社団法人 電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.
NLC2000-33, SP2000-81(2000-12)

〔招待論文〕

言語理解とロボットの行動制御
— 音声認識から音声理解へ —

田中 穂積

東京工業大学 大学院情報理工学研究科

〒 152-8552 東京都目黒区大岡山 2-12-1
03-5734-3046
tanaka@cl.cs.titech.ac.jp

あらまし

コンピュータによる自然言語の理解は，30年以上にわたって，人工知能の分野における興味深い研究テーマであったが，これらの研究成果をロボットの行動制御に応用した例は少ない．一方，コンピュータグラフィックス技術の発展により，3次元画像で表現されたソフトウェアロボットを容易に作成することができるようになった．このようなソフトウェアロボットは，マウスを用いて制御することもできるが，自然言語によって制御することが望ましい．本稿では，我々の研究室で開発をすすめているプロトタイプシステムの概要を示し，ソフトウェアロボットを自然言語や音声入力によって制御する際に考慮しなければならない様々な問題を考察する．

キーワード 言語理解，行動制御，ソフトウェアロボット，対話，音声理解

Understanding Natural Language and Controling Robot Actions
— From Speech Recognition to Speech Understanding —

TANAKA Hozumi

Graduate School of Information Science and Engineering,
Tokyo Institute of Technology

2-12-1, Ookayama, Meguro-ku, Tokyo 152-8552, Japan
+81-3-5734-3046
tanaka@cl.cs.titech.ac.jp

Abstract

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-dimentional image to be easily created in a computer and be handled through a mouse device. The 3-dimentional 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.

key words natural language understanding, action control, software robot, dialogue, speech understanding

1 はじめに

最近の音声認識技術の進歩には目覚しいものがある．市販の音声認識システムの多くは、単語連鎖に関する統計情報を利用して，音声を単語の列に変換している．音声認識システムが次々に音声を単語の列に変換している様子を見ていると，あたかも音声認識システムが，話した言葉の意味を理解しているように思える．しかし現在の音声認識システムは，話した言葉の意味を理解していない．音声を単語の列に変換しているに過ぎず，単語の列(文)を作り出す過程で，構文解析さえも行なっていない．人間と機械との会話で音声を用いるためには，これから音声認識システムを音声理解システムへとレベルアップする必要がある．それには，言語理解の機構に関する研究が必要になる．音声認識の研究者と自然言語処理の研究者との共同作業が今ほど求められている時代もないといえよう．

言語理解システムは，1960年代後半から1970年にかけて，MITでWinogradが行なったSHRDLUシステムの研究が良く知られている[Winograd 72]．これはコンピュータの中にシミュレートされたロボットが，自然言語による端末からの指令に応じて積木の世界で動作する．当時から，(自然)言語でロボットの行動を制御する研究の重要性は指摘されていたにもかかわらず，SHRDLUシステム以後，研究の進展は必ずしもはかばかしくなかった．これは，機械的なロボットにしろコンピュータ内部の仮想空間のロボットにしろ，当時はそれらの動作機能が限られていたこと，指令文を端末からいちいち入力する手間が問題であったこと，自然言語処理技術が未熟であったこと，計算機パワーが不十分であったことなどが原因であったといえる．それに加えて，我々の言語理解の機構の解明も不十分であった．

ところがコンピュータグラフィクス(CG)技術の進展により，機械的な制約を受けず，行動機能が豊富な3次元のソフトウェア・ロボットをコンピュータ内部に作りだし，それを自在に動かすことができるようになってきた．自然言語による指令にしたがって動作するロボットは，機械的なロボットに限られなくなってきたのである．これまでの代表的なヒューマンインターフェースとして，マウスがある．ところが，最近マウスに現れるメニューの数が増え，単純な指令であってもマウスをクリックする回数が増えてきている．音声によるマンマシンインターフェース機能の実現が再び求められてきているのは，このような事情による．

一方，すでに述べたように，最近音声認識システムの実用化が進み，音声で指令を与えることが可能になってきた．自然言語処理技術も1970年代と比較して格段の進歩がみられる．SHRDLU時代とは比較にならない程の計算パワーを我々は簡単に手に入れることもできる．以上のことから，音声認識技術と自然言語処理技術を統合した音声理解システム，CG技術，人工知能技術を集大成し，音声によりロボットの行動を制御するシステムを構築する土台が準備されつつあると言ってよい．SHRDLUシステムを越えたシステムの実現に再度挑戦する時期が到来しているのである．

ここで，ロボットといっても機械的なロボットについては，まだ可能な動作についての制約が大きい．そのため，ロボットに行動を指令する自然言語の言い回しを単純なものに制限せざるをえない．本稿で言及するロボットは，当面CGによる仮想空間内の3次元ソフトウェアロボットを仮定する．その方が，複雑な自然言語の言い回しによる指令を出すことができるからである．

