

## 解 説



## 音声言語情報処理の現状と研究課題

## 4. 協調から協応へ

## —自然な発話に対する新たなアプローチ—

岡田 美智男†

## 1. はじめに

「留守番電話であると気づいたとき、おもわず語りかけを諦めてしまう。そして受話器をそっと置いてしまう。いったい何を伝えようとしていたのだろうか。本当に話したかったことは、相手とおしゃべりの中から生まれてくることなのかもしれない。」<sup>1)</sup>

日常でのコミュニケーションを成り立たせている基本的な原理は何だろうか。情報理論や符号化技術の影響もあって、我々のコミュニケーションは無意識のうちに「コードモデル」の枠組みで捉えられていることが多い。それは、「お互いがコードの体系(モデル)を共有し、メッセージとしてエンコードされたコードを、一方ではそのモデルを利用してデコードしながら解釈する」という枠組みである。これは多くの音声認識や対話理解の手法にも反映されている。

我々が無意識的に用いている個々の発話には、その背後に何らかの意図や動機が反映されている。あるいは、その発話意図を達成するために何らかの計画立案がなされ、発話が形作られているのだという。聞き手の心的状態をいかにしたら自分の抱く目標状態に推移させることができるのか。相手の心を読み、ゴールを達成するための前提を調べあげる。その前提をクリアするために、サブゴールを潰していく。

この発話生成の枠組みはそのまま、話し手の発話のプランを理解したり、その背後にある意図を推定する問題の裏返しにもなっている。話し手がある意図を背景に、いくつかのサブゴールに展開しながら、合理的にそれを実現していると仮定す

れば、振舞いの途中の段階でそのプランの構造が見え、次に解決すべきサブゴールも予想がつく。次第に、話し手の発話の意図がおぼろげに見えてくる。相手の発話意図を探り当てるには、発話における計画立案の過程を試行錯誤しながら遡っていく。これは、いわゆるプラン推論と呼ばれるプロセスである<sup>4)</sup>。

コンピュータがユーザに対して協調的に振舞うためにも、我々の断片的な振舞いの中から、プランの構造を推測し、その背後にある振舞いの意図を探る。その意図に沿って、必要にして十分な情報を提供する。こうした振舞いが「協調的である」とされた。

ところがこうした原理に基づいて構築されたシステムと対話してみると、妙な気分になっていることに気づく。それは、協調的な振舞いを強いられるのはコンピュータではなく、我々自身なのではないかという感覚である。協調的な情報を提供してもらうためには、我々も心の準備をしながらコンピュータに向かう必要がある。そこでは合理的で手際のいい振舞いが求められる。これでは少しも気が抜けない。こうした「ユーザフレンドリな対話システム」に対して、我々はすこしも親しみを感じないのである。

発話をナビゲートしているものは何か、発話をレギュレートしているものは何か。我々がこれまで想定してきたモデルとの間に、何か本質的なミスマッチが存在するのではないだろうか。本稿では、「エコロジカルアプローチ」と呼ばれる視点からの、ことばや発話、コミュニケーションに対する新しい取組みの一端を紹介したい。

## 2. 発話をナビゲートするもの

我々が日常生活の中で何気なくおしゃべりをす。ここで用いられるような発話を「自然な発話

† From Cooperation to Coordination by Michio OKADA (ATR Media Integration & Communications Research Laboratories).

†† ATR 知能映像通信研究所

(spontaneous speech)』と呼び、新聞を声を出しながら読むような「朗読発話 (read speech)」と区別して考えている。普段、その違いはあまり意識されることはないのかもしれない。

### 2.1 プランとは何だろうか

計画立案や推論など、我々が日頃「慎重に行う行為」に関する研究は、1960年代のニューウェル (A. Newell), サイモン (H. Simon) らの物理記号仮説や GPS と呼ばれる問題解決のモデル化を契機としている。その流れは今日まで続いている。短期記憶や長期記憶などに代表されるように、「コンピュータの動作をメタファとして認知プロセスのモデル化を試みる」ところにその特徴がある。これは「情報処理アプローチ」とも呼ばれている。

