

## 言語理解と行動制御

### ー 音声認識・言語理解・コンピュータグラフィクス技術の統合 ー

田中穂積<sup>†</sup>

<sup>†</sup> 東京工業大学大学院情報理工学研究科 〒152-8550 東京都目黒区大岡山 2-12-1  
E-mail: <sup>†</sup> tanaka@cl.cs.titech.ac.jp

あらまし 「言語理解と行動制御」は、2001年度から5年間の予定で開始した学術創成研究の名称である。この研究は自然言語による指令により、仮想空間内の自律ロボットを動作させる学術を新たに創成することを目的としている。そこには、音声認識と言語理解とコンピュータグラフィクスに関する諸技術を統合化したシステムのプロトタイプシステムを開発することが含まれている。ロボットを実際に動作させるためには深い言語理解が必要になることがある。一例を挙げれば、「赤い玉を右に動かせ」という指令の場合、「もっと右」とほどの程度右なのかを決めなければ、ロボットの動作を映像として可視化できない。「赤い玉」と言われてもそれが(仮想)空間内に複数個あれば、どの赤い玉を動かすのかを決める必要がある。前者は Vagueness 解消問題、後者は Object Identification の問題であるが、いずれもこれまでの言語理解、対話理解の研究で見過ごされてきた問題である。本学術創成研究では、そのほかに、ロボットの行う Gazing, Gesturing など非言語的な動作のコミュニケーションに果たす役割の研究、ロボットを自律させるための計画立案能力や多様な動作が可能なロボットに関する研究、雑音下での実時間の音声認識、イントネーションの研究などを行う予定である。ソフトウェアロボットのプロトタイプシステムの試作とともにハードウェアロボットについても、同様な研究を行う。以下では本研究計画の概要を中心に述べる。

キーワード 音声認識、言語理解、CG、プランニング、ソフトウェアロボット、仮想空間

## Language Understanding and Action Control

Hozumi TANAKA<sup>†</sup>

<sup>†</sup> Graduate School of Information Science and Engineering 2-12-1 Oookayama, Meguro-ku, Tokyo, 152-8550 Japan  
E-mail: <sup>†</sup> tanaka@cl.cs.titech.ac.jp

**Abstract** The Natural Language Understanding and Action Control started in 2001 as 5-year research project of a Grant-in-Aid for Creative Scientific Research supported by the Ministry of Education, Culture, Sports, Science and technology, and Japan Society for the Promotion of Science. As Natural Language understanding (NLU) research environment has changed drastically in the past two decades. As better technologies in speech recognition, natural language processing and computer graphics are now available. Combining these technologies in this project, we are going to research into NLU more deeply by developing life-like animated agents who can understand commands in spoken language and perform actions specified by the commands. The experience to build a prototype agent system will teach us its strength and shortcomings and made us identify the requirements for future NLU system. The results obtained by this research will be naturally applicable to hardware robots, which indicate a grounding of the research in a real world.

**Keyword** Speech Recognition, Natural Language Understanding, Computer Graphics, Planning, Software Robot, Virtual Space

### 1. 目的

コンピュータが我々の身近な存在になるにつれて、我々の生活空間は最近大きな変貌をとげている。物理空間から、仮想空間に居住する割合が増えてきている。仮想空間を快適なものとするためには、(物理空間と同様に仮想空間内でも)自然言語を用いたコミュニケーションが可能であることが望まれる。最近では言語だけでなく音声

や映像を用いたマルチモーダルなコミュニケーションの手段が検討されている。これは、コンピュータグラフィクス(CG)技術と音声認識技術の著しい進展と無関係ではない。コンピュータ内の仮想空間に3次元のソフトウェアロボット(エージェント)を作り出して動作させることも容易になってきた。