本稿では音声を理解し動作する3次元ソフトウェアロボット構築に向けた一つの試みとその意義を説明するとともに，応用例と今後の研究課題を述べる．

2 プロトタイプシステム「傀儡」

2.1 「傀儡」の構成図

現在，我々が構築中のプロトタイプシステム「傀儡」の構成を図1に示す[Shinyama 99]．「傀儡」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い意味理解の結果(Semantic Representation／意味表現)を得る．すなわち，音声認識システムと自然言語処理部がカスケードに結合した構成になっている．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．図1では，このときにどのような知識が使われるのかも図示されている．

2.2 動作例

ここで簡単なプロトタイプシステムの動作例を示して，研究の意義の一端を具体的に説明してみたい．プロトタイプシステムでは，音声による指令にしたがって動作する3つの3次元のアニメーテッドエージェント（馬，鶏，雪ダルマ）[1]が仮想空間内に存在する．また青と赤い玉が二つずつ適当に配置されている．これらの映像をカメラが映し出している．この映像を映し出しているカメラもまたエージェントの一つである．カメラエージェント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いう動作を行なわせると，カメラが映し出す映像が連続的に変化する．

[1]以下では，ソフトウェアロボット，アニメーテッドエージェント，エージェントという言葉を同義と見做して使う．

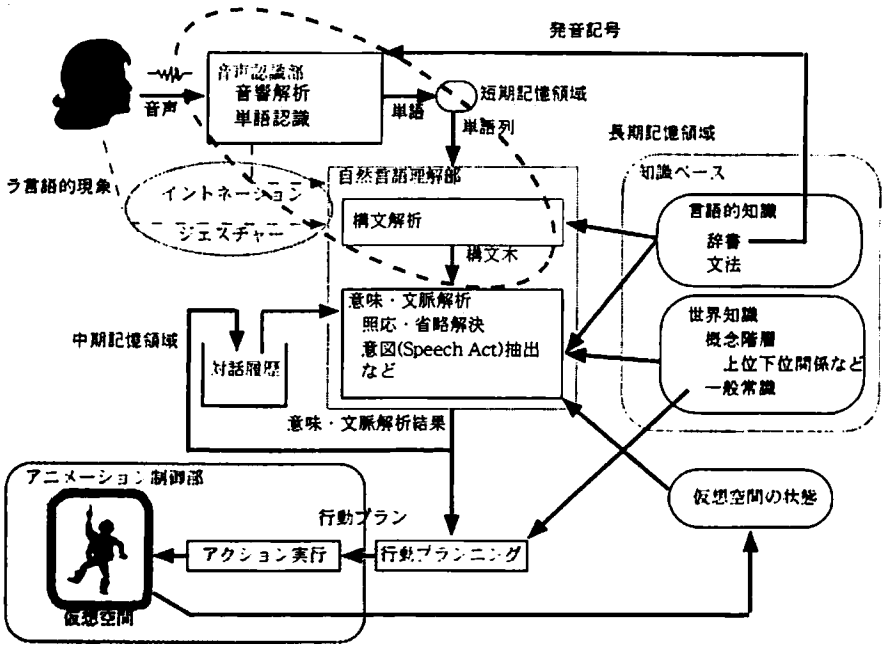

図 1: プロトタイプシステム「傀儡」の構成

指令 1：馬は赤い玉を押せ.

馬は自分の視野を考慮し，視野のもっとも近くの「赤い玉」に近付き，それを押す．この時どの程度の早さでどの方向にどこまで押すかについて，指令は何も直接指示していない．これを言語学者は Vagueness (漠然性) の問題と呼んでいる．システムは，指令の結果を映像化，視覚化しなければならないので，この Vagueness の問題を解決しなければならない.

指令 2：もう少し.

この指令には，いくつかの省略が含まれている．システムはそれまでの脈絡から,「馬」が先の「赤い玉」をもう少し (直前に動かした方向に)「押す」と解釈して，動作映像を作り出す．話し言葉の理解には，脈絡に依存した省略語の推定と意味理解がしばしば必要になる.

指令 3：鶏もそれを押せ.

鶏も，馬が押した「赤い玉」を，馬の押した方向にさらに押して移動させる．指令 3 に含まれる指示代名詞「それ」は先の「赤い玉」と解釈する.

指令 4：カメラは赤い玉に近付け.

先の「赤い玉」の大きさが大きくなる．この映像を作り出すために，カメラの位置変化の量に対する Vagueness の問題を解決しておかなければならない.

指令 5：行き過ぎ.

