

# 学術創成研究：言語理解と行動制御 Language Understanding and Action Control

田中穂積\*

Tanaka Hozumi

## 1 はじめに

我々はこれまで、生活空間の大半を物理空間で過ごしてきたが、最近、情報空間で過ごす時間が急激に増大してきている。このような情報空間を、使いやすく豊かな空間にするためには、物理空間と同様、情報空間においても、自然言語を用いた対話が可能なことが望まれる。

このような対話システムの先駆として良く知られているものに、1960年代後半から1970年代前半にかけてMITのWinogradが開発したSHRDLU(ロボット)がある[Winograd 1972]。SHRDLUロボットは端末から入力した英語の指令を理解し、仮想空間内の積木の世界でロボットに仕事を行なわせることができる。指令文に含まれる代名詞の指すものを同定したり、入力文の解釈に曖昧性が生じた場合には、積木の世界の様子を調べて曖昧性を解消すること、積木の世界の操作で障害となるブロックがあればそれを除去してから本来の仕事をおこなうという計画立案能力をもっている。当時のコンピュータ環境を考慮すれば、SHRDLUは画期的なシステムであったといえよう。しかしSHRDLUロボットの動作は単純で、対話文も単純な言い回しに限られていた。対話は鍵盤入力を通じて行なうものであった。理想を言えば、音声による対話が可能なことが望ましい。

現在では、音声認識技術、自然言語処理技術とも当時と比べて格段に向上している。CG技術の進歩にも目覚ましいものがあり、極めてリアルな3次元映像を作り出すことが可能になってきた。表情豊かで複雑な動作を行なう3次元ソフトウェアロボットを、仮想空間内に作るできるようになってきた。機械的な制約の多いハードウェアのロボットと比べてソフトウェアのロボットは、はるかに複雑な動作を行なうことができるので、複雑な言い回しの指令を出すことができる。また、ソフトウェアロボットとの対話は、ハードウェアロボットが進歩すればそれとの対話にそのまま置き換えることができる。我々は、CG技術、音声認識技術、自然言語処理技術を統合して、SHRDLUが扱うことが不可能であった次世代の対話理解システムの構築を目指す時期が到来したと考え、本研究計画を立案した。

## 2 目的

我々が考えているソフトウェアロボットは、先のSHRDLUロボットのように、棒でできたロボットではなく、目、口、顔、手、足をもつlife-likeなロボットである。自然言語による対話によりlife-likeなロボットの行動を制御する研究を通じて、言語理解と行動制御の研究を行なうことを目的にする。そこでは、これまでの研究のように、言語理解を記号から記号への還元としてとらえるのではなく、ロボットの行動として可視化しなければならないことが問題になる。可視化にあたってこれまで以上に深い言語理解の研究が必要になるので

ある。

第一に、音声認識の精度をさらに向上させること、またパラ言語的現象(イントネーションなど音声的現象だけでなく、nonverbalな現象(たとえば、gazing, pose, nodding, 表情なども含む))を考慮する必要がある。第二に、対話であるから状況に依存した言語理解を行なう必要がある。省略されたものがあれば、それらを推測しない限り、ロボットは行動を行なうことができない。指示代名詞についても同様である。第三に、ロボットの行動の可視化に伴う漠然性(vagueness)の問題を解決する必要がある。「もうちょっと右へ行って」という指令は、「どの程度右か」を決めない限り、行動の可視化ができない。第四に、対話の文は話し言葉であり、いわゆるill-formedな文が含まれる。たとえば、「あの、えーと、もう少し前、いや後ろに下がって」などのように、これらをどう処理する考えなければならない。第五に、話し手の真の意図を抽出する必要がある。たとえば、「もうちょっと右に行け」といって「あー、ちょっと行き過ぎ」といえば、それは「すこし左に戻れ」ということを意味している。これは言語哲学の分野では言語行為論として論じられてきた問題である。第六に、ロボットの動作については、行動計画を立てる必要がある。「部屋の外に行きなさい」といわれたら、マイクロには「ドアのところに行く」「ノブがあればそれを回す」「ドアを押しあるいは引いてドアを開ける」「外に出る」を計画する。マイクロには、たとえばノブをどう掴むかという掴み方の計画を立てなければならない。

音声認識、言語理解、CG技術を統合し、以上に述べた問題を解決することが、本研究計画の当面の目的である。言語理解を行動という視点からとらえることにより、言語理解の本質を明らかにすることが最終的な目的である。

本研究の応用として次のものを考えることができよう：ゲームなどentertainmentの分野、介護ロボット、マニュアル教授システム、手話システム、仮想空間のナビゲーション、情報家電....