発話そのものを行為と見なし、計画立案や問題解決の枠組みで捉える研究もその流れにある<sup>8)</sup>。これらは「熟考的なパラダイム」とも呼ばれるように、事前に外界の状況を把握して、計画立案によって最適な動作系列を得る。それから、この動作系列 (プラン) に従っておもむろに計画を遂行していく。このような慎重過ぎる行動は、間接的であり何かはがゆい。

我々の日頃の行動を考えてみるまでもなく、すべての行動がプランによってナビゲートされているわけではない。むしろ、我々は行き当たりばつたりの生活の中で、大過なくすごしているようだ。サッチマン (L. Suchman) らは、「プランというのは、ある前提条件と行為の結果とをもっともらしくつないだ説明にすぎないのだ」という立場をとる<sup>9)</sup>。

### 2.2 説明のための後付け

サッカーの様子をビデオに収録し、巻き戻しながら、各々のパスの意味付けをする。あるいは、その背後にあった監督の思惑を推し量る。何か監督の思惑で各選手が動いているように見えてしまう。そしてあとから、その作戦の優劣を議論する。これはサッチマンのいう「説明のための後付け」ということになるだろう。

対話をテープに録音し、それをあとから書き起こしてみると、何らかの構造が見えてくる。そして、その背後に何らかの規則の存在を感じてしまう。そして「ことば」に対しても、我々はまずその構造に着目してしまう。ある規則に従って談話

が生成されている、あるいは制御されているような視点で捉えてしまう。その規則から外れたもの、逸脱したものを例外と捉えてしまう。これは「構造」というものに対するトップダウンな視点である。

### 2.3 行為をナビゲートするもの

日頃から理屈っぽい人ほど、車の教習などで苦労する。教習所の中での車庫入れや縦列駐車練習を思い浮かべてみよう。一度後ろの様子を確認すると、その後はひたすらハンドルの動きに対応した前輪の回転角を思い浮かべ、ハンドルを右に回すか左に回すべきかを考え込む。そして一度決断がつくと、その後はひたすらハンドルを回すことに専念する。仔犬がふらふらと近付いてきても「行為」の途中では修正が効かない。

ところがすこし慣れてくると、ハンドルの動きに合わせた「外界の見えの変化」によって、車庫入れの操作はいとも簡単に行えるようになる。このとき車を動かしながら、ハンドルを「いちかばちか」だが、どちらかに回していることに気づくだろう。ハンドルをとにかく動かせば、その動きが外界の見えの変化として直接感じることができ。そしてそれはハンドルを動かすという行為、あるいは目的方向に車を移動させるという行為そのものを誘導する情報になる。車の運転を覚えるコツは、この「情報」に素直に沿うことなのである。

この「行為者に対して環境が提供する行為の可能性についての情報」は、「アフォーダンス (affordance)」と呼ばれている<sup>3)</sup>。ハンドルを動かしながら、その動きにもなって変化する「見え」から、行為の可能性をピックアップし、我々はその「アフォーダンス」に沿って車を移動させているのである。

### 2.4 発話における漸次的精緻化

次の発話は、かつて利用したことのあるレストランの中にあつた「おっきい太い黒い柱」を思い出しながら、その雰囲気を変えようとしている会話の一部である。

● 柱、黒い、黒い柱が、おっきい太い黒い柱が、ぬっとでてい...

ラフなデッサンからはじめて、絵の内容を段階的に精緻化していくように、思い出しながら「柱」のイメージを精緻化する。ラフなデッサン

が次の詳細なイメージを作るための叩き台となる。これはことばでイメージを伝えあうときも同じであろう。ここで「発話を引き出しているものは何か」という視点が重要になる<sup>7)</sup>。

最終的に伝えようとしたことは、「おっきい黒い太い柱」のことであり、はじめの「柱、黒い...」は結果的に「言い直されてしまったもの」である。しかし、この発話を加速しているものはじめに引き出された「柱」という言葉であり、それに続く発話をナビゲートする。すなわち、はじめに引き出された「柱」という発話が次の「黒い柱」の想起を促す。さらに、この「黒い柱」という発話が「おっきい太い黒い柱」というイメージの想起を促している。まず動くこと、その動きによって得られる情報が次の行動を促していく、次の行動をナビゲートする。発話しながら言いたかったことが次第に明確になっていく。

