

## 【招待講演】状況を考慮した言語理解にむけて

徳永 健伸†

† 東京工業大学 大学院情報理工学研究所  
152-8552 目黒区大岡山 2-12-1  
E-mail: †take@cl.cs.titech.ac.jp

あらまし 本稿では、自然言語処理研究の変遷を振り返り、我々の取り組みを例にあげて今後の展望について述べる。特に言語が使用される際の状況を考慮することが次世代の言語理解システムには重要であることを述べる。

キーワード 自然言語理解, 状況依存性, 対話システム

### [Invited talk] Toward Situated Language Understanding

Takenobu TOKUNAGA†

† Department of Computer Science, Tokyo Institute of Technology  
Meguro Ōokayama 2-12-1  
E-mail: †take@cl.cs.titech.ac.jp

**Abstract** This article reviews the progress of natural language processing research and outlooks the future research direction through introducing our research project.

**Key words** natural language understanding, situatedness, dialogue system expressions, Japanese

#### 1. 自然言語理解研究の変遷

自然言語処理の研究は、コンピュータが発明されるとほぼ同時に始まった。1957年のスプートニク・ショック以来、人工衛星の開発でソ連に遅れをとった米国がコンピュータを使ってロシア語の科学技術文献を英語に翻訳する可能性を摸索していた。当初は、辞書を作り、単語を置き換えることで比較的簡単に機械翻訳を実現できるという楽観的な見方がされていたが、現実はその簡単ではなかった。1966年に提出されたALPACレポートによって、より基礎的な研究に研究資金が投じられるようになった。これが自然言語処理や計算言語学と呼ばれている研究分野の本格的な始まりである。

1960年代、1970年代は対象領域を限定し、その対象領域に関する知識を手で構築するアプローチが中心であった。データベースに対する質問応答をするような対話システムもいくつか作成されたが、いずれも小規模なものであった。後述するSHRDLUなどもこの時期に作られた対話システムのひとつである。

1980年代に入って、機械翻訳に代表される実用的な言語処理システムが構築されるようになった。対象領域を実用的な規模に拡大するために、言葉の意味にまで立ち入った深い解析ではなく、統語的な構造の操作などの浅い解析が中心的なアプローチであった。処理に必要な知識は依然として人手によって構築

していた。この時期には、日本のコンピュータメーカーによる商用の機械翻訳システムが市場に現われた。

自然言語をコンピュータで処理するには様々な知識をコンピュータ上に構築する必要がある。知識の記述という観点からすると、それまで手続き的に記述していた知識を、宣言的に記述しようとするアプローチが主流となったのも1980年代である。折しも日本では第5世代コンピュータの開発プロジェクトがおこなわれており、知識を宣言的に記述する考え方は、このプロジェクトで採用された論理プログラミングのパラダイムに合致したものであった。言語学の分野でも、この時期、米国西海岸を中心にGPSG、LFG、HPSGなどの制約に基づく言語理論、あるいは単一化に基づく言語理論と呼ばれる理論が提案されている [1]。

知識を宣言的に記述しようとする試みは、知識の構築・管理のコストを少しでも抑えようとする動機から始まっている。1990年に入ると、これまで人手で記述してきた知識を大量の言語データから自動的に抽出しようとする研究が盛んになる。いわゆる統計的自然言語処理と呼ばれる研究の流れである。この背景には、コンピュータの記憶容量の増大と処理能力の向上というハードウェア的な要因と電子化テキストが普及したというソフトウェア的な要因がある。この時期に爆発的に普及したインターネット、特にWebの発展はこの電子化テキストの普及に拍車をかけた。

もうひとつの要因として統計的手法、特に HMM による音声認識技術の成功があげられる。音素列から文字列への翻訳において成功したモデルを単語列から品詞列の翻訳に適用することによって、言語処理においても知識を手で記述することなく、訓練データを用意するだけで、高精度の品詞付与をおこなう手段を手にすることができた。結果として、形態素・統語処理に関しては一般的に利用できるツールが多く存在し、その解析精度もアプリケーションによっては十分実用に耐えるほどに向上している。統計的手法を意味処理、談話処理にも適用しようとする試みはあるが、残念ながら、これらの処理については十分な精度のツールが手軽に使えるような状況には至っていない。

