

自然言語理解アニメーションシステム: K_3

船越孝太郎[†] 徳永健伸[†] 田中穂積[‡]

[†] 東京工業大学 大学院情報理工学研究科 計算工学専攻

[‡] 中京大学 情報科学部 認知科学科

1 はじめに

Winograd による SHRDLU [3] は自然言語理解システムの中で最も重要な位置を占めていると考えられる。積み木を操作するためのアームしか持たない SHRDLU は「身体化」されているとは言いがたいものの、会話エージェントが持つべきいくつかの特徴を確かに有しており、自然言語理解研究の大きな可能性を秘めていた。

近年、自動音声認識技術や自然言語処理技術が目覚ましく発展した。またコンピュータグラフィックスの飛躍的な進歩によって、複雑かつリアルな3次元アニメーションによるヒューマノイドエージェント（ソフトウェアロボット）を仮想空間内に実現できるようになった。これらの技術を組み合わせることにより、SHRDLU を超える言語理解システムを作ることができるであろう。それは、単に技術的に優れるだけでなく、SHRDLU では埋もれていたインタラクションにおける様々な問題を明らかにすることを期待できる。特に、従来の言語理解・会話エージェントの研究ではほとんど議論されてこなかった、行動と言語の関わり合いにおいて生じる多くの興味深い事項を浮き彫りにするであろう。

本稿では、我々の構築したアニメーションエージェントシステム K_3 (図1) を紹介する。実際に K_3 を通じて、言語理解における行動知識の利用、知覚的群化を考慮した参照表現理解、誤解の解消にともなう再解釈のための適切な記憶機構など多くの研究課題が見いだされている。

2 システム概観

K_3 のシステム構成を図2に示す。現在のシステムは、「青い玉をとって」のような、物理的な行為

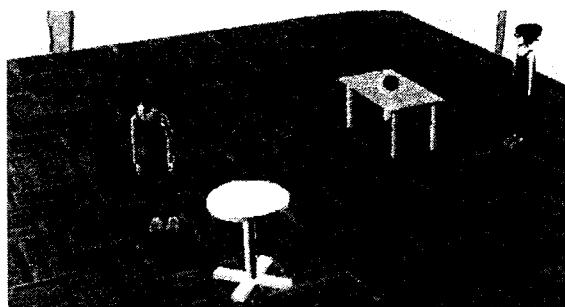


図1: K_3 システムのスクリーンショット

を指示する単文（日本語）を受け付ける。発話には省略・照応が含まれてよい。また、「そっちじゃない」といった、エージェントの誤解を修復する発話も受け付ける。

音声認識モジュールはユーザの音声入力から単語列を生成する。構文/意味解析モジュールは単語列を解析して格フレームを生成する。このモジュールは、助詞落ち、倒置、言い直しを含む発話を解析することができる [1]。この段階では、発話中の省略のために、全てのフレームスロットが埋められているとは限らない。また、省略がない場合でも指示対象物のインスタンスは特定されていない。

省略と照応の解決、そしてインスタンス特定は、談話解析モジュールによって行われる。照応解決とインスタンス特定はプラン知識を用いて統一的に行われる [2]。インスタンス特定に失敗した場合には、エージェントはユーザに問い返し、ユーザの回答を元に再解釈を試みる。

談話解析モジュールはユーザのゴールを抽出し、プランニングモジュールに渡す。プランニングモジュールは与えられたゴールから、適切なアニメーションを生成するためのプランを生成する。言い換えれば、プランニングモジュールはユーザのゴールをアニメーションデータに変換する。しかし、

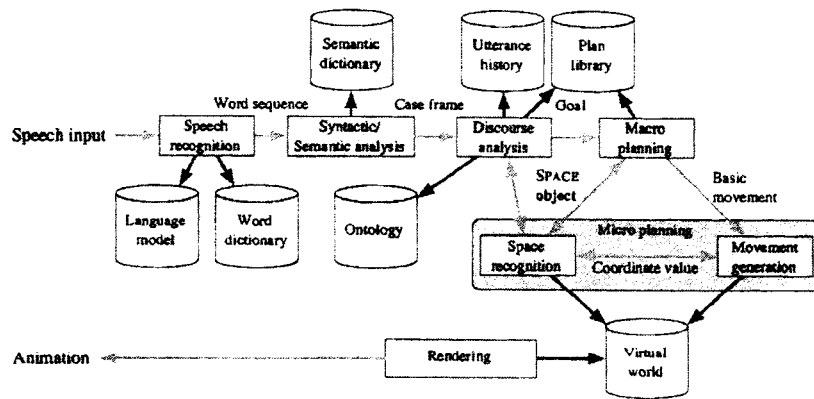


図 2: システム構成

ユーザのゴールとアニメーションデータの性質は大きく異なり、直接的な変換は容易ではない。ユーザのゴールがシンボルとして表現されるのに対し、アニメーションデータは一連の数値群として表現される。この隔たりを克服するために、我々は2段階でプランニングを行うアプローチを取る。この2段階のプランニングをマクロプランニングとマイクロプランニングとよぶ。マクロプランニングを行うマクロプランナはユーザのゴールから、シンボルレベルでのプランを生成する。このプランは基本動作 (basic movement) の配列である。基本動作とはエージェントが行うことのできる最小粒度の動作 (「見る」、「取る」など) である。マイクロプランニングを行うマイクロプランナは基本動作から、数値表現によるアニメーションデータ (エージェントの仮想空間内における位置と方向、および関節角) を生成する。

マクロプランニングの段階においても、プランナは物体の位置など、物理的な属性を知る必要がある。例えば、ボールを拾い上げるためにはエージェントは初めにそのボールに手の届く場所に移動しなければならない。そのためにはボールとエージェントの距離を計算する必要がある。しかし、この種の情報は仮想空間内の座標値として表現されており、マイクロプランナが対象とするものである。シンボル操作を行うマクロプランナには座標情報は直接扱えない。そこで我々は、SPACE オブジェクトとよぶ空間のハイブリッド表現を提案し [2]、これを用いて2つのプランナのインターフェイスを定義した。SPACE オブジェクトは、空間のシンボリックな性質 (例えば、ある対象の「周囲」をあら

わらす空間であること) と数値的な性質 (仮想空間内の特定の空間をあらわすこと) を同時に表現し、マクロプランナの要請に応じてマイクロプランナによって生成される。SPACE オブジェクトはポテンシャル関数を用いて空間を表現するため、空間を参照する言語表現のもつ漠然性 (vagueness) に対応できる。また論理演算との親和性が高く、マクロプランナが生成した空間表現 (例えば、「ボールの近くかつ机の前である空間」) を数値演算に直接変換できるという特性を持つ。マクロプランナは複数のSPACE オブジェクトを用いて空間の論理的な表現を合成し、基本動作のパラメータとしてマイクロプランナに渡す。

謝辞

本研究は文部科学省科学研究費補助金学術創成研究「言語理解と行動制御」(13NP0301)の一部である。プロジェクトの URL は <http://www.cl.cs.titech.ac.jp/sinpro/en/>.

参考文献

- [1] K. Funakoshi, T. Tokunaga, and H. Tanaka. Processing Japanese self-correction in speech dialog systems. In *Proceedings of the 19th International Conference on Computational Linguistics (COLING)*, pages 287–293, 2002.
- [2] T. Tokunaga, K. Funakoshi, and H. Tanaka. K2: Animated agents that understand speech commands and perform actions. In *Proceedings of the 8th Pacific Rim International Conference on Artificial Intelligence*, 2004.
- [3] T. Winograd. *Understanding Natural Language*. Academic Press, 1972.