションの目的もさまざまである。詩や小説が与えるような芸術的感動を与えることを目的とするプレゼンテーションを自動的に生成する工学的手法というものは多分存在しない。一方、事実を無駄なく分かりやすく伝えることや、論拠をともなった理性的な主張を行うことが目的であるのなら、その達成の仕方には何らかの規則性もしくは指針があると思われるし、それをういた自動生成も可能性のあることとなる。たとえば、装置の使い方を説明するマニュアルや会社の収支状況を報告するレポートなどは自動生成の適当な候補となる。とくに、これらでは装置のスイッチなどを示す図面類や数値データを集計したグラフ類が頻繁に用いられるので、さまざまなメディアの積極的な利用という点からもマルチメディアプレゼンテーション自動生成の有望な適用先となっている。

個々のプレゼンテーションを手作りすることと比べて自動生成に価値があるためには、その枠組みがさまざまな伝達内容に対して適用可能であること、つまり枠組みの汎用性が必須である。対象となる装置が変わっても、今年レポートでも来年のレポートでも同じ枠組みで生成できることに

<sup>1)</sup> 本稿では誤解を招かない範囲でメディアとそのメディアを用いた表現であるメディアオブジェクトとを区別せずに扱う。また、メディア、モード、モダリティなどの用語についても厳密な区別をしない。

よって、必要となるたびにプレゼンテーションを作成することに対して優位となる。このときのもう1つの利点は、その枠組みが完成した後であれば、ある伝達内容が与えられてからそのプレゼンテーションが得られるまでのターンアラウンドがきわめて短くなることである。このため、道路情報や気象情報など、刻々と変化する伝達内容に対する緊急性の高いプレゼンテーションの生成はとくに重要な応用先として期待される。

伝達内容の変化だけでなくそれ以外の状況の変化に対しても適応的なプレゼンテーションが行えることが自動生成の意義を高めることになる。たとえば、それをみる相手が初心者であるか専門家であるかに応じて、利用できる資源(メディアの種類やその量)に応じて、さらに概要を知らせたいのか詳細が必要かという目的に応じて、内容と形式とを変化させた生成ができることが重要である。また、対話的な場面を考えると、相手の応答やこれまでの経緯に応じてさまざまなプレゼンテーションを行う必要がでてくる。このように直面する状況があらかじめ数え上げられないような場合は、自動生成が大きな重要性をもつことになる。

### 3. メディア間の関係と協調

マルチメディアプレゼンテーションが伝達意図や状況に照らして適切であるためには、各メディアがその特性を活かした適切なものであることが重要であるが、それに加えて、プレゼンテーション全体が関連のないメディアの寄せ集めではなく、それぞれがお互いに協調して伝達意図に貢献していることが必要である。メディアが適切に協調しているかは、以下の3つの関係に照らして考えなければならない<sup>2)</sup>。

**内容の関係** たとえば、グラフとその説明テキストのように複数のメディアが同じデータに依存していることによって生じるメディア間の関係をいう。自動生成の場合、与えられた伝達意図に基づいて各メディアの内容が決定されるわけであるから、この関係は自動的に成立するわけであるが、それが認知できるようになっていることが必要である。さらに、ある伝達意図を達成するために、メディアの特性に応じた適切な役割分担がなされているか、各メディアで伝達される情報に不足がないか、必要以上の冗長さが無いか問題となる。

**空間の関係** いわゆるレイアウトに関する問題であるが、自動生成においては、伝達すべき情報の分節の仕方、まとめ方の問題として捉えられる。たとえば、ある情報を1枚のグラフで伝達するか、それとも複数枚を用いるのか、その場合に順序をどうするかということは、ほかのメディアにおける分節の仕方や表現の順序に影響を与える。各メディアにおける分節の間に適切な関係が成り立っていることが必要である。

**時間の関係** いわゆる同期の問題で、音声、点滅などの動的な視覚効果、アニメーションなどの時間メディアを含んだプレゼンテーションでは常に問題となる。各メディアには、それぞれ推奨される情報伝達の順序があり、それらの調和をとることが必要である。