応用システムについていえば、情報検索、情報抽出、自動要約、QA システムといった新しい応用分野に言語処理技術を適用しようとする試みが始まった。また、対話システムについても音声言語処理技術と組み合わせて音声対話システムという形で研究が盛んになった。以下、本稿では特に対話システムに焦点を当て、人間とコンピュータのインタラクションという観点から自然言語処理研究の課題と展望について述べる。

## 2. 対話システムの例

自然言語によって人間とインタラクションするシステムは対話システムと呼ばれ、これまでも多くの研究がおこなわれてきた。自然言語処理の研究の歴史の中でもっとも初期に作られ、まったく異なる方向性を持っている代表的なシステムに ELIZA [2] と SHRDLU [3] がある。

ELIZA は Weizenbaum が作成した精神科医を模倣するプログラムである。以下は ELIZA の対話の例である。

人間	: Men are all alike.
ELIZA	: In what way?
人間	: They're always bugging us about something or other.
ELIZA	: Can you think of a specific example?
人間	: Well, my boyfriend made me come here.
ELIZA	: Your boyfriend made you come here

ELIZA で使われている技術は、あらかじめ用意したキーワードを手がかりにした単純なパターンマッチであり、人間の入力を理解しているわけではない。入力中に用意したキーワードに合致するものがなければ、この例の最後の ELIZA の発話のように、人間の発話をおうむ返しにして人間がキーワードを含む別の発話をするのを待つ。ELIZA は、このように単純なメカニズムのプログラムではあったが、被験者を使った実験によると、Weizenbaum の意図に反して、人間が真剣に ELIZA と対話をする現象が観察された。これは、精神科医のカウンセリングという、特に達成すべき明確な目的がない対象領域を設定していることが主な理由であると考えられる。最近ではシーマンなどのビデオゲームにも同様の手法が使われているし、emacs エディタの doctor コマンドも ELIZA の派生である。

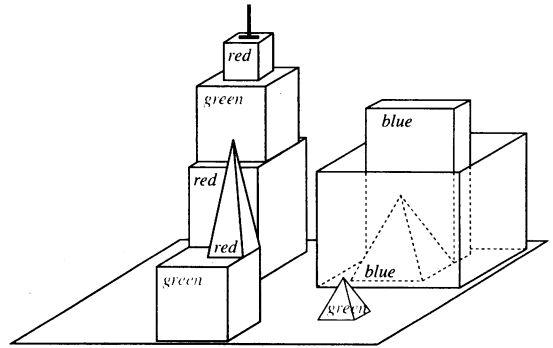


図 1 SHRDLU の積木の世界 [3]

1970 年代に Winograd によって開発された SHRDLU は、様々な色や形の積木の世界において、人間がロボットアームに自然言語で指令を出し、世界の状態を変えることができるシステムである。人間は世界の状態や過去の履歴に関してシステムに質問をすることもできる。対話する人間に特に明確な目標がない ELIZA に対し、SHRDLU では、人間が意図する通りに世界の状態を変化させるという目的がある。SHRDLU は当時として考えうる自然言語処理のほとんどの処理を網羅し、多くの言語現象を扱うことができた対話システムである。また、グラフィック・ディスプレイを備え、世界の変化を動画として確認できるという点でも画期的であった。

言語によるインタラクションという観点からみると、これらの 2 つのアプローチの用途が異なることは自明であろう。暇つぶしや娯楽のためのシステムなら ELIZA 的アプローチで十分かもしれないが、ある特定の目的を達成するためには、SHRDLU のように話者の意図を把握し、それに応じて適切な動作、あるいは発話をする必要がある。

