

コンピュータが拓く新しい言語世界

田中穂積・徳永健伸

(たなか ほづみ)

(とくなが たけのぶ)

省略の多い発話を的確に理解して求められた仕事をこなすロボット、互いに声を掛け合いながらゲームをするロボットのサッカーチーム。それらを可能にする対話理解システム実現には、どんな課題があるか。

一 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。自然言語を理解するシステム、対話理解システムの実現が求められているのはそのためである。自然言語による対話が可能になれば、情報空間は誰でも利用可能な空間になり、われわれの生活空間の一部として一層身近な存在になる。

月刊
『言語』
Vol. 31, No. 3
02, 3月号

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、一九六〇年代後半から一九七〇年代前半にかけて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こ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しいものがあり、極めてリアルな三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三次元ソフトウェアロボットを、仮想空間内に作ることが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、(ゲームなどで見られるように)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。したがって、複雑な言い回しの指令を出すことができる。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て良い。

二 一九八〇年代、九〇年代の対話理解システム

一九八〇年代に実用化を目指した対話理解システムは、データベースの検索などを、自然言語で行なうものであった。

データベースという極めて狭い世界を対象としていたが、実用化されて現在まで生き延びているシステムは皆無である。データベースの検索ロボットは、先の SHRDLU ロボットとは次の点で大いに異なる。データベース検索ロボットは、言葉で人との間のコミュニケーションを行なう機能を持っているが、SHRDLU ロボットのように、仮想世界に存在する物を動かしたり、自分で歩くなどの機能は持っていない。

一九八〇年代に入り、コンピュータとの対話では、言語の他にパラ言語学的な現象も無視できないことが明らかになってきた。この種の研究では、わが国の研究がもつと注目されて良い。一九八〇年代の初めに北海道大学の戸田正直教授が行なった「ねね」プロジェクトでは、端末上のスクリーンに人の顔を映し出し、システムが行なう応答の内容に応じて、表情を変化させることが考えられていた。野心的な試みではあったが、当時のCG技術の限界を越えていた。そのため、SHRDLUのようなデモンストレーション効果の大きいシステムの構築には至らなかった。わが国では、これは一九九〇年代前半のソニーの長尾確(現名古屋大学)らの研究を経て、マルチモーダルインタフェースの研究へと発展した。マルチモーダルインタフェースの研究では、システムからの応

答に、音声や映像を総動員する。一方仮想現実感の研究では、システムと人との間で交わされるアイコンタクトの問題も対話の場面で取り込むことが議論されている[Rickel 2011]。

一九八〇年代には、適切な応答文生成の立場から、対話する人の知識を推測して、専門家であれば専門家なりの、専門家でなければ、冗長であつても丁寧な応答をするユーザモデルの研究も行なわれた。

三 次世代対話理解システム

対話により仮想空間で仕事を行なうソフトウェアロボットが解決すべき研究課題を見るために、次の対話例を考えてみよう。それにより、これまで以上にロボット側に深い言語理解が求められることが分かる。

- (1) 人：「窓のカーテンをちょっと開けて下さい」
- (2) ロボット：「窓のところに行き、カーテンを掴んで引いて開ける」
- (3) 人：「もう少し」
- (4) ロボット：「もう少し開ける」
- (5) 人：「開け過ぎだよ」

- (6) ロボット：「カーテンを少し閉める」
- (7) 人：「少し部屋の空気が汚れてきたね」
- (8) ロボット：「閉まっている窓に手をかけ、窓を開ける。」

最初の指令では、ロボットが窓のカーテンを開けるためには、窓のところに行き、カーテンを掴んで引いて開けるという計画を事前に立てて、それに従った行動を起こす必要がある。それをマクロな行動計画とよぶことにする。カーテンの掴み方にもさまざまな方法がある。カーテンを掴むためには、手をあらかじめ開いて、カーテンを掴み、そのままカーテンを引いて開けるという計画を立ててそれを実行する。これをミクロな行動計画とよぶことにする。「窓のカーテンを開ける」という指令は、このようなマクロとミクロな行動計画とその実行をとまなつてはじめてロボットが動作し、言語（指令）を理解したとみなすことができる。