前文より省略語が「カメラ」であり，その最終位置を基準として,「赤い玉」から「遠ざかる方向に動かす」という意味であると解釈しなければならない．文字通りの意味ではなく，その背後にある発話者の本当の意図を抽出しなければならないのである．これは言語行為論として哲学者らが研究してきたが [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[2].

以上の動作例から，映像化するためには，これまで以上に深い言語理解が必要になることが分かるだろう.

3 研究課題

これから解決すべき具体的な研究課題の主なものを列挙する．大別して 3 つの研究課題がある.「言語理解機構に関する研究」,「行動機能が豊富な 3 次元ソフトウェアロボットの構築」,「言語と行動制御に関する研究」である．やや理解が困難と思われる課題については，補足説明を付記した.

- 言語理解の機構に関する研究
  - 音声理解の研究 (音声認識と構文解析の統合化モデルの構築)

[2] この指令の解釈は，現在のプロタイプシステム「傀儡」にはまだ実装されていない．後述するが，今後の重要な研究課題である.

    * 図 1 で示したような，音声認識の結果を構文解析部に送るカスケード型のシステムではなく，音声認識と構文解析を融合して行なうシステムの構築を目指す．図 1 ではこれを点線で囲んで示してある．
  - 構文解析技術の研究
    * 人間の話す言葉は必ずしも (書き言葉を中心とした) 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 文とよぶが，この種の文の構文解析技術もここに含まれる．Well-formed な文の構文解析技術はほぼ成熟の段階にあると言って良い．(「話し言葉の分析と対話の研究」の項参照．)
  - 意味理解機構の研究
    * 間接言語行為論にもとづく発話意図の抽出の研究や，意味表現形式，言語理解のための知識ベースの構築もここに含まれる．(「言語と行動制御に関する認知理論の構築」の項参照．)
  - 文脈理解の研究
    * 対話状況に応じた省略語の補完，指示語の対象の推定．
  - 言語生成に関する研究
  - 話し言葉の分析と対話の研究
    * 冗長語，繰り返し，割り込み現象などの分析を行なう．
  - 対話理解の計算モデルの構築
- 行動機能が豊富な 3 次元ソフトウェアロボットの構築
  - ソフトウェアロボットの基本動作機能の実現に関する研究

    モーションキャプチャの利用だけでなく，ニュートン力学の世界でさまざまな動作が可能なエージェントの研究を行う．
  - 言語理解とパラ言語現象（身振り，手振り，表情，イントネーションなど）との関係分析
  - 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
    * これは感性情報処理の研究課題でもある．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 (Emotionally Intelligent Virtual Actor と呼ぶこともある [Elliott 98])，これは人工生命の研究とも接点を持つことになる．
  - ソフトウェア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．その実時間映像化についても考えたい．
- 言語と行動の統合に関する理論の構築
  - 言語理解に基づく行動計画の立案と学習に関する研究
    * 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ためには，少なくとも「ドアのところまで歩いて行き，ドアにノブがあればそれを回して引いたり押したりしてドアを開けて外に出る」ということを知っていなければならない．移動経路に障害物があれば，それを避けるという動作も必要になる．こうした一連の行動手順 (行動計画) を知らなければ，それを学習する必要がある．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では無視されてきた重要な研究課題である．
  - ソフトウェアロボットの協調動作に関する研究
    * 複数のソフトウェアロボットに協調して仕事を行なわせる場合には，ロボット相互がコミュニケーションしながら仕事を遂行する必要がある．そのためには，個々のロボットに自律性を持たせることが必要になる．これもまた，行動の側面から重要な研究課題であるが，困難な研究課題になるだろう．
  - 言語と行動制御に関する認知理論の構築
    * 話し言葉には，言葉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促すということがよくある．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であり，ウェイターはそれに答える動作を行なわなければ,「気が利かないウェイターだ」ということになる．この問題は哲学者により，間接言語行為とよばれて研究されてきた．哲学者らはこの問題にトップダウンにアプローチしてきた．

トップダウン的なアプローチにおける問題は，発話のさまざまな現象を網羅的に分析した理論の構築が難しい，ということである．むしろボトムアップにこの問題にアプローチすることも考えるべきであろう．具体的な対話データを収集，分析し，そこから得た知見，理論を実際に動作する計算モデルに組み込み，それを動作させることによりさらに新しい問題の抽出と解決を行なう，というサイクルを考えた研究が望ましいと考えるからである．