ELIZA や SHRDLU 以降も多くの対話システムが研究されているが、残念ながら SHRDLU のような総合的なものは少なく、SHRDLU が扱っていなかった特定の言語現象に焦点を当てたものが多い。特にデータベース中の情報を対話を通じて検索したり、旅行計画の立案をしたりする、いわゆる情報探索型の対話システムが最近の対話システム研究の中心である。情報探索型の対話にも、システムから質問に対して質問で答える問い返しや、すでに決定した事項の変更に関する対話など、興味深い現象は多いが、SHRDLU のように人間とシステムが世界を共有できる類のものはほとんどない。

また、マルチモーダルな対話システムも多く研究されているが、対話システムの中でグラフィカルな情報を言語表現と効率的に組み合わせることや、システムを擬人化して表情を持たせることによる心理的な効果を狙うものなどが中心で、人間とシステムが世界を共有するためにグラフィカルな情報を利用しようとするシステムは非常に少ない。現在のコンピュータ・グラフィックスの進展を考慮すると、人間と世界を共有できる対話システムの研究をおこなう環境が整いつつあるといえる。

### 3. 言語理解と身体性

例として部屋の家具の配置を相談している以下の会話を考えよう。

- |                                    |
|------------------------------------|
| (1) 夫： このソファは君の右の壁の方がいいんじゃない？      |
| (2) 妻： (振り返って) この出窓のあたり？           |
| (3) 夫： (場所を指差しながら) いやその辺。          |
| (4) 妻： ちょっとここには入らないんじゃないの。         |
| (5) 夫： (子供に向かって) メジャーをお母さんに取ってあげて。 |
| (6) 子： どこにあるの？                     |
| (7) 夫： 机の引き出しにあるだろ。                |
| (8) 子： あった。                        |
| (9) 夫： それをお母さんに渡して。                |

このような会話を対話システムによっておこなおうとしたら、システムには少なくとも以下のような能力がなければならない。

- a) 相手の位置やオブジェクトの位置を把握し、前後左右の空間的な位置関係を正しく理解できること。

この例のような空間的な関係を理解するためには適切な参照枠 [4] の設定が必要となる。たとえば、2人が向い合って「君の右」といった場合、話者から見て「君の右」なのか、聴者から見て「右」なのか、参照枠の設定によってはまったく逆の場所を指すことになる。

- b) 言語情報とパラ言語情報<sup>(注1)</sup>を統一的に扱えること。

単に言語表現だけでなく、言語表現と視線の移動や指差し動作との同期を考慮しないと、(2)や(3)のような発話は正しく理解できない。

- c) 漠然性が扱えること。

ソファを置く正確な場所がどこであるのかは、言語表現の上では漠然としているが、実際にソファを置く際には厳密な位置を決定する必要がある。

- d) 協調作業ができること。

この例では対話を通して複数の人間が協調的に計画を立てている。また、実際にソファを移動しようとするれば複数の人間が協調して動作をしなければならない。

この例からわかるように、システムとの対話を実現するためには、システムがその世界の状況に根ざしていなければならないことがわかる。これまでのように言語を単に記号として扱い、言語処理が閉じた世界の記号処理に留まっていたのでは、このような対話は扱えない。このように記号に実世界に根ざした意味を与えることは記号のグラウンディングと呼ばれている。また、これは、近年、認知科学や人工知能の分野で、「身体性」というキーワードをもって語られる概念と関係がある。身体性を重要視する研究者は、身体をもたない知能はありえず、知能とは世界とのインタラクションによってはじめてもたらされると主張している。

(注1)：ここでは、音響的な情報だけでなく、発話にともなうジェスチャ、表情などの非言語的情報も含む。

### 4. 状況に即した言語理解のための研究課題

前節では、対話システムが状況に根ざした対話をおこなうために必要な能力について述べた。ここでは、それをさらに詳細化し、状況に即した言語理解をおこなうシステムを実現するための研究課題について述べる。