これらの振舞いは、発話の内容を考えながら、それと同時並行に口を動かすような「漸次的な発話生成」の概念<sup>9)10)</sup>ではなく、むしろ「発話そのものを思考の道具に使う」といった感覚に近い<sup>11)</sup>。

はじめから綿密に計画されたプランを持たず、「行為の可能性」に引き出されるようにして発話を開始し、その最初の発話が新たな「行為の可能性」を用意する。あるいは、新たな環境を構成していく。あらかじめ用意されていたプランは、行動選択の際に参照されるリソースの1つにすぎない。このような視点と、「概念の想起やプランそのものが発話をナビゲートする」と考えるトップダウンからの視点とは、ちょうど逆の関係にある。発話をナビゲートするものが基本的に異なるのである。

### 2.5 監督の指示に気を取られていた

これまで「ことば」や対話のメカニズムに対する我々のアプローチは、「トップダウンな見方」にやや偏してきたともいえる。その意味で、サッカーの試合を観戦しているとき、我々は「監督の指示」だけに気を取られていたのかもしれない。そういう視点から発話生成のモデルが考えられてきた。そして、こうしたメカニズムにより発話が生成されることを前提に、発話を解釈するためのモデルが考えられてきた。具合の悪いことに、こうした視点に慣れ親しんでしまうと、「偏った見

方をしていること」にだれも気づかなくなってしまう。

日常的な発話の中で、ボトムアップに組織化される発話の役割、その組織化のメカニズムに対して、我々はおっと目を向けていく必要がある。そして、「柱、黒い、黒い柱...」という口ごもりながらのこうした発話は、ボトムアップに組織化する発話の挙動を調べる、その貴重な手がかりを与えているともいえる<sup>12)</sup>。

### 3. 発話をレギュレートするもの

我々は「ある秩序だった振舞い」を見たとき、2つの側面からの捉え方をしている。1つは、「協調」あるいは「調整」ということばに代表されるように、お互いが意識的に調整しあって、秩序を維持しているという見方である。一方、この「協調」、「調整」に対して、「お互いの相互作用の中から、結果として秩序が生じた」とのボトムアップな見方ができる。お互いに協調しあう、調整しあうとは考えない。むしろ振り返ってみると「協調している」かのように見える、あくまで事後的なもの捉え方である。ここではコーディネーションをボトムアップな視点で、「協応する」という立場から捉えていく。

#### 3.1 「言い誤り」「言い直し」をどう捉えるか

我々の日常の発話の中での「言い誤り」、「言い直し」はどうだろうか。自然な発話における非流暢性 (disfluencies) は、とかく常軌から逸脱したものと捉えられやすい。文法から「逸脱したもの (ill-formed sentence)」という視点から分析されがちである。たとえば「言い誤り」とは、発話意図と実際の発話とが乖離した行為のしそこない、すなわち「スリップ (slip)」の一種である。そして「言い直し」とは文字どおり、その「言い誤り」に対する修復操作であると分析されてきた。ここでも「秩序だった振舞いをどのような視点から捉えるのか」によって、そこで生じるスリップ (slip) や非流暢性 (disfluencies) の解釈が違ってくる。

たとえば、この「逸脱する」という言葉には、「ふつうはあまり逸脱することはない」というニュアンスがある。先験的に与えられた文法なり、言語的な構造に沿った形で発話が産出され、その発話意図が伝えられる。そのコード化の過程でた

またま誤った、逸脱したことと捉えられやすい。こうした言語観の下で、自然な発話を解析するシステムをデザインすると、この「文法からの逸脱」を修復するメカニズムを考えてしまう。

ところが、こうしたシステムに向かうとすこし緊張する。適当に聞き流してくれないことに気づく。こちらが言い直すとその言い直しを解析しようとして、いろいろなルールを適用しはじめる。困ったことにはシステムがしばらく沈黙してしまう。いつの間にか、「言い直す」ことを禁じられているような、いつも意識的な発話を強要されているような気持ちになる。

