

言語理解とロボットの行動制御

Understanding Natural Language and Controlling Robot Actions



東京工業大学 大学院情報理工学研究科
教授 田中穂積

Hozumi TANAKA, Professor
Department of Computer Science
Tokyo Institute of Technology

Understanding natural language by a computer has been one of interesting and challenging research themes in the field of artificial intelligence for more than 30 years. Many valuable theories and technologies of natural language processing have been emerged but they have not been applied to control robot actions. Along with the recent developments of computer graphics enables a 3-dimensional image to be easily created in a computer and be handled through a mouse device. The 3-dimensional image is considered as a software robot that can perform various actions mechanically impossible. It is preferable for us to use natural language to control the software robot easily. Showing our prototype system developed in our laboratory, we will discuss what kinds of problems exist to control a software robot by natural language or speech inputs.

1 はじめに

我々が知的と感じる人間行動(学習, 推論, 問題解決, 言語理解など)の原理を解明し, それらをコンピュータ上に実現することは人工知能研究の究極の目的である。アラン・チューリングは, 壁の向うにいる人間と機械のそれぞれに対して, 人間が英語や日本語などの自然言語による会話やゲームを行なうことで人工知能の実現を判定しようとするチューリングテストを提案している。両者の応答の差から, 壁の向うの人間と機械の見分けがつかなければ, その機械は人工知能であると判定するのである。しかし, このテストには重要な視点が一つ欠けている。壁の向うの相手が見えないため, 相手の動作をみて人工知能が実現したかどうかを判断することが許されていないからである。自然言語理解(以下, 言語理解)能力についても動作という視点は重要である。言語理解を通じて, 相手に動作や行動を促したり(行動制御), 逆に他人からそれを促されたりしながら社会生活が営まれていることを我々は知っている。人間の動作や行為は言語理解と密接に関係しているのである。

ところがこれまでの言語理解の研究では, チューリングと同様に, 動作や行為という視点が欠落していた。従来の研究では, 機械翻訳, 情報検索, 文書分類・要約など, 主として自然言語で書かれた文書を対象としていたからである [Allen 95][?][?] [田中 99]。それに対して, 行動という視点から言語理解の機構を解明することも, 人間の知能, 知的行動の原理の一端を明らかにしようとする立場から重要なことである。この最終目的に至るために以下の研究を行なっている:

コンピュータの内部に機械的な制約を受けない, 行動機能が豊富な3次元のソフトウェアロボット(機械)を構築し, それを自然言語の対話により仮想空間内で動作/制御させることを通じて, 言語理解の機構をコンピュータ上に構築する。

このような状況設定では, 言語によってロボットの行動を制御するので, 言語と行動との関係について具体的に研究することが可能になる。言語と行動の関係は, 言語哲学の分野でも研究が進み, いくつかの理論も提案されている。本研究では, 具体的に動作する計算モデルの構築を通じて, そ

これらの理論を実証的に研究し、言語理解に関する新しい問題を発見し、これらを解決するための言語と行動に関する新しい理論の構築を目指している。

本研究と類似した研究は80年代後半から90年代の前半にかけて米国のペンシルバニア大学で行なわれている [Badler 93]。しかしそこでの研究は、ニュートン力学を考慮したソフトウェアロボットの動作機能の実現技術に関する研究に重点が置かれていた。当時のCG技術の水準が未熟であったこともあり、十分な成果を挙げるに至っていない。言語と行動の関係についても研究がなされていたが、言語関連の有力研究者が他大学に移動したこともあり、本格的な研究に着手することなく現在に至っている。後述する言語行為の問題については未着手である。Badlerらの研究は工学的な立場を明確にしており、認知科学的な視点をまったく欠いている。