#### 4.1 空間的關係の把握

前後左右や上下など物体間の空間的な関係とその言語表現の間関係については哲学や認知科学などの分野で数多くの研究がある。たとえば、互いに向い合っている状況で「君の右」と言った場合、話者の視点に立つのか、聴者の視点に立つのかで、解釈が逆になってしまう。これは単に視点だけの問題ではなく、参照物体(上の例では「君」)自身に方向性があるかないかなど、様々な要因が言語理解に関係する<sup>(注2)</sup>。

このように空間的な関係の解釈を絞り込むために制約を設定することを参照枠を設定するという[4]。参照枠を決めるモデルは認知科学の立場からいくつか提案されている。たとえば、Levelt [5] は座標系と参照物がそれぞれ話者であるか話者以外であるかによって、参照枠を3種類に分類しているし、Retz-Schmidt [6] は、参照物自身が方向性を持つかどうかという要因と視点からやはり3種類に分類している。

このように認知科学の研究は参照枠を分類することに主な興味があるが、その具体的な手続きについては教えてくれない。その他にも、Herskovits は参照枠の分類よりも、それを決定する要因を中心にこの問題を整理し、座標系の原点、軸の順序、軸の方向の3つの要因によって参照枠が決定されるとし、最初の2つの要因については決定方法を述べているが、肝心の軸の方向の決定については明確な答えを出していない[7]。実世界あるいは仮想世界の上で実験システムを構築し、これらの要因について実証的に明らかにしてゆくことは工学の役割であろう。

#### 4.2 パラ言語情報

人間同士の対話では、暗黙のうちに多くの情報が言語表現以外の手段によって伝わっている。たとえば、表情や手の動き、視線、あるいは声の調子などの音響的な情報は言語で表現された情報を補完する役割を担っている。音声の韻律情報も言語を理解する上で重要な手がかりとなる。たとえば、文字列で表現すれば同じ発話でも文末のイントネーションが上がるのか下がるのかによって、疑問文になったり宣言文になったりする。Cassellら[8]は話しをする人間の動作をビデオに撮影・分析し、人間と同じような振舞いをソフトウェアエージェントで実現する試みをおこなっている。

また、自然な視線の動きの実現、コンピュータアニメーションにおいて発話と口唇の動きを同期させる口唇同期(Lip Sync)と呼ばれる技術、表情の生成などは、人工知能における仮想エージェントの研究分野では活発に研究されているテーマである。たとえば、長尾らは実際に音声対話システムに表情生成を組み込み、音声認識に失敗したことを、顔をしかめることによってユーザに伝える実験をおこなっている[9]。その結果、音声認

(注2)：「君」を「ボール」に置き換えるとこのような曖昧性は生じない。

識の失敗を言語表現で伝えるより、ユーザの主観的な評価は改善されたと報告している。これらの研究は、コンピュータグラフィックスや音声認識の技術の進歩を前提としており、最近のこれらの研究分野の進展によって始めて可能になったものである。

#### 4.3 曖昧性と漠然性

曖昧性は言語解析の様々な段階で問題になる。たとえば、前節の親子の対話例において、発話(9)で使われている代名詞の「それ」が「メジャー」を指しているということは、人間ならばすぐにわかるが、コンピュータでこれを同定するのはそれほど容易ではない。テキスト中の既出のオブジェクトにはメジャーの他にも机や引き出しやソファなど「それ」で指されそうなものはいくつもある。これは参照の曖昧性と呼ばれ、これを解決する処理は照応の解消と呼ばれている。

この例は指示対象を対話を書き起したテキストの中に見つけることができるが、その場の状況を考慮しないと解消できない照応もある。このような照応は一般に外界照応と呼ばれる。たとえば、以下のような対話を考えよう。

客:	(大根を指差して)これ、ください。
八百屋:	一本でいいの?
客:	はい。
八百屋:	じゃ、200円ね。

「これ」の指示対象は、その場の状況には存在するが、この対話を書き起したテキスト中には現われない。この例でもわかるように、外界照応は言語を記号の中に閉じた系として考えていたのでは扱えず、言語と状況との関連を考えて始めて浮き彫りとなる問題である。