### 3.2 ベルンシュタイン・アプローチ

歩きながら本を読む、思わず口ごもる、相手の話に聞き耳をたてたり、勝手に聞き流す。いわゆる「情報処理アプローチ」では、こうした無意識になされる振舞い、高度にコーディネートした流暢な振舞いを説明するための十分な手段を持っていない。無意識に行えていることを意識的な手段で説明しようとしてしまう。我々の日常での振舞いや日常で使われる「自然な発話」をコンピュータで扱うことの本質的な難しさ、その要因は、実はこうしたところに潜んでいる。

我々が流暢に振る舞うとき、そのたびにその秩序だった運動の系列が記載された「運動プログラム」が呼び出され、筋肉はそのプログラムにそのまま従うだけなのだろうか。この「運動プログラム」のような行為に先だって、ある瞬間ごとの配置を指定する枠組みでは、制御すべき膨大な自由度に対して破綻をきたしてしまう。こうした「中枢による運動の制御」の問題点を最初に指摘したのは、ロシアの生理学者ベルンシュタインである<sup>13)</sup>。

ここでの問題は「こうしたシステム内部の膨大な自由度をどのようにして切り盛りするか、制御すべき独立なパラメータをいかにして減じるか」である。そして、この問題に対してベルンシュタインによって予見された仮説は「要素間での相互作用によって自由度を減じる協応メカニズム」の存在であり、これを「シナジー (synergy)」と呼んだ。

### 3.3 協応構造

こうした運動制御に対するベルンシュタイン・アプローチは、1980年以降になってケルソ

(S. Kelso) などのハスキンス研究所の研究者たちに受け継がれることとなった。そして、調音運動におけるダイナミクス、歩行運動、タイピングなどの打鍵運動の観測の中から、運動制御における「機能的なシナジー」、すなわち「協応構造 (coordinative structure)」の存在が次々に確かめられてきた<sup>14)</sup>。

流暢な振舞いでは、「協応構造」が運動制御の1つの単位となり、この「協応構造」の自律性に負った形で、意識の関与を最小に抑えながらの、意図に沿った制御が行われる。こうした視点から、これまでの個々の自由度に対して指定を与える「中枢による運動制御」のパラダイムを見直してみると、その多くは過制御 (over-specification) であるといえる。その意味で、この「協応構造」の概念は、不完全指定 (under-specification) での制御のメカニズムを支える重要な概念の1つであり、また無意識に行われる流暢な振舞いを説明するものである。

### 3.4 対話をレギュレートするもの

ベルンシュタイン問題は、身体の制御の問題にとどまらない。ロボットの制御方略や発話の生成、あるいはコミュニケーションの方略など、我々の身近なところにも潜んでいる。

たとえば、対話者間の相互作用の中から創発的に生じる対話におけるコーディネーションやヒューマンコンピュータインタラクションにおけるコーディネーションを、こうした「協応構造」の枠組みで説明しようとする試みもある<sup>1)</sup>。対話者間で協応構造が形作られるとき、その間で交される発話の持つ役割、発話を解釈すること、発話を産出することの基本的な意味が異ってくる。「聞き流せる」とはどういうことか、「聞き耳をたてる」とはどういうことか。「協応構造」の概念はまた、知覚的な側面や情報伝達の方略に対する新しい視点をも与えてくれる。

## 4. ボトムアップな情報戦略

### 4.1 対話を愉しむ

テニスをしていると、相手の気持ちや性格までも何となく伝わってくることもある。ボールを打ちあっていると「一体感」を感じて、初対面でこれまで深く話し込んだことがなくとも、次第に相手の人柄が伝わってくる。このとき相手から飛ん

でくるボールの物理的なスピードを測定しても、そのような情報は得られないであろう。

我々が日頃何気なく行う「おしゃべり」にも、このテニスの話に似たところがある。このおしゃべりの目的は、必ずしもお互いの情報交換や相互の発話の理解にあるとは限らない。ある時間や場を共有することで満足したり、聞き手が耳を貸してくれているだけで満足する場合も多い。そうして、ちょっとしたことばのやり取りの中で、相手の気持ちや調子を伺い知ることができる。これは、高度にスキル化されたコミュニケーションの方略であると考えられるが、言語的にまだ未発達と思われる幼児や乳児にも見られる。