以下、メディアはそれぞれ固有の特性をもつことと、それに配慮しこれらの関係を成立させる必要があることを簡単な例を用いて考えてみる。

(1) 録音された用件を聞くためには、本体右側中央にある白い留守ボタンを押します。

この文は留守番電話機の取り扱い説明の一節で、このテキスト単独で適切なプレゼンテーションとなっている。これによって、(a) ボタン B1 を押すという行為が録音された用件を聞くという目的を達成するための手段であること、もしくは、録音された用件を聞けることがボタン B1 を押すという行為の効果であることと、(b) ボタン B1 が本体右側中央にある白い「留守ボタン」という名前のボタンであること、が伝えられている。ここで、(b) の情報は、ボタン B1 がどれであるかを相手に同定させるための表現である参照表現の意味内容となっており、そのために必要十分なものが選択されている<sup>2)</sup>。

メディアの特性ということで、同じ内容をテキストではなく音声で伝達する場合を考える。このときは、以下の(2)に示した表現の方が自然であると感じる人が多いと思う。少なくとも、音声対話では、参照表現を名詞句としないで、(2)のように節とする場合が多いことが報告されている。このように同じ意味内容を伝達する場合でもメディアによって適切な表現というものは異なるので

<sup>2)</sup> 厳密には、B1 の名前が「留守ボタン」であることは相手も B1 を同定する助けにはならない。これは今後の参照のためのラベル付けである。

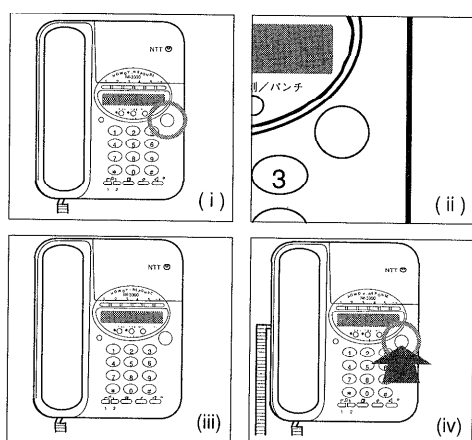


図-1 図面による留守ボタンの参照

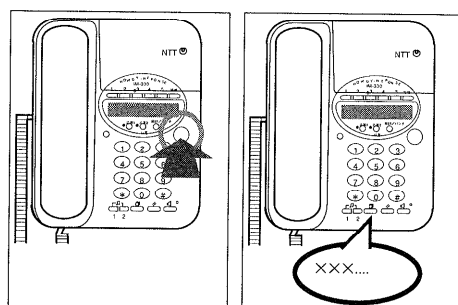


図-2 図面による行為-結果の関係の表現

ある。

(2) 録音された用件を聞くためには、本体右側中央に白い留守ボタンというボタンがありますので、それを押します。

(3) は、(b) が伝達されていないので、そのみでは不適切な表現である<sup>☆3</sup>。(3) が適切であるためには、(b) の伝達によって達成される目的、つまり B1 の参照をほかのメディアによって分担させなければならない。たとえば、図-1(i) を添えることで適切なプレゼンテーションとなる。言い換えれば、図-1(i) は、B1 の参照のための適切な表現となっている。この場合、テキストと図面とで適切な役割分担が行われ、内容の関係が成り立つことになる。

(3) 録音された用件を聞くためには、留守ボタンを押します。

ちなみに、テキスト同様、図面にもその目的に照らして適切なものとそうでないものがある。

☆3 この表現の以前に B1 のラベル付けが完了していれば問題はない。逆にそのような場合は (1) が冗長で不適切となる。

たとえばこの場合、図-1(ii)、図-1(iii)は適切とは言いがたい。図面を用いてある対象を参照するためには、その対象が描かれているだけでなく、その位置などを明らかにする背景も描かれており((ii)にはこれが欠けてる)、さらにその図面中での対象物が注目されているものであるかが明らかになっていなければならない((iii)ではこれが行われていない。(i)では、サークリングによってそれが明らかにされている)。

(3)と図-1(i)の組合せでは、テキストによって(a)が伝達され、図面によって(b)の伝達と同様の目的が達成され、しかもその間の関係が理解できるようになっている。この例では、2つのメディアに対して重複なく役割分担が行われている。一方で、(1)と図-1(i)の組合せを考えると、テキストによる(b)の伝達と図面とが同じ役割をもっているために冗長な印象を受ける。ただし、重複による冗長さは常に排除されるべきものではない。たとえば、図-1(iv)では、矢印というメタオブジェクトを導入することで「押す」という動作を図面上でも表現しているが、これを(3)と組み合わせても違和感はない。