## 3 組織

工学者、哲学者、言語学者の参加を得て以下の組織で研究を進めている。

- 言語と行動に関する認知理論班  
土屋俊(千葉大)、山田友幸(北大)、辻幸夫(慶応大)、山梨正明(京大)、楠見孝(京大)、丸山直子(東京女子大)
- 言語理解の機構班  
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)、牧野正三(東北大)、河原達也(京大)、古井貞熙(東工大)、鹿野清宏(奈良先端大)、田中穂積(東工大)
- ロボット班

\*東京工業大学 情報理工学研究所 計算工学専攻

北橋忠宏 (関西学院大)、白井良明 (阪大)、小林哲則 (早大)、佐藤誠 (東工大)、中嶋正之 (東工大)、原島博 (東大)、広瀬啓吉 (東大)、小林隆夫 (東工大)

● 研究顧問

長尾真 (京大総長)、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 (日本女子大)

#### 4 「傀儡」システム

コンピュータパワーと最近の音声認識技術、自然言語処理技術、CG 技術の著しい向上により、それらを密に統合し、SHRDLU が扱うことが不可能であった新しい言語理解システムの構築を目指すべき時期が到来したということを目頭に述べておいた。本研究では、そのフィジビリティ研究として、先の三つの技術を密に結合した対話理解システムのプロトタイプ「傀儡」の開発を行なった [Shinyama 2001] [Tanaka 2001][Tanaka 2002]。

「傀儡」は話し言葉を対象としている。特に、話し手の視点と聞き手の視点とを考慮した「右」、「左」、「前」、「後」などといった相対的位置関係を理解し、仮想空間内に存在する四つのソフトウェアロボットに動作(仕事)を行なわせることができる。動作の結果はアニメーションとして見ることができる。ソフトウェアロボットが可能な動作は、「行く」(仮想空間内の特定の場所へ移動する)、「向く」(特定の方向を向く)、「押す」(仮想空間内に配置された玉を、指定した場所まで押す)の三つである。

図1に「傀儡」の構成図を示す。点線で囲んだ枠は、おおまかなコンポーネントである。音声認識部では、音声で与えた話し言葉の指令を認識し、単語の列を出力する。言語解析部では、単語の列からなる文を解析し、仮想空間内の状態を調べて、話し言葉に含まれる曖昧性の解消や照応表現の同定や省略語の補強を行なうとともに、意味表現を生成する。さらに、意味表現から発話の意図を推測する。そして、指令にしたがう動作を実行する手続きを生成し、アニメーション生成部に送る。アニメーション生成部では、その手続きにしたがってロボットに動作をさせ、仮想空間の状態を変更する。

「傀儡」で注目すべきことは、3次元の仮想空間を映し出すカメラもまたソフトウェアロボットであると考えていることである。カメラワークを音声による指令で制御できるのである。それにより、3次元の仮想空間のナビゲーションを音声で行なうことも可能になり、多数の工学的応用例を考えることができることを2.で述べた。

「傀儡」システムで明らかにされた諸問題については、[Shinyama 2001]、[Tanaka 2002] を参照されたい。

#### 5 おわりに

一対多 broadcast モードの対話の研究の重要性についても触れておきたい。ソフトウェアロボット同士の対話の例として、サッカーというスポーツゲームを考えてみたい。サッカーでは、チームメイト同士で競技中に互いに声を掛け合いながらプレイをしている。これは典型的な一対多 broadcast モードの対話になるが、興味深いことは、その対話が敵側の選手にも聞こえることである。敵側の選手は、対話の内容から、先回りしてボールをインターセプトしようとしたり、防御を固めるなどの行動を引き起こす。サッカーなどで行なわれる broadcast モードの対話には、このように人工知能研

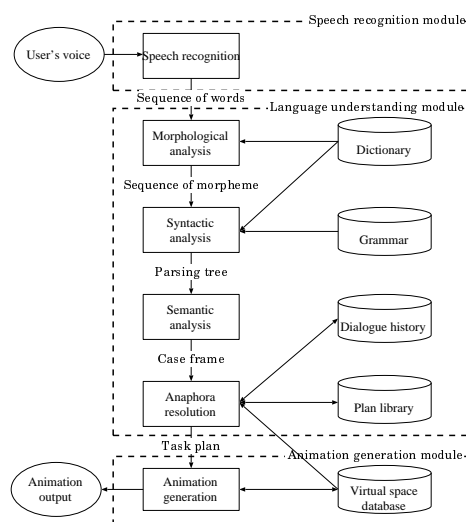


図1: 「傀儡」システムの概要

究の立場からも興味ある研究課題が含まれている。将来ロボカップでも、チームメイト間での声の掛け合いが考慮されることもあるかもしれない。それはまた、broadcast モードの対話に新しい研究の場を提供することにもなるだろう。

#### 参考文献

- Badler, N. I., Phillips, C. B., Webber, B. L., Simulating Humans - Computer Graphics Animation and Control, *Oxford University Press*, 1993.
- Cassel, J. et al., Embodiment in Conversational Interfaces: Rea, *Proceedings of CHI'99 Conference*, pp. 520-527, 1999.
- Rickel, J., Intelligent Virtual Agents for Education and Training: Opportunities and Challenges, in Angelica de Antonio, Ruth Aylett and Daniel Ballin (Eds.), *Intelligent Virtual Agents-Third International Workshop, IVA 2001 Madrid, Spain, September 2001 Proceedings*, Springer, pp. 15-22, 2001.
- Shinyama, Y., Tokunaga, K. and Tanaka, H., 自然言語を理解するソフトウェアロボット: 傀儡, *情報処理学会論文誌*, vol.42, no.6, pp.1359-1367, 2001.
- Tanaka, H., 言語理解と行動制御に関する研究, 平成 12 年度科学研究費補助金 (創成的基礎研究) 報告書, Mar, 2001.
- Tanaka, H., 言語理解と行動制御に関する研究, 平成 13 年度科学研究費補助金 (創成的基礎研究) 報告書, Mar, 2002.
- Weiss, G. ed., Multiagent Systems, *The MIT Press*, 1999.
- Winograd, T., Understanding Natural Language, *Academic Press*, 1972.