これまでのマルチモーダルな対話システムでは、基本的に言葉による質問に対して、言葉で応

答する質問応答システムの構築に主眼がおかれており、仮想空間内に存在するソフトウェアロボットを実際に動作させ、仕事をさせようとする研究は、1970年代初頭のSHRDLUシステムを除き皆無であったと言ってよい。MITのWinogradが開発したSHRDLUシステムは[2]、人間からの英語による指令に対して、ソフトウェアロボットが積み木の世界で動作し、積み木を積み直したり、現在の積み木の世界の様子を答えることが可能であり、当時のAI研究における画期的な成果の一つであったといえる。

しかし当時はコンピュータパワーが十分でなかったこと、CG技術、音声認識技術が未熟であったことなどから、ロボットと言っても一つの逆T字型の棒が上下左右に移動するだけのもので、動作する人間に似たロボットとよべる代物ではなかった。指令は音声ではなく、タイプライタから打鍵するシステムであった。このような制限はあったものの、ロボットは自律しており、それが動作するためのPlanning(計画)能力を有していたこと、「その」などの指示代名詞の指すものを同定する言語理解能力も有していたことは特筆してよい。その後の言語理解の研究はSHRDLUの限界に挑戦することになったと言ってよい。しかし言語理解システムとして、SHRDLUほどデモンストレーション効果が高く、AIの研究と自然言語処理の研究に大きなインパクトを与えたシステムの例を筆者は知らない。

最近になって、コンピュータパワーの著しい向上、それに伴う音声認識技術とCG技術の進展により、SHRDLUシステムを超える言語理解システムの研究に取り組むことが可能になってきた。我々の生活空間が仮想空間にまで広がってきたというような背景から、新しいタイプの言語理解の研究が益々必要になってきたように思われる。映像を含む高度なマルチモーダルな対話システムとして、仮想空間内のソフトウェアロボットが言葉による対話だけでなく、実際に動作し仕事をすするシステムの構築に関する研究は、ハードウェアロボットを物理空間(実空間)で動作させることにも役立つ[1]。

こうしたことを見越して、我々は「言語理解と行動制御」に関する研究が必要であるとして、2001年より5年間の学術創成研究として開始した。自然言語の指令により、仮想空間内に存在する自律ロボットを動作させるためには新しい学術を創成する必要があると考えたからである。音声認識と言語理解とコンピュータグラフィクス

の技術を如何に統合したらよいか、またエージェントの行う動作と言語理解とを結びつけるためには、どのような問題があり、それをどのように解決しなくてはならないかを探るためには、これまでにない新しい学術の創生が必要になる。本研究計画には、プロトタイプシステムの作成が含まれている。それは、開発した新しいアルゴリズム・システム・理論・方法を、プロトタイプシステムを実際に動作させて、有効性を確認したかったからである。

## 2. 研究組織

言語理解とロボットの行動制御の関係深く探るために、学際的な研究が必要であると考えて、現在21大学、27教授・助教授からなる以下の研究組織で計画を推進している。

研究代表者：田中穂積(東工大)

1. 基礎研究班：土屋俊(千葉大)、山田友幸(北大)、辻幸夫(慶大)、山梨正明(京大)、楠見孝(京大)、丸山直子(東京女子大)
  2. 言語処理班：田中穂積(東工大)、辻井潤一(東大)、白井清昭(北陸先端大)、奥村学(東工大)、黒橋禎夫(東大)、松本裕治(奈良先端大)、徳永健伸(東工大)、乾健太郎(奈良先端大)、飯田仁(東京工科大)、石崎雅人(東大)
  3. 音声処理班：古井貞照(東工大)、牧野正三(東北大)、河原達也(京大)、鹿野清宏(奈良先端大)
  4. ロボット班：中嶋正之(東工大)、北橋忠宏(関学院大)、白井良明(阪大)、小林哲則(早大)、佐藤誠(東工大)、広瀬啓吉(東大)、原島博(東大)、小林隆夫(東工大)
- 顧問：長尾眞(NICT)、辻三郎(阪大名誉教授)、白井克彦(早大)、野家啓一(東北大)、井出祥子(日本女大)