メディアはさまざまな特性をもち、それによって伝達できる情報も異なる。たとえば、図面では参照と述定の区別が困難なので「留守ボタンは本体右側中央にある白いボタンです」と「本体右側中央にある白いボタンが留守ボタンです」というテキストでの違いを表現することはまず不可能である。また、目的-手段関係のような抽象的な関係を表現することも難しいため、(a)の伝達を図面のみで行うことは簡単ではない。ところが、行為-結果の関係は図面によって表現する、もしくは暗示することができる。図-2のように2枚の図面を並べればよい。興味深いのは、この図面にもなるテキストとして、(3)よりも(4)の方がふさわしいことである。これが空間の関係で、あるメディアでの情報の捉え方、その分節の仕方、空間的配置が、ほかのメディアの適切性を左右するのである。

(4)留守ボタンを押すと、録音された用件が聞こえます。

あるメディアの空間的配置など、内容というより表層表現にかかわる要因がほかのメディアでの伝達内容に影響を与えること、つまり、マルチメ

ディーププレゼンテーションの生成において内容と表現とが密接に絡み合い、内容の関係や空間の関係を成立させることの例が、(5)と図-1(i)の組合せである。このように対象そのものの特徴でなく、ほかのメディアでの対象の表現を利用して、それを参照することを cross-modal 参照と呼ぶ。

(5)録音された用件を聞くためには、図面中央に示された留守ボタンを押します。

最後に、時間の関係を示すために(6)を考える。これはテキストではなく、(2)と同様、時間メディアである音声によって伝達されるとする。

(6)録音された用件を聞くためには、本体、右側中央の、この留守ボタンを押します。

この音声では、「この」という表現からも察せられるように、動作との同期が期待される。この表現は、一種の cross-modal 参照で、それが発声された時点で言及されている対象が指し示されていることを伝えている。(6)の音声に人間が動作をとまわせるとすれば、およそ以下のようなになる。「本体」と発声する時点で、電話機本体に軽く手が添えられる。「右側中央の」で手が動き出し指差し動作へと変わっていき、「この」の時点で留守ボタンが指差されている。指差しは音声を終了するまで続けられる。これらの指差し動作と音声との対応付けを明確にするために「本体」の後と「この」の前には短いポーズがおかれる。人間の指差し動作を模したアニメーションが音声に同期するプレゼンテーションであれば、まさにこのような時間の関係が必要となろう。

#### 4. 生成と協調の機構

前章でみたようなメディア間の協調を実現するためには、マルチメディアプレゼンテーションを生成するためのさまざまな機能が有機的に連携することが必要となる。必要な機能は以下のようにまとめることができる。

**内容企画** 与えられた伝達内容・伝達意図から、プレゼンテーションに含める意味内容を具体化して決定する。前章の例では、留守ボタンを相手に同定させる必要があること、そのためにはその位置や形状を伝える(言語的視覚的を問わず)必要があることなどを決定する。

**メディア協調** 企画された意味内容をどのメディアによって伝達するかを決定し、各メディア生

成(後述)にその生成を割り当てたり、生成されたメディアの表現の間の調和を図るための交渉を行ったりする。留守ボタンの参照をどのメディアで行うか、両方のメディアで表現されている留守ボタンをどのように関連付けるかを決定し、それを達成するよう、各メディア生成に表現の修正などを依頼する。

**メディア生成** 与えられた意味内容から、それを表現するテキストなり画像なりを生成する。一般にはメディアごとに生成器が設計される。つまり、テキスト生成器、音声生成器、グラフィクス生成器などが、それぞれのメディアの生成を行う。留守ボタンを同定させるための図面を生成したり、留守ボタンを押すことで録音された用件を聞くことができるという意味内容を表現するテキストを生成したりする。

これらの機能は、それぞれ独立したモジュールであることもあれば、1つのモジュールが複数の機能をもつ場合もある。これらの機能をどのように実現し関係させるかによってどのような協調が実現できるかが決まる。以下、いくつかの例を紹介する。