- 音声対話によるソフトウェアロボットの行動制御システムの試作

図 2 に本研究課題と他の学問分野との関連図を示す．

4 応用分野

我々の過ごす生活空間は，物理空間から情報空間へとシフトしつつある．情報空間で過ごす時間的な割合が増えている．情報空間へは，好むと好まざるとにかかわらず，コンピュータを介して入ることになる．最近ディジタル・ディバイドが大きな問題になっているのはそのためである．本研究の直接的な目的と応用は，人間と機械との間に自然で柔軟なインタフェース機能を実現することにある．それにより，いわゆるディジタル・ディバイドの問題の解決に寄与し，２１世紀に向けて人間と計算機システムとが共生する社会の実現の基礎として重要な役割を果たすことが期待できる．

工学的な立場から見た本研究の応用例の主なものをあげてみる．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，マイクロマシンで置き換えれば，それらを対話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対話が可能なら，介護ロボットの動作制御が可能になる．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，カメラへの指示を対話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．エンターテイメントの世界では，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．言葉で対話する新しいゲームの世界を作り出すこともできよう．

5 おわりに

80 年代後半から 90 年代前半にかけて，類似の研究が米国のペンシルバニア大学で行なわれている．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（CG 技術）に重点が置かれていたが，本年から言語と行動の視点を取り入れた研究に着手しようとしている．当面は副詞と動作との関係（「激しく動く」など）を扱うこととしており，本格的な言語理解の観点からの研究に着手する段階に至っていない．程度の副詞の理解などは，動作様態に関わるものであるので，特に重要である．

本研究の意義を別の視点から眺めることもできる．我々は言語を通じて，相手に行動を促したり，逆に他人から行動を促されて社会生活を営んでいる．人間の動作や行為は言語と密接に関係している．これまでの言語理解の研究では，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][Mani 99][長尾 96][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は，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．

最近言語学の分野でも，行動という視点から従来の言語学を根本的に見直そうとする動きがある．その一つが認知言語学である．言語を理解したかどうかは，相手の行動をも含めて判断する必要がある．

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できないからである．本研究は，言語理解の結果をソフトウェアロボットの行動をディスプレイ上の映像として見ることができる．そのために深い言語理解を必要とするので，新しい意味でのチューリングテストの場を提供したと見做すこともできる [Badler 99]．

最後に，SHRDLU システムとの主な相違点について述べておきたい．SHRDLU システムでは，構文解析，対話状況に応じた指示代名詞，省略語の補強，推論を含む行動計画の一部が含まれている．その他の問題，たとえば

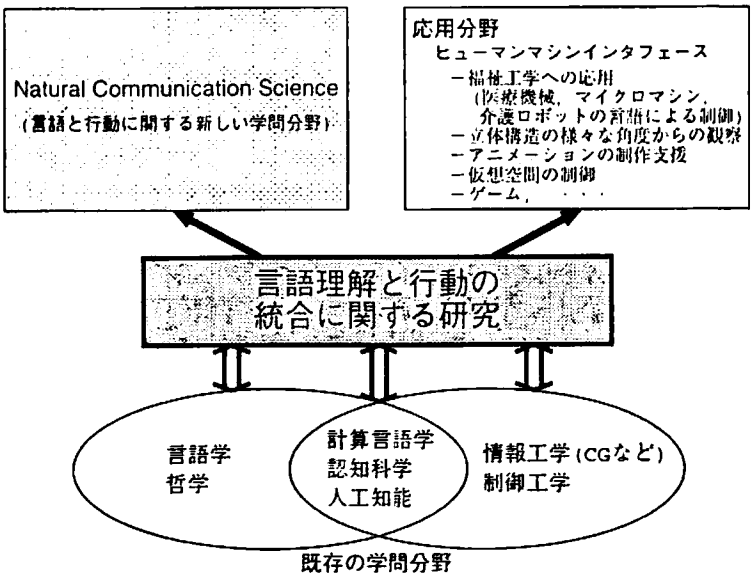

図 2: 他の学問分野との関連図

本研究で取り上げた，音声による対話，言い替えると「話し言葉」を扱っていない．間接言語行為や学習の問題なども扱っていない．SHRDLU を越えた言語理解の研究が求められているのはそのためである．

本研究のように，音声理解，言語理解は，言語学・哲学・認知科学的また工学的にも重要な研究課題であり，文系・理系の枠を越えた学際的な研究組織の下で研究を進める必要があるだろう．

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).

[Elliott 98] Elliott. C.: *Hunting for the Holy Grail With "Emotionally Intelligent" Virtual Actors*. SIGART Bulletin · Summer 1998. pp. 20-28 (1998).

[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).

[Shinyama 99] Shinyama,Y., Tokunaga,T. and Tanaka,H. *Processing of 3-D Spacial Relations for Virtual Agents Acting on Natural Language Instructions*. Proceedings of the 2nd Workshop on Intelligent Virtual Agents, pp.196-206, (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).

[Winograd 72] Winograd,T.: *Understanding Natural Language*, Academic Press (1972).