言語処理の研究において言語の曖昧性は中心的な課題であり、多くの研究がおこなわれてきたのに対して、漠然性に関する研究は非常に少ない。上述した参照の曖昧性のように、一般に言語処理における曖昧性の解消は、多くの候補の中から正しいものを選択する離散的な過程としてとらえることができる。これに対して漠然性の解消は数え上げることができない候補からもっともらしい答を見つける連続的な過程であるといえる。たとえば、前節の対話例の発話(3)においてある場所を「その辺」という表現で指示しているが、この場合、指示されている場所を数え上げてその中から正解を選ぶという処理は適切ではない。もちろん最終的に「その辺」に物を置く場合には、正確な位置は一意に決まることになるが、対話の中でやりとりされる「その辺」という表現が指示する場所にはかなりの幅が許容されている。

ここで興味深いのは、「その辺」という言語表現が記号的なものであるのに対して、それが指示する場所は連続的な広がりがあるという点である。このようなミスマッチは、言語を記号の中に閉じた系として考えている限り現われてこない問題である。この例からもわかるように、言語を使用するシステムを実/仮想世界に置いて、そのシステムとの間で言語や行動を通じてインタラクションしようとする、このような漠然性を扱うことが不可欠となる。

このような問題は古典的な人工知能の分野では記号処理を基

礎としたプランニングによっておこなわれてきた。「その辺」という表現を単に記号として扱っている限り、古典的な手法は使えるかもしれないが、この例のように特定の状況の中で、言語表現から具体的な空間的な位置関係を計算しようとする連続的な座標系の計算も不可欠となってくる。このためには従来の古典的な記号処理と空間座標のような連続量のギャップを埋める枠組が必要となる。

#### 4.4 協調作業

複数の人間が協力して作業をおこなうためには、お互いが共通の基盤を共有する必要がある。前節の対話例において夫が子供にむかって「それ(メジャー)をお母さんに渡して」と言っているが、子供がメジャーを母親に渡すためには、子供が「渡す」動作をするだけではなく、母親の方も同時に「受け取る」動作をしなければならない。この例にはもうひとつ興味深い点がある。発話(9)は表面的には子供に向けられたものであるが、実際にはその場にいる母親にも聞こえていて、母親に対するメッセージも込められている。多くの対話システムでは1対1の対話を扱うものが多いが、このように複数の人間が同じ場を共有するような例では、ひとりの発話が必然的に複数の人間に聞こえることになる。このような場合、特定のひとりに対するメッセージの場合もあれば、複数、あるいは全員に対するメッセージの場合もある。誰に対するメッセージなのかは状況に依存し、これを判断することが必要となる。

### 5. 次世代自然言語理解システムへの展望

身体性を持ち、言語を通して人間とインタラクションできるエージェントの研究が近年注目を集めている[8]。これらは「身体を持つ会話エージェント(embodied conversational agents)」と呼ばれている。これらの研究で重要な点は、単にコンピュータグラフィックスによって精緻なアニメーションを生成するだけではなく、その多くが言語能力を重要視していることである。

このような研究分野は、必然的に、コンピュータグラフィックス、音声言語処理、計算言語学、認知科学、哲学、言語学などの関連分野を含む学際的なものとなる。著者らのグループでもこれらの関連分野の研究者を組織し、2001年度から5年間の予定で「言語理解と行動制御」という研究題目で研究をおこなっている(文科省科学研究費補助金 学術創成研究 13NP0301)<sup>(注3)</sup>。以下、我々のプロジェクトで具体的におこなっている研究課題について紹介し、次世代の自然言語理解システムへの展望とする。

我々のプロジェクトでは、これまで記号の世界に閉じておこなわれてきた言語理解の研究を、実/仮想世界とのインタラクションを導入することによって、より状況に根ざした言語理解に発展させることを主な目的としている。特に言語理解の結果として生じるエージェントの行動を重要視している。ただし、これは単に言語を解析した結果を視覚化するというだけの意味ではない。