**割当に基づく協調** 上記の機能をそれぞれモジュールとし、内容企画部が決定した伝達内容をメディア協調部が受け取り、それをそれぞれのメディア生成器に割り当てるという方法である。各メディア生成器はその割当に応じた生成を行うことのみを考え、その後の通信は行わない。メディア協調部は単に割当を行うだけでなく、内容企画部が決定した意味内容の構造をプレゼンテーションの構造を考慮して再構成する場合もある。洗練されたメディア協調の枠組みに基づいて割当が行われれば、この手法は内容の関係を保った十分整合的なプレゼンテーションを生成する。ただし、あるメディアの表層表現に呼応するような協調は不可能である。

**表層に基づく協調** 割当に基づく協調とは対照的に、1つの主導的なメディアの表層表現もしくはそれへの意味的な注釈によってほかのメディアの生成を制御する方法である。たとえば、図面や地図に示された対象物を音声で次々と説明していくような状況において、「まず国道X号を北上して〇〇市に向かいます。そこで…」という発声とともに、地図中の対応する道路や街を点滅さ

せる。この方法はあるメディアの表層表現に密接にかかわる協調の実現に有効である。ただし、情報の流れが一方的なので主導的なメディアの生成に一定の配慮をしないと、たとえば、あちこちの対象物がばらばらと点滅したり、ある対象物が点滅している時間が短くなりすぎるようなことが生じうる。

**黒板モデルによる協調** 内容企画によって得られた伝達すべき意味内容を共通のデータ構造とし、それを媒介としてメディア生成やメディア協調が動作する枠組みである。メディア割当は得られた内容に対して、どの部分をどのメディアで表現するかを注釈付けることで行われる。各メディア生成器は自分が生成するようにと注釈された部分の生成を行うが、このとき、割当に基づく協調の場合と異なり、自分が生成すべき部分だけでなく、それを包む文脈(その表現の目的や位置付け)も参照できるので、より適切な表現が選択できる。さらに、それぞれのメディアの生成の過程、たとえばテキスト生成器がどの部分を一文として生成したかなどを注釈付けていくことによって、それによって新たに生じる制約、たとえば、その分節に応じた生成をしなければならないという制約をほかのメディア生成器に伝播することができる。

**漸進的プランニングによる協調** あるメディアの表現に関する制約が、表現方法だけでなく、意味内容にも影響を及ぼすことがある。部品の説明を図面に含めようとした場合、そこにどれだけの内容が記述できるかは図面の空き領域に制約される。加えて図面中でどれだけの内容が伝えられたかはそれにとまなうテキストで伝えるべき内容に影響する。このような問題を真剣に考えると、内容企画、メディア協調、メディア生成のすべてを同時並行的に行い、さまざまな決定やその決定の結果を細かい単位でやりとりし、その影響に対処するという漸進的な枠組みが必要となる。内容企画とメディア協調をまとめてプランニング過程として実現し、これを漸進的に制御することでこの枠組みが実現できる。このプランニングでは、内容決定やメディア割当のレシピを参照して目的を達成するための行為を徐々に詳細化していく。メディア生成はその結果を順次受け取り、漸進的に文や図面の生成を試み、その結果を返す。細かい単位でのバックトラックが繰り返され、適切なプ

レゼンテーションが生成されていく。

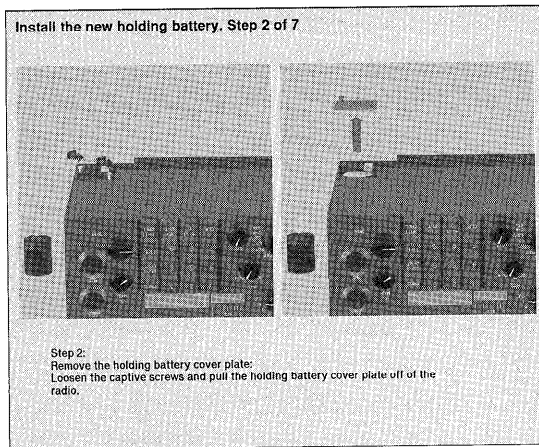
**複数候補生成に基づく協調** バックトラックを用いることなく各メディア生成間の調整を図るもので、先に複数の候補をあげて、それらの比較検討を通じて最適なものを選択するという枠組みである。各メディア生成器は、一般に同じ意味内容に対して複数の表現を生成することができ、メディアに閉じた適切性の基準をもっている。たとえば時間に関する制約であれば、音声は短く簡潔な方がよい、グラフィクス中の項目の説明順序は左から右というような空間的順序に一致するのがよい、などである。各メディア生成器が自分に割り当てられた意味内容の複数の表現(もしくは一部だけが決定された表現)を、それぞれを採用した場合の時間的制約と優先順位とともに提案し、すべてのメディアが提案する制約を満たし、全体として最も高い優先順位となるような表現を選ぶことによってメディア間の協調が図られる。