## 3. 研究課題

具体的な研究課題を列挙する。

1. 自然言語による行動指令とそれに伴うロボットの行動制御
  - a) 指示代名詞の指示するものの決定(Resolution of anaphoric relation)。
  - b) 発話で省略されたものの推定(Ellipsis handling)。
  - c) 指示物体の決定(Object identification)。(これは、後述するように、a)と微妙に異なる。)
  - d) 不明確性の処理(Vagueness handling)。
  - e) 談話の管理(Discourse management)。

- f) 空間位置の言語表現とその理解 (Space understanding)。
  - g) 発話意図の理解 (Understanding speech act)。
  - h) 信念のモデル化 (Belief modeling)。
  - i) 行動に伴う計画立案と実行 (Planning and task execution)。
  - j) 衝突回避アルゴリズム (Collision avoidance algorithm)の開発。
2. 実環境でのハードウェアロボットの行動制御
- a) 物体認識 (Object recognition)。
  - b) 物体の把握制御 (Object grasping with haptical control)。
3. 快適なユーザインタフェース
- a) 話し言葉の処理 (Spoken language analysis)。
  - b) マルチエージェントシステムにおける一対多の対話処理。
  - c) 非文 (Ill-formed sentence analysis)の理解。
  - d) 非言語表現 (Non-verbal expression)の理解。
  - e) パラ言語現象 (para-linguistic phenomena)の処理。
  - f) 雑音環境下での音声認識 (Speech Recognition under noisy environment)。
  - g) 実時間音声認識 (Real-time speech recognition)
4. プロトタイプシステムの試作
- a) CGによる Life-like な3次元ソフトウェアロボットの作成。
  - b) 多種多様な動作が可能なソフトウェアロボットの開発。
5. 言語理解と行動制御に関する基礎研究
- a) 言語行為論の研究
  - b) 認知言語学から見た言語と行動
  - c) 空間理解の認知科学研究
  - d) 話し言葉の言語学的研究

まず1について、言語理解とロボットの行動との関係を明らかにして、仮想空間内でソフトウェアロボットを行動させるのであるが、ロボットの行動は動画として画面に表示される。そのために、より深い言語理解の研究が必要になる。

1のc)は、「赤い玉」が世界に複数個存在している場合、「赤い玉」という言語表現は、どの「赤い玉」を具体的に指しているのかを決める同定問題である。仮想空間内でロボットが具体的な行動を起こすためにも、この同定問題は必ず解決しておく必要がある。これがc)に述べた問題である。a)の照応指示関係の解

決では、どの「赤い玉」であるかまで同定することは普通しない。たとえば機械翻訳では、物体同定まで行う必要はなく、「赤い玉」を単に the red ball と訳すことで済ましてしまうことができるからである。ところが物理空間であれ仮想空間であれ、ロボットを動作させそれを可視化するためには物体同定の問題を避けて通るわけにはいかない。

d)の問題は、これまでの言語理解の研究で十分に扱われてこなかった問題であるが、ロボットに動作させようすると、この問題が顕在化する。この問題を少し詳しく説明してみたい。たとえばロボットに対して「右に曲がれ」という動作指令を出したとき、ロボットはどの程度「右」に曲がる動作をしたらよいか。右に曲がる角度はアナログ値であろうから、それこそ無限の可能性がある。ロボットはその無限の可能性の中から妥当な値を決めて右に曲がる動作をしなければならない。ロボットの動作は映像として視覚化しなければならないので、この問題の解決は重要である。これは vagueness の解消問題と呼ばれている。