2 プロトタイプシステムの概要

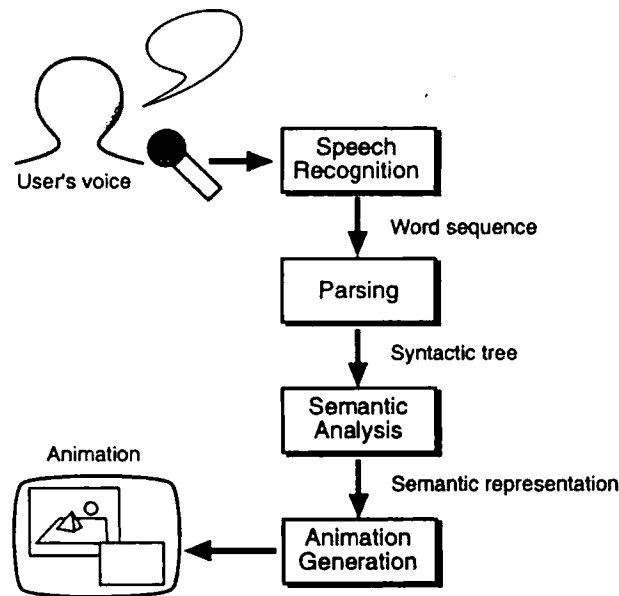


図1 プロトタイプシステムの構造

我々の開発したプロトタイプシステムの構造を図1に示す [新山 99]。コンピュータ内のソフトウェアロボットに対して音声で指令を出すと、音声認識システムはこれを単語の列に変換する。音声認識システムとして市販の製品を用いているが、この市販のシステムは単語の列からなる文の意味を一切理解していない。これは一見不思議のように思えるが、ほとんどの市販の音声認識システムは統計的な言語モデルを用いて音声から単語の列を作り出す。いわゆる構文解析や意味解析と言った、いわゆる自然言語処理技術はまったく使っていないことから、市販の音声認識システムが文の意味を理解していないことが分かると思う。人間は音声認識と言語理解を同時に行なっていると推察されるが、このような原理に基づく音声認識システムは実用段階に至っていないのである。

そこでわれわれのプロトタイプシステムでは、音声認識システムの出力である単語の系列を文と見做し、それに対して構文解析や意味解析を行なって意味理解の結果 (Semantic Representation / 意味表現) を出力する。この意味表現をさらにロボットの動作指令に変換し、ロボットの動作映像を最終的に得ている。

このプロトタイプシステムは、最近の音声認識技術と自然言語処理技術、それにコンピュータグラフィクス (CG) 技術の進展により初めて可能になったといえる。コンピュータ内部に機械的な制約を受けない、行動機能が豊富な3次元のソフトウェアロボットを作り、それを実時間で自在に動作させることが可能なCG技術が幾つか生まれている。現在のCG技術では、あらかじめ決まった動作を人間にさせ、それを3次元の映像としてコンピュータに取り込むことにより、リアルな動作映像を再現する技術、すなわちモーションキャプチャ技術が用いられることもあるが、行動機能は自ずと制限され豊富であるとは到底いえない。たとえばニュートン力学の世界で自由自在に動作することが可能で豊富な動作機能をもつソフトウェアロボットの開発は、依然として将来の研究課題として残されている。そこでとりあえず、映像を映し出しているカメラもまたソフトウェアロボットであるとして、カメラに指令を出すことができるシステムをプロトタイプシステムとして試作している。カメラの動作は、カメラの映し出す映像の変化として知ることができる。

3 動作例と言語理解と行動に関する問題

プロトタイプシステムの動作の概略を説明する。まず始めに、3人のソフトウェアロボットが仮想空間内で演奏している映像をカメラが映し出しているところから出発する。このカメラロボットC1に対して音声による指令を出して、カメラの位置を変えるとともに、そのカメラが映し出す映像が連続的に変化してゆく。これを動作例とともに以下に示す。

動作例：

指令1： ドラマーの後ろに回り込んでください。

(C1が映し出す映像がドラマーを中心に連続的に変化し、ドラマーの後ろ姿を映し出す)

指令2 もうちょっと後ろに。

(ドラマーの姿が少し小さくなる)

指令3 もうちょっと。

(ドラマーの姿がさらに小さくなる)

指令4 ちょっと行き過ぎ。

(ドラマーの姿が若干大きくなる)

.....