## 5. いくつかのシステム

前章で述べたような協調機構を用いて、さまざまなマルチメディアプレゼンテーションの自動生成が試みられている。本章ではそのうちの代表的ないくつかを紹介する<sup>\*)</sup>。

**テキストとグラフや表の協調** 組織の運営状況や製品の販売状況についてのレポートを作成するシステムが、Roth や Arens らによって報告されている。たとえば、今年度の営業部の支出はなぜ予想を上回ったのかについて、テキストとグラフを用いて説明する。どの情報にテキストを用い、どれに図表類を用いるかという判断を割当に基づく協調で実現している。メディア生成では、ある情報の伝達にどんなグラフや表を用いるかを判断する。このとき、時空間の点を表現する場合は量を含意してしまう棒グラフなどは使えないというような表現する内容そのものの特性と、何のためにそれを伝達したいのかという作り手側の意図、たとえば、量を正確に伝えたいのであれば数値を示す表が適しているが、相互比較が目的であればグラフの方がよいなどが考慮される。割当に基づく協調だけでは対処できない問題もある。たとえば、1枚の図表で複数の情報を伝えることができ

<sup>\*)</sup> さらに詳しく知りたい方は、文献3),4)を参照いただきたい。



文献6)より引用

図-3 COMETによるプレゼンテーション

るが、どの情報を含めるかの決定にはテキストとの関係が考慮されなければならない。たとえば、テキストを読んでいって、図-1に関連した記述があり、次に図-2に関連した記述があり、その後また図-1に関連した記述があるのでは空間の関係が適切であるとは言い難い。また、1枚の図表に限っても、図表中では名前順に並べられている組織がテキスト中では支出の多い順に説明されるようでは問題である。これらに対処するためにメディア生成器間での情報交換が行われる。

図表を生成するのではなく、すでにある図表について説明するシステムもある。XTRAは、税金申告の書式について、「ここには年収を記入します」というような記入欄への指差し動作をともなった説明テキストを生成する。

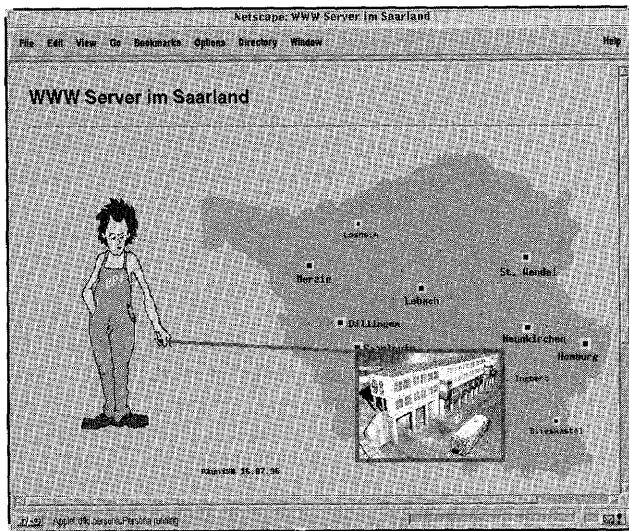
テキストと図面の協調 テキストと図面の組合せでは、装置の操作説明を題材としたシステムが提案されている。WIP<sup>5)</sup>は、エスプレッソマシンの取り扱いを対象としたシステムであり、その後、方式の汎用性を示すためにモデムや芝刈り機の説明にも適用されている。COMET<sup>6)</sup>は、軍用無線機の取り扱いや維持管理の方法を説明する(図-3)。この2つは、テキストと図面との協調に関してさまざまな問題を提起し、それらへの回答を与えている。COMETは基本的に黒板モデルによる協調を用いているが、cross-modal参照を生成するために、一部、内容企画へのバックトラックを行っている。WIPは、漸進的プランニングによる協調を提案している。この漸進性はテキストや図面の表層生成にまで達しており、伝達す