このように、ロボットとの対話では、可視化されたロボットの動作から、ロボットが言語を理解したかどうかを、我々は直ちに判断することができる。言語理解と行動とを表裏一体の関係をなすものとしてとらえることは、認知言語学が目指す研究の方向とも一致している。行動を含めた言語理解の評価基準は、人工知能が実現したかどうかを知るためのいわ

ゆるチューリングテストよりも厳しい評価基準となる。チューリングテストでは、壁の向こうに人と人工知能を置いて両者の行動を比較するので、壁の向うで行なわれる行動を観察するわけにはいかないからである。

ロボットの行動を可視化することにより、「カーテンをちよつと開ける」にしても、どの程度開けたら良いかというvaguenessの問題を解決しなければならなくなる。これもまた、これまでの対話理解システムでは見過ごされてきた問題である。

第二の「もう少し」という文には、主語も目的語も動詞も欠けている。前文よりわれわれは省略語の推測を行ない、「あなた（ロボット）は、もう少しカーテンを開けてください」という指令の文であると解釈できる。対話で使われる話し言葉は、書き言葉以上に省略を多用する。省略語の推測だけでなく、「それ」「これ」といった照応表現の指すものも同定も必要になる。

第三の「開け過ぎだよ」という文にも先と同様な省略語の推測が必要になるが、文の意図することは、文字通りの意味ではなく、「少しカーテンを閉めてくれ」という意味に解釈しなければならぬ。これは、言語行為論の中で間接言語行

為として論じられてきたものに他ならない。対話には間接言語行為がしばしば含まれるので、その意図を見誤ると、気の効かないロボットであるということになる。

第四の文も、典型的な間接言語行為である。常識的には、その意図が「窓を開ける」ことであるとロボットは解釈し、それに沿った行動計画を立てて実行しなければならぬ。

以上をまとめると、次世代対話理解システムでは、困難ではあっても、省略語の推定、照応指示関係の同定、間接言語行為などの意図推測を行なう研究に本格的に取り組む必要がある。次節で説明するように、次世代対話理解システムでは、複数ロボットを相手にした対話も扱う必要がある。

四 複数ロボット相手の対話

これまで説明してきた対話システムは、すべて人とロボットとの間の一対一の対話であった。しかし、対話は一対一の対話に限られるわけではない。講義に見るように、教師と学生との間の一対多の対話などは、日常頻繁に経験するにもかかわらず、これまでの対話システムではまったく取り上げられていないし、研究もなされてきていない。人と複数ロボット相手の対話を以下では「一対多 Broadcast モードの対話」

とよぶことにする。一対多 broadcast モードの対話の特徴は、対話の内容が聞き手全員に聞こえることである。

一対多 broadcast モードの対話で注意すべきことは、おそらく人からの指令が主語を欠き、対話の相手が直接分らないことだろう。人からの指令がどのロボットに向かつてなされたかが分からなければ、ロボット間で動作の衝突、混乱が生じる。対話当事者が誰であるかを各ロボットが知ることにより、こうした混乱を避けることができるので、対話の相手を推論する問題は極めて重要である。これは、一対多 broadcast モードの対話に固有の新しい問題である。一対一の対話では対話の相手は一人に限られるので、この問題は生じない。今後、一対多の対話例を大量に収集して、この種の新しい問題の発掘と分析を進める必要がある。

工学的な視点から一対多 broadcast モードの対話モデルの意義について考えてみよう。直ちに考えられることは、仮想空間内に複数のロボットの存在するゲームなどのエンターテインメントの世界への応用である。動画の作成にも応用可能である。仮想空間内の複数のロボットのそれぞれを演技者と考えれば、あたかも我々が監督者になった気分演技者に指令を出して演技をさせることも可能になるからである。

五 「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、これらを密に統合し、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冒頭に述べておいた。そこで、われわれは最近、新しい研究プロジェクトを立案し、そのフィー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]。