このC1動作例から、言語理解と動作に関するさまざまな問題の一端を示すことができる。

指令1に関しては、「回り込む」という動作にはさまざまな「回り込み方」がある。そのいずれか一つを決めて、カメラの位置を変化させなければ映像が作れない。指令2では、カメラがどの程度後ろに引くかを決めなければならない。これらは、言語学者が漠然性 (vagueness) とよんでいる問題である。さらに、指令2は、指令1のカメラの最終位置を基準として、ドラマーから遠ざかる方向に動かなければならない。これは、脈絡に依存した言語理解が必要になることを意味している。さらに、指令2では前文より省略語を推論し、「ドラマーのもう少し後ろに下がる」という意味であると解釈しなければならない。指令3でも指令2と同様な言語理解が必要になる。指令4では、文字通りの意味解釈とは異なる行動/動作をカメラに要求している。「カメラに、ドラマーの方にもう少し近づく」ことを要求しているからである。これは言語行為論として哲学者らが研究しているが [Austin 62][Searle 69]、言語理解と行動の研究では、こうした問題がただちに顕在化する。

我々の開発した現在のプロトタイプシステムでは、指令1から指令3までは理解することができるので、括弧で囲まれた記述どおりの映像を得ることができる。しかし指令4で説明した言語行為に関する問題が扱える段階にまだ至っていない。

4 応用分野

われわれは、言語理解と、音声認識技術、CG技術の連携をはかり、言語によりソフトウェアロボットの行動を制御するシステムを開拓することを目的にして研究を行なっている。ソフトウェアロボットは、コンピュータ内部の仮想空間に存在し、言語による人間からの指示を理解し、それにしたがって仮想空間内で動作する。このようなことを可能にする技術をベースにして、工学的な見地から以下に述べるさまざまな応用が考えられる [Badler 99]。

ソフトウェアロボットを、これまで熟練した専門家が操作していた医療機器、マイクロマシンなどで置き換えれば、それらを音声によって操作することが可能になる。手足が不自由で機器の操作ができなくても、話すことが可能なら、たとえば介護ロボットの動作を制御することができる。さらに、分子構造、都市構造、宇宙の構造など、ミクロからマクロなレベルに至るさまざまな立体モデルを映し出すカメラもまたソフトウェアロボットとして考えることができるので、カメラへの指示を音声によっておこなうことにより、さまざまな角度から立体構造を容易に観察することも可能になる。このように本格的なシステムが実現すれば、これは他の学問分野にもインパクトを与える可能性があると思われる。これらは自然言語によるヒューマンマシンインタフェースの実現例にもなっている。エンターテインメントの世界に目を向けて、ソフトウェアロボットをアニメの人物であると仮定すれば、アニメーションの制作支援にも応用できる。これはアニメーションの制作現場を変えることにもなるだろう。

5 今後の研究の進め方

すでに述べたように、本研究を通じて従来見過ごされてきた言語理解の新しい研究課題を発見することも可能である。

たとえば、ソフトウェアロボットは、個々の単語の概念を理解していたとしても、「部屋の外に出る」という行動を行なうのに十分とはいえない。「部屋の外に出る」という概念には、少なくとも「ドアのところまで歩いていき、ドアにノブがあればそれを回して引いたり押ししたりしてドアを開けて外に出る」ということが含まれている、ということを知らなければならない。こうした一連の行動手順（行動計画）を知らなければ、それらを学習しなければならない。このように、言語理解と行動の関係を研究することにより、学習の問題が直ちに顕在化してくる。これは文書理解の研究からは気がつきにくい言語理解に関する重要な研究課題であろう。

また、言語の文字通りの意味ではなく、その背後に真の意味があり、それに基づいた行動を行なう必要があると言うことを第3章で述べておいた。たとえば食堂で「喉が乾いた」とウェイターに言えば、「喉を癒すものが欲しい」と言う意味が真の意味だから、ウェイターはそれに答える動作をしなければならない。この言語行為もまた本研究の延長上で本格的に取り組む必要がある。

われわれは、今後の研究の進め方として、以下の4つの具体的な研究課題に取り組むたいと考えている。

- 言語理解の機構に関する研究
 - 構文解析技術の研究
 - 意図抽出を含む意味理解機構の研究
 - 脈絡を考慮した文脈理解の研究
 - 対話の分析と対話理論の構築
- 行動機能が豊富な3次元ソフトウェアロボットの構築
 - ニュートン力学の世界でのソフトウェアロボットの動作機能の実現に関する研究
 - 豊富な表情制御と、音声合成と口唇動作の同期に関する研究
 - ロボット動作の実時間映像化に関する研究
- ソフトウェアロボットの行動制御理論の構築
 - 意味理解結果をロボットの行動計画へ変換する機構の研究
 - 複雑な行動手順の学習理論に関する研究
- 言語と行動に関する認知理論の構築
 - 言語行為論
 - 言語理解機構の認知科学的モデル