べき内容が揃う前から、プランニングが始められ、テキストや図面の出力が開始される。

これらのシステムにおいては、テキストだけでなく、図面についても意図に基づいた、部品からの合成が行われる。つまり、ちょうど、テキストが与えられた意図を伝達するために単語から文法規則を用いて合成されていくように、図面も単に選ばれるのではなく、対象のワイヤフレームモデルに対して構図を決定し、矢印などのメタオブジェクトを加え、名称などの注釈を付与するというような操作の組合せによって生成される。これらは、ある作図操作の選択と、それによって変化した図面が蓄積された制約を満たすかの検査とを繰り返すことで進められる。たとえば、ある部品にその名称を注釈付ける場合、まずその近傍にテキストを表示するという手段が選択され、そのような手段をとった結果、その図面で示される必要のあるほかの部品が隠れてしまわないか、つまり必要な部品がはっきりみえるという制約に違反していないかが検査される。もし検査の結果に問題があるようであれば、別の注釈手段、たとえば、図面の空白部分に名称を示し、そこから部品へと矢印を引くなどの代替手段が検討される。

時間メディアの協調 音声、点滅などの動的な視覚効果、アニメーションなどの時間メディアを扱った初期のシステムは、表層に基づく協調を行うものがほとんどで、たとえば、道案内の音声と同期して、表示された地図の中で現在説明している道路やランドマークを点滅させたり、領土内に配置された軍事施設の説明を音声と地図、さらには表を用いて行うというのが、MayburyやNealらによって提案されている。

最近のシステムでは、漸進的プランニングや複数候補生成によるより緊密な協調が行われている。これらのシステムでは音声の中のある単語の発声、図面の表示、説明項目への指差し動作、説明項目の点滅などの視覚効果の実施等々をイベントとしてとらえ、それらイベント間の時間的制約を調整する形で協調が行われる。この時間的制約はAllen流の時間論理<sup>7)</sup>を用いて記述されることが多い。これは2つのイベントの時間関係を13種類の述語で表現するもので、たとえば、イベントxが終了した瞬間にイベントyが始まる関係をx meets y、イベントyが始まった後にイベント



文献9)より引用

図-4 WWW環境での擬人化エージェントPPP



文献11)より引用

図-5 Cassellらによる擬人化エージェントのジェスチャ

xが始まり、yが終わる前にxが終わる関係を x during y と表現する。音声中の連続する単語は (近似的に) meets の関係をもって発声されるし、指差し動作と指差される対象の表現の発声は during の関係となる<sup>★5</sup>。このようなイベント間の制約を重ね合わせて、全体の順序が決定される。なお、Allen 流の時間論理では、量的な制約、たとえば、指差し動作は3秒以上続けなければいけないなどが表現できないため、この点を拡張した制約解消機構が使われることもある。

★5 ただし、人間の行う教示説明でこれが成り立っているかには疑問がある。

このようなシステムの例である PPP<sup>8),9)</sup>は、WWW環境中でさまざまなプレゼンテーションを指差し動作と音声とで行う擬人化エージェントである(図-4)。MAGIC<sup>10)</sup>は、テキストによる注釈を含むグラフィクスを表示し、そこに描かれた項目をハイライトするなどの視覚的効果と音声とを組み合わせ、医療データベースの情報から患者の生体情報や病歴などのレポートを生成する。

擬人化エージェント 前出の PPP もそうであるが、人間のように手足と顔をもった擬人化エージェントは典型的なマルチメディアプレゼンテーションであり、音声や視線や表情やジェスチャを協調させる必要がある。

PPPはジェスチャとして主に指差しに関するものを扱っているが、Cassellらの擬人化エージェントは、それ以外のさまざまなジェスチャを生成する<sup>11)</sup>。動作の選択は音声への注釈に一定の規則を適用することで行われる。注釈はそこで伝達される情報の新旧、つまり今まで話題になった情報であるか否かとそれが何を指しているか、たとえば具体的な対象物なのか行為なのかからなる。これらに基づいて、新情報に対してのみジェスチャがともない、その指示内容が具体物であればその形状を表現するアイコン的な動作が、具体的な身体行為であれば、そのパントマイム的な動作が行われる。図-5にその様子を示すが、左上は小切手を表現し、左下はサインするという行為を表現している。右の2つは申し出をする、待つなど、抽象的な行為にともなうジェスチャである。これらは表層に基づく協調であるといえるが、彼らの最近の研究では、音声とジェスチャの役割分担についてより詳しい研究がすすめられている。たとえば、「彼が公園に歩いて行った」ことを伝えるために、音声では「行った」とだけ伝達し、人差し指と中指を下に向けて交互に動かすジェスチャをともなわせるなどの事例の収集と検討が行われている<sup>12)</sup>。