「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」（特定の方向へ向く）、「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。

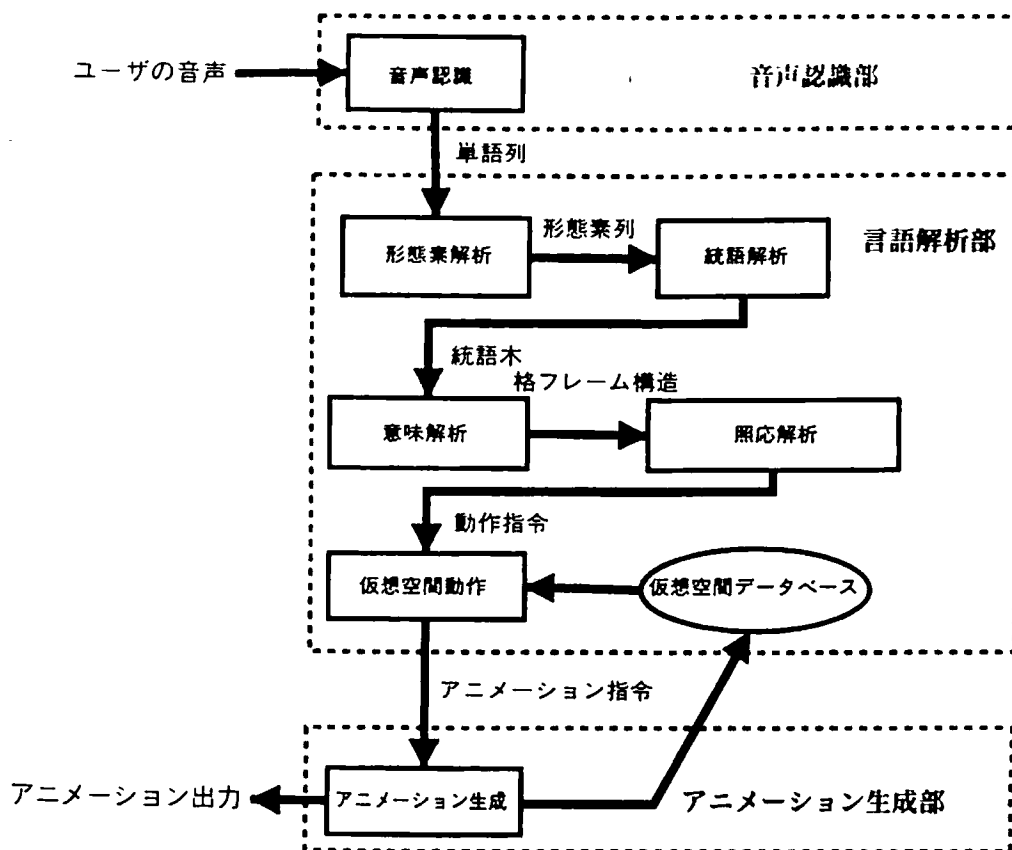


図1 「傀儡」システムの構成

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、三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、三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よう。

「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 2001]を参照されたい。

六 おわりに

自然言語による対話理解の研究を概観し、次世代の対話システム構築に関する研究課題を説明した。一対多 broadcast モードの対話の研究の重要性についても触れた。しかしそこでは、人とソフトウェアロボットとの対話であり、ソフトウェアロボット同士の対話は考えられていない。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究の立場から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】

- [Rickel 2001] Rickel, J.: "Intelligent Virtual Agents for Education and Training: Opportunities and Challenges," in Angelica de Antonio, Ruth Aylett and Daniel Balin (Eds.), *Intelligent Virtual Agents - Third International Workshop, IVA 2001 Madrid, Spain, September 2001 Proceedings*, Springer, pp. 15-22, 2001.
- [Shinyama 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 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」 二〇〇一年三月
- [Winograd 1972] Winograd, T.: *Understanding Natural Language*, Academic Press, 1972.

(田中穂積／東京工業大学／計算言語学)
(徳永健伸／東京工業大学／計算工学)