Vagueness の解消問題を複雑にするもう一つの要因は曲がる角度が、ロボットの置かれている状況に依存して決まることである。直線の道の中で左の壁にぶつかりそうになったときの「右に曲がれ」と言うときの「右」と、道の直角のコーナーで「右に曲がれ」というときの指令の「右」とは、おそらく異なる。後者は「直角に右に曲がれ」という意味に解釈しなければならないだろう。さらに、ロボットの行った「右の」解釈が適当でなく、「もうちょっと右に」という指令に含まれる「右」の解釈も状況に依存した vagueness の解消問題が含まれる。このように vagueness 解消問題は、動作の可視化にあたって意外に困難な問題を内包している。この問題の解決なしにはロボットの行動は映像化できないのである。

状況依存性は1のa),b),c)にも共通して現れる。その意味で、言語理解と行動制御の立場から、解決すべき大きな問題の一つは、状況依存の言語理解を行うことであるとして過言ではないだろう。そのために、e)の談話管理をきちんと行うことが必要になる。

1のf)は、ロボットに道案内をするときに含まれる地図情報教示などの言語表現に関する問題である。「まっすぐ道路を進み、道路の右手に見える郵便局の右の建物」という指令に含まれる「右」と、「郵便局を出た右の建物」という指令に含まれる「右」とでは、その解釈が異なる。ロボットの視界(状況)を考慮しなければ正しい「右」解釈は得られない。このよう位置情報もまた、ロボットの視界を考慮した状況依存の解釈をしなければならない。

g)は、自然言語による指令が、文字通りの意味では

ない解釈が必要になる場合があるということである。たとえば、「右に曲がれ」と言う指令の後で、「行き過ぎ」という発話が行なわれた場合には、発話者の真の意図は「行き過ぎ」という文字通りの解釈ではなく、「曲がり過ぎなので戻りなさい」という指令であると解釈すべきである。これは典型的な「間接言語行為」とよばれる言語現象であり、一般に解決が困難な問題である。

h)は自律したロボットの持つべき機能の一つである。

i),j)は、2. のハードウェアロボットの行動制御にも共通する研究課題である。ソフトウェアロボットといえどもこれらの問題を解かなければならない。i)には、状況変化に対応した再計画の問題も重要な研究課題である。j)は、CGの問題として解かなければならないので、物理的な衝突がないだけに、そのシミュレーションを事前に行わなければならない、物理空間とは異なる問題を解かなければならない。

3. の a)の話し言葉には、話し言葉特有の、書き言葉に含まれない言語現象が多々見られる。日本語の場合特に省略が多く、助詞落ちなどの言語現象が良く見られる。話し言葉については言語学者も十分な分析が行われているわけではない。話し言葉の言語学的な検討も本研究計画の大きな研究課題である。これが5.の d)である。b)について、ソフトウェアロボットは格好の multi-agent システムの研究を行う場を与えるということを指摘しておきたい。ハードウェアロボットと比較して、ソフトウェアロボットを作り出すことは、いとも簡単に可能であるので、multi-agent システムの研究を行うのに都合が良いからである。一対多の対話モデルとして、ロボット相互間での対話が行えるようなモデルの構築も考えたい。複数個のロボットが協調して仕事を行う問題も、Multi-agent system の枠組みの中で考えることができよう。

c)の非文の処理であるが、話し言葉による指令には、「あー」とか「あの一」などという語が発話途中で挿入されたり、「言い直し」などが頻繁に起こる。このような文であっても処理を続けることが可能な頑健な処理システムを構築することが非文処理の目的である。d),e)は、これまでのマルチモーダルシステムにおいても、円滑なコミュニケーションを行うために一定の役割を果たすことが知られている。ロボットの行動に、これらの非言語表現による応答を含めることは、ハードウェアロボットに比べて容易であろう。

4 のプロトタイプシステムは[3]、言語理解と行動制御に関わるさまざまな問題を、プロトタイプシステムの試作を通じて発見したり、開発したさまざまなアルゴリズム・理論・手法の有効性を検討するためのテストベッドとして役立つものと思われる。本研究計画の応用として、さまざまなものが可能であることが明らかになるだろう。1 の i), j) で述べた行動に伴う計画立案と実行、衝突回避アルゴリズムの研究も、プロトタイプシステムの試作を通じて有効性を検証したい。3 の b)に関しては、マルチエージェントシステムに関する研究のテストベッドとして、このプロトタイプシステムを利用して、複数ロボットの協調動作の研究を行う予定である。