## 6. おわりに

マルチメディアプレゼンテーションの自動生成について、とくに、言語とそのほかのメディアとを組み合わせる場合のメディア間の協調についての研究動向を述べた。本稿では触れられなかったが、音声とさまざまな動作とを組み合わせることで、対話するマルチモーダル入力に関する研究も活発となっている。これらの技術と組み合わせることで、対話的なマルチメディアプレゼンテーションの生成、マルチモーダル対話システムの構築へとという展開が期待できる。また、マルチメディア一般の研究においても、同期の記述についてのモデルなど、さまざまな提案がなされている。これらの成果を積極的に取り入れていくことも重要である。さらに、何がよいプレゼンテーションであるかという問題も完全には解決していない。この点についてはメディアの特徴やコミュニケーションの方略についてのより深い洞察が求められていくだろう。

謝辞 本稿の図-3、図-4、図-5 はそれぞれ、Dr. S. K. Feiner & Dr. K. R. McKeown, Dr. E. André, Dr. J. Cassell のご厚意により掲載させていただきました。ここに感謝いたします。また、本稿の執筆に際して貴重なコメントをいただいた本誌編集委員、巖寺俊哲氏、中野有紀子氏、浅野久子氏に感謝いたします。

## 参 考 文 献

- 1) Hovy, E. H. : Automated Discourse Generation Using Discourse Structure Relations, *Artificial Intelligence*, 63, pp.341-385(1993).
- 2) Steinmetz, R. and Nahrstedt, K. : *Multimedia: Computing, Communications & Applications*, Prentice Hall(1995).
- 3) Sullivan, J.W. and Tyler, S.W. (eds.) : *Intelligent User Interfaces*, ACM press(1991).
- 4) Maybury, M.T. (eds.) : *Intelligent Multimedia Interfaces*, The AAAI Press / The MIT Press (1993).

- 5) Wahlster, W., André, E., Finker, W. and et al. : *Plan-based Integration of Natural Language and Graphics Generation*, *Artificial Intelligence*, 63, pp.387-427(1993).
- 6) Feiner, S.K. and McKeown, K.R. : *Automating the Generation of Coordinated Multimedia Explanations*, *IEEE Computer*, Vol. 24, No. 10, pp.33-41(1991).
- 7) Allen, J. F. : *Towards a General Theory of Action and Time*, *Artificial Intelligence*, Vol. 23, No. 2, pp.123-154(1984).
- 8) André, E. and Rist, T.: *Coping with Temporal Constraints in Multimedia Presentation Planning*, *Procs. of AAAI-96*, Vol. 1, pp.142-147(1996).
- 9) André, E. and Rist, T.: *Planning Referential Acts for Animated Presentation Agents*, *Procs. of ACL/EACL-97 Workshop on Referring Phenomena in a Multimedia Context and their Computational Treatment*, pp.67-72(1997).
- 10) Dalal, M., Feiner, S., McKeown, K. and et al.: *Negotiation for Automated Generation of Temporal Multimedia Presentations*, *Proc. of ACM Multimedia 96*, pp.55-64(1996).
- 11) Cassell, J., Pelachaud, C., Badler, N. and et al.: *Animated Conversation: Rule-based Generation of Facial Expression, Gesture & Spoken Intonation for Multiple Conversational Agents*, *SIGGRAPH 94*, pp.413-420(1994).
- 12) Cassell, J. and Prevost, S. : *Distribution of Semantic Features Across Speech & Gesture by Human and Machines*, *Procs. of The Workshop on the Integration of Gesture in Language and Speech*, pp.253-269(1996).

(平成9年11月5日受付)



加藤 恒昭 (正会員)

1959年生。1981年東京工業大学電気電子工学科卒業。1983年同大学総合理工学研究科電子システム専攻修士課程修了。同年日本電信電話公社(現NTT)横須賀電気通信研究所に入所。自然言語処理、対話処理、マルチモーダルコミュニケーションに関する研究に従事。現在NTT情報通信研究所知的通信処理研究部主幹研究員。工学博士。電子情報通信学会、人工知能学会、言語処理学会、ACL各会員。