Austin や Searl らの言語行為論に見られるように、言語の使

(注3) : <http://www.cl.cs.titech.ac.jp/sinpro>

用(発話)も行為の一種であると考えることができる。逆に発話に対して、物理的な動作や音調などのパラ言語的な手段によって対応することもできることを考えると、ある種の行動は言語の使用と同類であるともいえる。このように人間の活動において言語と行動は密接な関係にあるにもかかわらず、これまで言語処理は言語を閉じた記号系として扱い、ロボティクスでは行動を単なる制御系の問題として扱ってきた。知的なエージェントを実現するためには、言語と行動を統一的に扱う必要があると我々は考えている。

この目的を達成するために、我々は仮想世界中に存在する複数のエージェントと音声対話によってインタラクションできるプロトタイプシステム  $K_2$  を作成し、これをテストベッドとしていくつかの研究テーマに取り組んでいる。

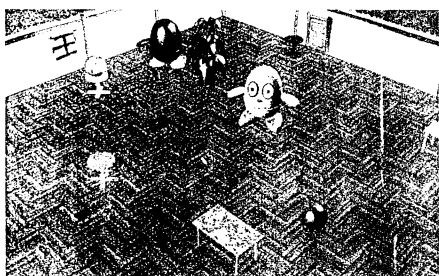


図2  $K_2$  のスクリーンショット

図2はプロトタイプシステムのスクリーンショットである。この図では、仮想空間中に黄色と黒色の2体のエージェントと色のついた机とボールが置かれている。人間は音声入力によってエージェントに指令を出し、エージェントはそれにしたがって世界の状態を変化させる。エージェントの行動と世界の変化の様子はアニメーションによって人間に提示される。

現在、このプロトタイプシステムを使って以下の項目について研究をおこなっている。

- (1) 音声入力における言い直しの扱い [10]
- (2) 状況情報とプランを利用した照応の解消 [11]
- (3) 空間的な漠然性の表現とそれを利用した行動計画 [12], [13]
- (4) 構成的な動作の辞書の構成 [14]

以下、各項目について概略を述べる。詳細については各文献を参照して欲しい。

### 5.1 音声入力における言い直しの扱い

キーボード対話に比べると音声対話では言い直しが多く観察される。頑健な音声対話システムを実現するためには、言い直しを扱う必要がある。英語の言い直しを扱うモデルとして Nakatani と Hirschberg の Repair Interval Model (RIM) [15] がよく知られている。我々は日本語の言い直しの現象を分類整理し、日本語でよく観察される倒置現象も扱えるように RIM を拡張し、システムに実装している。

言い直しを処理するための機構はスタックを用いた文節間の係り受け解析をおこなうパーザに埋め込まれている。言い直しの処理は係り受け解析と同時平行的におこなうので、言い直さ

1. [(うどん)>
2. [(うどん)|(いや)>
3. [(うどん)|(いや)|(そば)>  
(言い直しの処理:「うどん」を「そば」に置換)
4. [(そば)>
5. [(そば)|(食べた)>  
(係り受けの処理:「そば」を「食べた」にかける)
6. [((そば)食べた)>
7. [((そば)食べた)|(彼と)>  
(係り受けの処理:「彼と」を「食べた」にかける)
8. [((そば)(彼と)食べた)>
9. [((そば)(彼と)食べた |(えつと)|(温かい)|(やつを)>  
(係り受けの処理:「温かい」を「やつを」にかける)
10. [((そば)(彼と)食べた |(えつと)|(温かい) やつを)>  
(言い直しの処理:「そばを」と「やつを」に置換)
11. [((そば)(彼と)食べた)>

図3 言い直しの解析例

れる部分と言い直している部分の間の統語的・意味的構造を比較し、より正確な修正が可能となっている。例として、以下の入力を解析する様子を図3に示す。

うどん いや そばを 食べた 彼と えつと 温かい やつを

図3において、“|”がスタックの底，“|”が文節の区切り，“>”がスタックトップを表わす。言い直しの検出は、「いや」や「えつと」などの編集表現の除いたスタックトップの2つの要素を構文的・意味的に比較することによっておこなう。ただし、言い直しの検出は決定的にはできないの実際には複数の仮説を保持し、すべての仮説について処理をおこなう。