5.は、学際的な立場からの言語と行動制御に関する基礎理論の構築を目指す。図1に、本研究計画の学際的研究の相互の関連図を示す。

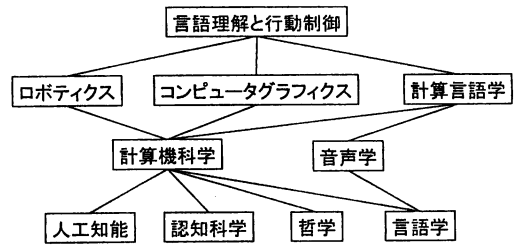


図1 本研究に関連する学術分野

### 3. システムアーキテクチャ

次頁図2に音声認識、言語理解、CG技術を統合したシステムの構成図を示す。矢印の向きに情報が処理される。概略は次の通り。

まず人間の音声が入力されると音声認識モジュールにより、単語の系列に変換される。変換された単語の系列は形態素解析(morphological analysis)を経て品詞付きの形態素列に変換される。このとき、主として辞書の情報を用いる。品詞付きの形態素列は次に文法に適合した文ならその構文構造(parse tree)を作り出す。構文構造は意味解析(semantic analysis)を経て意味構造を作り出すが、その構造は、次の「あの」とか「その」などの指示代名詞が何を指すかを定める対象物同定を行う処理(anaphora resolution)を施す。Vaguenessの解消もこの時行うだけでなく、過去の対話履歴(dialogue history)を使ったり、現在の対話履歴を残すことなどを行う。そして過去の行動計画を参照しながら、行動計画を立案する。実際の行動は、仮想空間の現在の状況を見ながら障害物との衝突を避けることなどを行い最終的なロボットの行動を動画として画面に表示する。

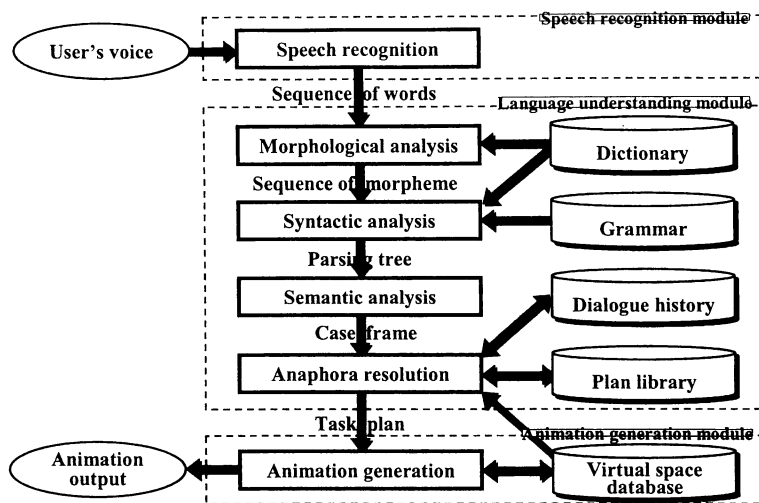


図2 システムアーキテクチャ

#### 4. 研究成果と考察

幾つかの成果例を挙げれば、日本語音声認識システムの開発、並列処理による実時間音声認識の新しい方式の開発、非文処理の新しいアルゴリズムの開発、空間位置表現の理解アルゴリズム、並列動作の言語表現の分析、話し言葉の特徴と分析、実ロボットの言語理解と行動に関する研究、ロボットの多様な動作を作り出すCG技術、計画立案とその実行に基づく照応解決アルゴリズム、などが挙げられる。その一部は、以下の発表で報告されるだろう。個々の詳細な研究成果は、以下を参照してほしい。