上記した研究課題の相互関連図を以下に示す。

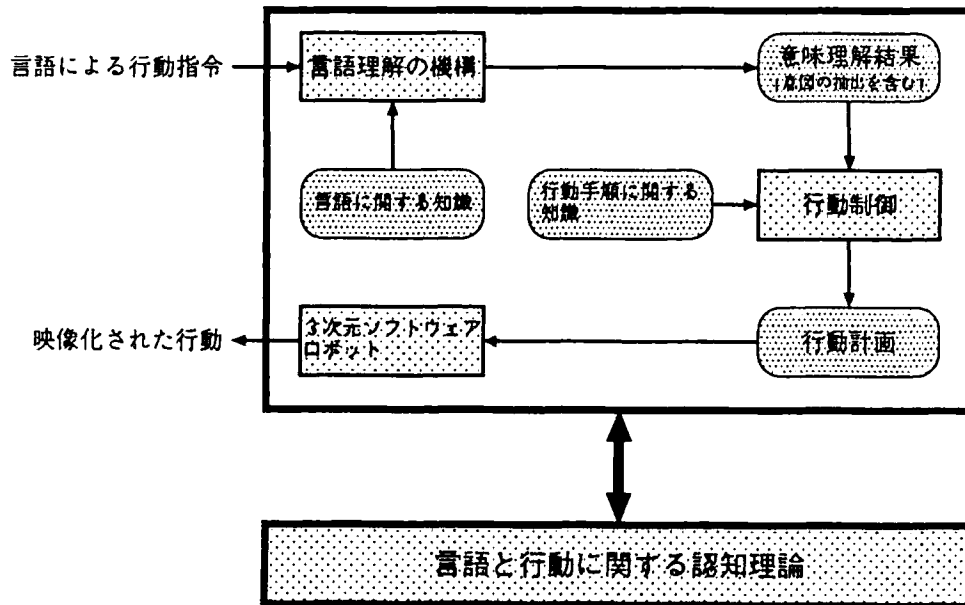


図2 研究課題の相互関連図

言語と行動に関する認知理論には、意図抽出を含む意味理解機構の研究と意味理解結果をロボットの行動計画へ変換する機構の研究や、対話理論、複雑な行動手順の学習理論の研究が特に深く関係している。人間の話す言葉は必ずしも文法にかなったものでないことが多い。たとえば、「えーと、もう少し前、いや後ろに下がって」などと言うことも稀ではない。このような文を ill-formed sentence とよぶが、この種の文の解析技術も重要になる。これは構文解析技術の高度化と密接に関係している。ソフトウェアロボットに複数の動作を同時に行なわせる言語表現がある。たとえば、「手を振りながら別れる」などという表現である。これはロボットの行動計画への変換の問題として考えることができる。行動機能が豊富な3次元ソフトウェアロボットの構築を進め、カメラだけでなく仮想空間中のロボットを言語によって制御するモデルの構築を進める過程で、さまざまな問題を解決していきたいと考えている。

6 おわりに

前章まで明確に述べなかったが、言語理解と行動に関する研究は緒についた段階にある。言語学、哲学、計算言語学、認知科学、人工知能、情報工学、制御工学が関わる学際的な立場から研究を推進すべき研究課題であり（図3参照）、未解決の問題が山積している。複数のソフトウェアロボットに協調して仕事をさせたりすることも、困難ではあるが興味ある研究である。これまでは、言語によるソフトウェアロボットへの一方的な指令を考えてきた。これだけでも問題山積の状態であるからそれに対応するのに精一杯であるといえるが、将来的にはソフトウェアロボット自身が言語で応答するだけでなく、感情を持ち表情豊かに応答することが可能なら、これは人工生命の研究とも接点を持つことになるだろう。