### 5.2 状況情報とプランを利用した照応の解消

照応解消に関する研究は盛んにおこなわれてきたが、その多くはテキスト中の文脈照応を、主に表層に表われる言語的手がかりによって解消するものであった。たとえば、センタリング理論では、談話の話題がテキスト中の「センター」と呼ばれる要素によって維持され、センターが照応や省略の先行詞になりやすいと仮定している [16]。各要素がセンターになる可能性は各要素の文法的な役割によって優先順位が付けられ、センターは談話の進行とともに、いくつかの制約によって管理される。

センタリング理論は対話にも適用されているが、音声対話では発話の境界が明確でないこと、いわゆる「文」を構成しない断片的な発話が多いことなどが問題となる。特に断片的な発話では多くの省略が生じるので、表層的な手がかりだけでは不十分なことが多い。日本語ではテキストにおいても多くの省略が生じることが知られており、亀山らはセンタリング理論を日本語の省略解消にも適用しているが、音声対話における省略とテキストにおける省略を同等に扱えるかどうかについては十分な研究がされていない。

我々の想定している対象領域では、発話中には多くの省略や照応が使われ、その場の状況を参照しないと、これらの発話を理解することは難しい。ここで扱っている世界は単純で、エージェントの行動も限定されているので、センタリング理論のよ

うな表層的な手がかりを利用した手法でも照応が解消できることも多い。しかし、以下の例でわかるとおり表層的な手法には明らかな限界がある。

(1-1) 黒いエージェントは赤い球を押して

(1-2) 青い球の前に移動して

(1-3) (青い球を) 押して

発話 (1-3) で省略されている目的語は「青い球」であり、これはセンタリング理論でも扱える。しかし、最初と最後の発話の動詞だけを変えた以下の例では、3番目の発話で省略されている要素は変わってくる。

(2-1) 黒いエージェントは赤い球を取って

(2-2) 青い球の前に移動して

(2-3) (赤い球を) 置いて

文法的な構造や各要素の役割は例 (1-1)–(1-3) も (2-1)–(2-3) も同じなので、この違いは表層的な手がかりだけでは説明できない。

これらの発話の背後にあるユーザのプランを考慮するとこの例は以下のように説明できる。あるオブジェクトを「置く」ためには、前提としてそのオブジェクトを持っていないなければならない。例 (2-1)–(2-3) の場合、発話 (2-1) によってエージェントは赤い球を持つことになり、(2-3) で置く対象となるのは現在持っている赤い球であると考えるのが自然である。一方、例 (1-1)–(1-3) の場合、一般にあるオブジェクトに対して操作をするためにはその近くに行く必要があるという知識があれば、青い球の前に移動するのはそれを押すための動作であると理解できる。

我々が対象としているシステムでは、ユーザは自分のゴールを達成するためにシステムと対話する。特に Cohen が指摘しているように、音声対話では、発話が断片化しやすく、いくつかの発話によってあるゴールを達成する傾向がある [17]。したがって、ユーザの発話はゴールを達成するプランの一部に対応していることが多く、これらの関連したプランの一部の中では照応や省略が起りやすくなると考えられる。このような観察に基づき、我々のシステムでは、ユーザの発話の背後にある部分プラン (プラン演算子) を同定し、その部分プラン同士の前提条件と効果の関係に基づいて照応を解消する手法を採用している。

### 5.3 空間的な漠然性の表現とそれを利用した行動計画

すでに述べたように古典的な人工知能の行動計画では、空間の位置を表現するのに記号が使われてきた。しかし、位置を指示する言語表現は漠然としており、それによって指示される位置もある程度の広がりや許される。このような位置の漠然性を古典的な行動計画の手法で扱うために、我々のシステムでは記号表現と位置のもっともらしさを表すポテンシャル関数を組み合わせたオブジェクトを使っている。これによって、ある程度の空間的表現に対応できることを明らかにしている。