母親の幼児に対する話しかけは「マザリーズ」と呼ばれる。この話しかけに乳児はニコニコしながら手足を動かしたり、「う〜ご〜」と言葉にならない声を出して応える。乳児は母親との相互作用の中で「対話を愉しんでいる」ようにも見える。お互いが喜びを感じあえる。まだ十分に語彙を獲得していない幼児とも、我々はおしゃべりを愉しむことができる。ほとんど言語的な情報が行き来しないにも関わらず、相手のことがよく分かった気持ちになる。これはコミュニケーションにおける特殊な状況ではないだろう。むしろ、言語的なコミュニケーションが成り立つ前提として、より根源的なものではないだろうか。

#### 4.2 コードモデルを超えて

2つのシステムが局所的な相互作用によって、全体としての自由度の小さなシステムを形成しているとき、「情報」の流れを考えてみると面白いことに気づく。それは、コードモデルで説明されるようなものではない。

お互いにばらばらに動いていた「振舞い」が局所的に相互作用し、2つのシステムが一緒に動き出す。これはテニスでも何でもいい。2つのシステムが自己組織化すると、系全体での自由度が減じられる。マクロスコピックに考えると確かに情報は伝わっている。そして、我々は何らかの情報をピックアップしているといえる。しかし局所的に考えると、そこにはボールのやり取り（ボールを介した相互作用）があるだけで、そのボールの性質を局所的に見ても、全体としての情報の流れはつかめない。シンボルを介したレベルでの情報伝達ではないのである。

こうした視点で考えると、自然な発話とは、対話者間における個々の「振舞い」の局所的なやり取りを媒介する「手掛かり (clue)」、あるいは「合い図 (cue)」でしかない。そして、対話とは「局所的な手掛かり (clue) のキャッチボールにより、2つのシステムが自己組織化した状態、協応構造を形成した状態」といえるだろう。

この2つの（それぞれが多くの下位システムを有した）システムがボトムアップに自己組織化するための局所的な相互作用は、言語的なことばのやり取りとは限らない。手の動きやうなずき、身体の動きなども局所的なやり取りを媒介する「手掛かり (clues)」の1つなのである。マクロに現われた、我々の日常的な感覚でのことばや対話の形は、あくまで結果としてのもので、「丁寧にコード化して送っているように見えるにすぎない」という見方もできる。

#### 4.3 1つのシステムを作る

「書く」という行為は表現意図を書き言葉としてコード化する過程である。その過程はボトムアップな要素を含む。ところが情報伝達のモードとしては「コードを伝える」というトップダウンな戦略をとる。このとき、ことばを産出する（捻り出す）ときのボトムアップな過程は見えない。これを解釈する側ではトップダウンな産出過程を仮定しても、その解釈は行えた。これはトップダウンな行動様式を持つこれまでのコンピュータでそのまま扱えた。

ところがこのトップダウンでのコミュニケーション（すなわちコードモデル）を想定して作られたシステムで、「ボトムアップな情報戦略で伝えあう日常での（ということはほとんどの）コミュニケーション」を扱おうとすると、本質的なミスマッチを生じることになる。基本は「対象との間で1つのシステムを作る」、すなわち「協応構造を作る」ことなのである。それは現象を傍観者として観測するのではなく、むしろ一緒に作り出すことなのである。

#### 4.4 エコロジカル・アプローチ

こうした「頭の中」での翻訳、解釈過程を介在させない直接的な知覚論、すなわちギブソン (J. J. Gibson) の論考<sup>3)</sup>を中心とした知覚観は、「エコロジカル・アプローチ」として知られている。「アフォーダンス」という秘妙な言葉の哲学的な

意味を深く考えなければ、論旨はきわめて単純である。すなわち、「知覚とは対象と1つのシステムを作ること...」である。

ギブソンの残した思弁的な議論を、様々な心理的実験から検証していく、あるいは数理的に説明する試みが現在もなお進行中である。たとえば、クグラー (P. Kugler)、ターベイ (M. Turvey) らは<sup>2)</sup>、非線形ダイナミクスにおける自己組織化の考え方を援用し、「アフォーダンス」や情報の意味、運動発達の問題を説明しようとしている。筆者らも、この流れに沿ってコミュニケーションや対話、ことばに対する新たな枠組を構築しようと考えている<sup>1)</sup>。