<http://www.cl.cs.titech.ac.jp/sinpro/Reports/Report2003e.pdf> (2003eを2002eとすれば、2002年度の英文報告書がえられる。) [4][5]

これまでの研究成果の一部は、プロトタイプシステムに集約されている。図2のアーキテクチャに基づくプロトタイプシステムの動作をビデオによるデモンストレーションとともに説明する。

プロトタイプシステムの試作を通じて明らかになったことを幾つか挙げる。一対多の対話モデルでは、さまざまな問題が生じる。仮想区間に複数のロボットが存在するので、人間からの指令がどのロボットに対して発せられたかを、個々のロボットは判断しなければならない。これは、対象物同定以前に行うべきことであろうが、この問題を本格的に扱った研究を筆者は知らない。ロボット相互間の対話が可能な状況で、こ

の発話者同定の問題はどのような意味をもつことになるだろうか。

第2に、一つのロボットの仮想空間内での行動は、そのロボットの行動だけで閉じるものではない。行動は、副作用として他のロボットに別の行動を引き起こすことあるからである。たとえば、あるロボットが玉を押すという行動は、他のロボットが、押すという行動を注視する(gazing)という行動を引き起こすだろう。そのほうがロボットの行動としては、人間の行う行動にちかくなるように思われる。ソフトウェアロボットの存在する仮想空間の世界では、このようなことを考慮しなければならない。

次に、ロボットの行う動作により、仮想(実)空間内の状況は動的に変化する。次のようなことも考えられる。複数のロボットが同時に同じ玉を掴もうと、その玉に近づくとといった状況である。この時、一早くその玉に近づいたロボットがその玉を持ってしまったとき、取りそこなったロボットはどう行動したらよいか。取りそこなったロボットは、おそらく状況が変化したことを認識して、改めて動作計画をたて直さなければならない。これはAIにおけるre-planningの問題になる。自律したロボットは、状況変化を常に検知する能力がなければならない。

ここで、言語理解とロボットの行動制御の立場から、ソフトウェアロボットとハードウェアロボットの比較をしてみよう(表1参照)。この表1から、本研究を行

うとし、ハードウェアロボットをベースにするか、ソフトウェアロボットをベースにするか、相互の利点、欠点などが見て取れると思う。言語理解に関する限り、ソフトウェアロボットをベースにして研究することに利点があるように思われるが読者はどう判断されるだろうか。

	Hardware Robot	Software Robot
多様な行動パターン可能? (Variety of actions?)	No	Yes
環境の検出感知能力? (Perception of environment?)	Yes	No
多エージェント・システムに 拡張可? (Multi-agent system?)	No	Yes
ニュートン力学の世界のシミュレーション必要? (Simulation of Newtonian World?)	No	Yes

表1 ソフトウェアロボットとハードウェアロボットとの比較

最後に本研究の応用例を幾つかあげて結びとしたい。

1. ゲームなどの娯楽用システム
2. 介護ロボットシステム
3. 教授システム(装置の操作を動画として教授するマニュアルなど)
4. 手話システム
5. 地図などのナビゲーションシステム
6. 情報家電

など、多様な応用が考えられよう。

#### 文 献

- [1] 田中穂積, 徳永建伸, “ロボットとの対話--人工知能からのアプローチ”, 情報処理, 44, 12, pp.1247-1252, 2003
- [2] T. Winograd, Understanding Natural Language, Academic Press, New York, 1972.
- [3] H. Tanaka, K. Tokunaga, Y. Shinyama, in Life-Like Characters, ed. H. Predinger and M. Ishizuka, pp.429-443, Springer-Verlag, Berlin Heidelberg New York, 2004.
- [4] URL:  
<http://www.cl.cs.titech.ac.jp/sinpro/Report2002e.pdf>
- [5] URL:  
<http://www.cl.cs.titech.ac.jp/sinpro/Report2001e.pdf>