### 5.4 構成的な動作の辞書の構成

アニメーションを生成するためには、エージェントの動作を定義した辞書が必要となる。しかし、すべての動作を定義することは不可能なので、基本的な動作を定義して、その他の動作は基本動作から構成的に作り出すような機構が必要である。基

本動作をどのように定義するか、あるいはそもそも基本動作なるものが定義できるのかについては哲学の分野でも長い論争がある。基本動作を一般的に定義することは困難なので、我々は工学的な立場に立ち、エージェントとの対話により室内の物体を移動するタスクに限定して動作の辞書の構築を目指している。具体的には、まず、このタスクでおこなわれる対話に含まれる動詞を洗い出し、既存の動詞辞書から得られるこれらの動詞の統語・意味属性を参考にして基本動作を選択するアプローチを採用している [14]。現在は人手によって基本動作の動きを記述しているが、将来的にはモーションキャプチャーなどによって基本動作の収集をおこなえるシステムを開発する予定である。

### 文 献

- [1] S. M. Shieber: "An Introduction to Unification-based Approach to Grammar", CSLI, Stanford University (1986).
- [2] J. Weizenbaum: "Eliza—a computer program for the study of natural language communication between man and machine", *Communications of the ACM*, **9**, 1, pp. 36–45 (1966).
- [3] T. Winograd: "Understanding Natural Language", Academic Press (1972).
- [4] S. C. Levinson: "Space in Language and Cognition", Cambridge University Press (2003).
- [5] W. Levelt: "Speaking: From Intention to Articulation", The MIT Press (1989).
- [6] G. Retzsch-Schmidt: "Various views on spatial prepositions", *AI Magazine*, **9**, 2, pp. 95–105 (1988).
- [7] A. Herskovits: "Language and Spatial Cognition. An Interdisciplinary Study of the Prepositions in English", Cambridge University Press (1986).
- [8] J. Cassell, J. Sullivan, S. Prevost and E. Churchill Eds.: "Embodied Conversational Agents", The MIT Press (2000).
- [9] K. Nagao and A. Takeuchi: "Speech dialogue with facial displays: Multimodal human-computer conversation", *Proc. of 32nd Annual Meeting of the Association for Computational Linguistics (ACL-94)*, pp. 102–109 (1994).
- [10] K. Funakoshi, T. Tokunaga and H. Tanaka: "Processing Japanese self-correction in speech dialog systems", *Proceedings of the 19th International Conference on Computational Linguistics (COLING)*, pp. 287–293 (2002).
- [11] 徳永, 関谷, 田中: "対話システムにおけるプランベースの照応解析", *情報処理学会第 65 回全国大会講演論文予稿集* (2003).
- [12] T. Tokunaga, T. Koyama, S. Saitô and M. Okumura: "Bridging the gap between language and action", *Intelligent Virtual Agent - 4th International Workshop IVA 2003*, Vol. 2792 of LNAI, Springer, pp. 127–135 (2003).
- [13] T. Tokunaga, T. Koyama and S. Saitô: "Classification of Japanese spatial nouns", *Proceedings of 4th International Conference on Language Resources and Evaluation (LREC 2004)*, pp. 1829–1932 (2004).
- [14] T. Tokunaga, M. Okumura, S. Saitô and H. Tanaka: "Constructing a lexicon of action", *the 3rd International Conference on Language Resources and Evaluation (LREC)*, pp. 172–175 (2002).
- [15] C. Nakatani and J. Hirschberg: "A speech-first model for repair identification and correction.", *Proceedings of 31st Annual Meeting of ACL*, pp. 200–207 (1993).
- [16] B. J. Grosz, A. K. Joshi and P. Weinstein: "Centering: A framework for modeling the local coherence of discourse", *Computational Linguistics*, **21**, 2, pp. 203–226 (1995).
- [17] P. R. Cohen: "The pragmatics of referring and the modality of communication", *Computational Linguistics*, **10**, 2, pp. 97–146 (1984).