アフォーダンスの議論は、もともと「視覚」の問題に限らない。石を拾いあげて、無意識に手の中で遊んでみたり、上下に振ってみる。これは「ヘフティング」として知られている動作である。腕の動作などを総動員して知覚システムを組織し、行為の可能性 (たとえば、どこまで遠くに投げられるかなど) を知覚するのである。

## 5. まとめ

本稿では、日常の何気ない振舞いをヒントに、そして「エコロジカルアプローチ」の視点から、音声言語処理の直面している問題や我々の発話やコミュニケーションにおける基本的な行動原理を探る新たな試みの一端を紹介した。

歩きながらも本が読める、思わず口ごもる、人の話に聞き耳をたてたり、勝手に聞き流す。こうした振舞いの背後にあるメカニズムを考えると、日頃、我々が当たり前のように行っていることが「情報処理機械」であるコンピュータにとっては最も不得手なこととして映ってくる。それは「目」「口」そして「耳」を持ちはじめたコンピュータシステムが次第に自立し、我々の社会に解け込んでいくとき、はじめに超えなければならない壁なのである。そして、このブレークスルーを見出す鍵は、意外にも「思わず口ごもる」ような、我々の日常の振舞いの中に隠されているようである<sup>1)</sup>。

## 参考文献

- 1) 岡田美智男: 口ごもるコンピュータ, 情報フロンティアシリーズ, 共立出版 (1995).
- 2) Kugler, P. and Turvey, M.: Information,

Natural Law, and the Self-assembly of Rhythmic Movement, Lawrence Erlbaum (1987).

- 3) ギブソン (古崎 敬他訳): 生態学的視覚論—ヒトの知覚世界を探る, サイエンス社 (1985).
- 4) Allen, J. and Perrault, C.: Analyzing Intention in Utterances, Artificial Intelligence, Vol. 15, pp. 143-178 (1980).
- 5) Grosz, B. and Sidner, C.: Attention, Intentions, and Structure of Discourse, Computational Linguistics, Vol. 12, No. 3, pp. 175-204 (1986).
- 6) Suchman, L.: Plans and Situated Actions, Cambridge University Press (1987).
- 7) 岡田, 栗原, 大塚: 最小指定の枠組に基づく自然な発話の生成機構のモデル化, 人工知能学会研究会資料 SIG-SLUD-9302-8 (1993).
- 8) Appelt, D.: Planning English Sentences, Cambridge Univ. Press (1985).
- 9) Levelt, W.: Speaking: from Intention to Articulation, MIT Press (1989).
- 10) Kempen, G. and Hoenkamp, E.: An Incremental Procedural Grammar for Sentence Formulation, Cognitive Science (11), 2, pp. 201-258 (1987).
- 11) McNeill, D.: PSYCHOLINGUISTICS: A New Approach, Harper&Row (1987).
- 12) 岡田, 大塚: 「口ごもるコンピュータ」の研究—自然な発話におけるパフォーマンスのモデル化—, 人工知能学会全国大会論文集 pp. 407-410 (1993).
- 13) Bernstein, N. A.: The Coordination and Regulation of Movements, Pergamon Press (1969).
- 14) Fowler, C. A.: Current Perspectives on Language and Speech Production: A Critical Overview, In R. G. Daniloff (eds.), Speech Science, Taylor&Francis (1985).
- 15) 岡田美智男: 「聞き耳をたてるコンピュータ」の実現にむけて, 月刊言語 94-8, 大修館書店 (1994).

(平成7年8月7日受付)



岡田美智男 (正会員)

1960年生。1982年宇都宮大学工学部電子工学科卒業。1987年東北大学大学院工学研究科博士課程修了。工学博士。同年、NTT基礎研究所情報科学研究部。1995年より、ATR知能映像通信研究所主任研究員。音声言語処理、認知科学、分散協調AI、生態学的心理学などに興味をもつ。著書に「口ごもるコンピュータ」(共立出版)、「AI奇想曲」(共著, NTT出版)など。人工知能学会, 日本認知科学会, 日本ソフトウェア科学会など各会員。