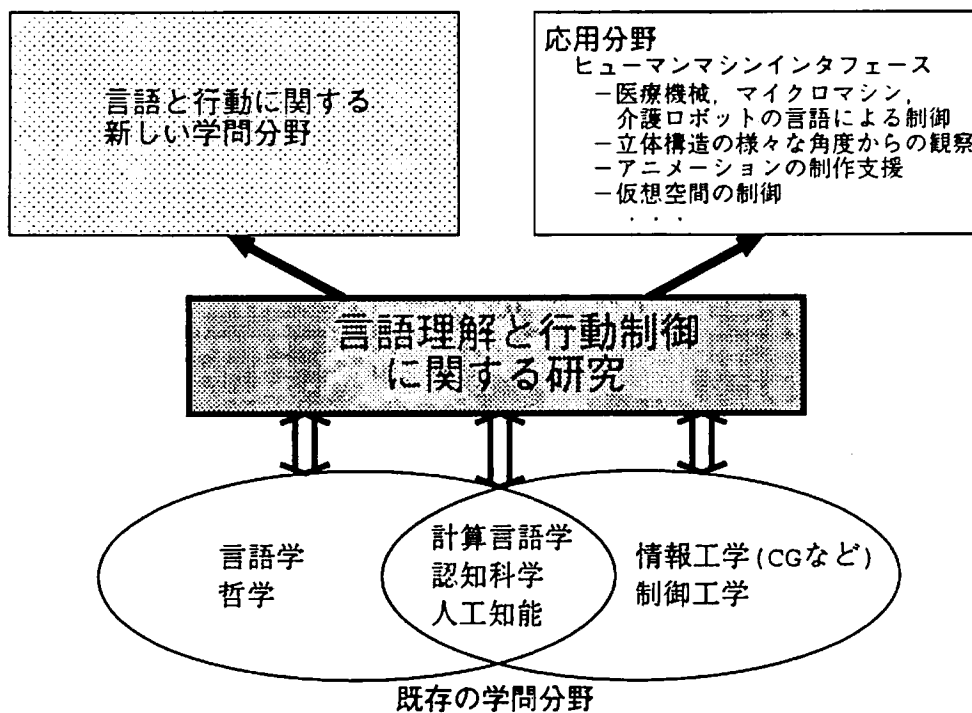


図3 学際的な研究の必要性

ソフトウェアロボットは最近ではエージェントと呼ばれることが多く、エージェントに関する研究は人工知能の国際会議や、エージェントに絞った新しい国際会議が開催されている。ところがいずれの会議でも言語とエージェントの行動を結び付けた研究発表はまだ皆無に近い。しかし、このテーマは言語学的にも哲学的にも認知科学的にも工学的にも重要な研究課題であり、今後、関連学会での研究活動だけでなく、学際的な学問領域においても活発に研究されることを期待したい。

本稿で述べた研究に興味を抱いたのは、あたかも我々が映画監督になった気分で、仮想空間中の俳優ロボットを動かしてアニメーションを制作したいということであった。考察を進めるうちに学問的にも極めて重要で興味ある問題が山積していることに気がついた。いつ適うか分からないとはいえ筆者の夢からこの研究が始まっている。

参考文献

- [Allen 95] Allen, J.: *Natural Language Understanding*, The Benjamin/Cummings Publishing Co., Inc. (1995).
- [Austin 62] Austin, J.L.: *How to Do Things with Words*, Harvard University Press (1962). 坂本百大訳: 言語と行為, 大修館書店, (1978).
- [Badler 93] Badler, B., Phillips, C. and Webber, B.: *Simulating Humans: Computer Graphics Animation and Control*, Oxford University Press (1993).
- [Badler 99] Badler, N.I., Palmer, M.S., and Bindiganavale, R.: *Animation Control for Real-Time Virtual Humans*, Comm. of the ACM, Vol. 42, No.8, pp.65-73(1999).
- [Mani 99] Mani, I. and Maybury, M.T.: *Advances in Automatic Text Summarization*, MIT Press (1999).
- [Searle 69] Searle, J.R.: *Speech Acts*, Cambridge University Press (1969). 坂本百大, 土屋 俊訳: 言語行為, 勁草書房, (1986).
- [長尾 96] 長尾 真 (編) : 自然言語処理, 岩波書店 (1996).
- [新山 99] 新山, 秋山, 鈴木, 徳永, 田中: 自然言語を理解するアニメेटドエージェントのための3次元仮想空間における位置の表現と処理, 人工知能学会全国大会 (第13回) 論文集, pp.217-220, (1999).
- [田中 99] 田中穂積 (監修) : 自然言語処理-基礎と応用-, 電子情報通信学会 (1999).