

エージェントとの対話における空間表現の漠然性の扱い

徳永健伸[†] 小山智史[†] 齋藤豪[‡] 奥村学[‡]
東京工業大学 東京工業大学
大学院情報理工学研究科[†] 精密工学研究所[‡]

概要

曖昧性の解消は、自然言語処理における中心的な関心事であったのに対し、漠然性を扱う研究は、ほとんどおこなわれてこなかった。本稿では、仮想空間中のエージェントとの対話において生じる言語表現の漠然性、特に空間表現の漠然性を扱うことを目的とした空間の表現形式を提案する。エージェントとの対話システムは、言語処理という記号的な処理と同時に、言語理解の結果をアニメーションとして生成するためのアナログ的な処理を含んでいる。本稿で提案する表現形式は記号的な側面と連続数値的な側面の2つの側面を合わせ持ち、この性質により、言語処理とアニメーション生成の2つの処理のギャップを埋めることが可能となる。

1 はじめに

言語行為論が主張するような「言語の使用は行為である」という観点からの研究はこれまで盛んにおこなわれてきたが、それに比べると言語理解の結果どのような行為をおこなうかという観点からの研究は少ない。もちろん、対話システムでは、発語行為の結果、産出される言語表現を理解し、その結果として発話行為をおこなうという意味では、理解の結果として発語行為をおこなっていることになる。しかし、身体的な行為までその射程に入っていることは少ない。近年、知能における「身体性」の役割が重要視されるようになっており [12]、これに呼応して、身体を持ち人間と自然言語を通して対話できるエージェントの研究が最近盛んにおこなわれている [1, 2, 3]。このようなエージェントと対話するためには、従来、自然言語処理の分野で盛んに研究されてきた曖昧性の問題とともに漠然性を扱うことが必要となる。

以下の人間 (H) とエージェント (A) の会話を例にとり、曖昧性と漠然性の違いを見てみよう。

H: 机の前にボールがあるよね。

A: はい。

H: それを机の上に置いて。

人間の2番目の発話中の「それ」は先行する発話中の「机」か「ボール」を参照している。このような状況で「それ」が指す正しい先行詞を決定することは、照応の曖昧性の解消と呼ばれ多くの研究成果が報告されている [5]。実際にエージェントがボールを机の上に置く動作をする場合、ボールを置く位置を正確に決定する必要があるが、発話中では「机の上」と指定しているだけで、正確な位置については述べられていない。曖昧性に比べ、この場合は、決定すべきボールを置く位置に無限の可能性があることになる。このように、身体行為をおこなうエージェントとの対話において、漠然性を扱うことは不可欠である。特に本稿では、この例が示すような空間的位置の漠然性を中心に扱う。

この例が示すように、曖昧性の解消は有限個の可能性からの選択であり、記号操作と親和性がある。一方、漠然性の解消は連続的な無限の可能性からのもっともらしい点を選択することを要求し、これは記号操作とは合い入れない性質を持つ。漠然性のこのような性質が、本質的に記号操作である言語処理の研究で敬遠されてきた理由のひとつであると考えられる。

これまでの多くの対話システムでは、対象世界のオブジェクトとその関係を記号的に表現した離散的空間を扱ってきた。離散的な空間は記号的プランニングと親和性があり、プランニングによって知的な振舞いができるシステムを構築することができた。しかし、同じ枠組で連続的な空間を扱おうとすると破綻が生じる。上述の例で、ボールを置く位置を決定するのに、記号的プランニングを用いると原理的には無限の位置を記号表現する必要が生じる。

以上のような考察から身体的行為をおこなうエージェントとの対話では、記号的な側面と連続的な側面を合わせ持つ空間の表現が必要であることがわかる。このような表現を通して記号システム (言語処理) と連続システム (身体行為) のギャップを埋めることが可能となる。言語による空間表現を可視化する研究が過去にもいくつかおこなわれている。たとえば、Olivier らは定性的・定量的両側面を持った空間表現を提案している [8]。しかし、彼らの主な目的は空間表現の可視化であり、我々のように行為のプランニングという動的な環境での使用は考慮されていない。本稿では、身

体行為をおこなう知的なエージェントとの対話において、記号的なプランニングと連続的な身体行為のインタフェースをとるための空間表現を提案する。

2 システムの概要

提案手法を評価するために、我々はプロトタイプシステム K_2 を実装している。図1は K_2 のスクリーンショットである。 K_2 の仮想空間中には2人のエージェントといくつかのオブジェクトがおかれており、音声入力によってエージェントに指令をして仮想空間中のオブジェクトを操作できる。現在のシステムは、「机のところまで行け」、「もっと」などの断片的な発話にも対応している。また、照応や省略を含む発話も扱うことができる。エージェントが話者の指令を理解した結果は、3次元のアニメーションとして提示される。



図1: K_2 のスクリーンショット

図2に K_2 のアーキテクチャを示す。話者の音声入力は音声認識によって単語列に変換され、さらに構文・意味解析を経て格フレームに変換される。発話中に省略や照応表現が含まれる場合は、この段階では格スロットは空のまま次の談話解析に送られる。談話解析では省略・照応の補完とともに仮想空間中の参照オブジェクトの同定をおこなう。

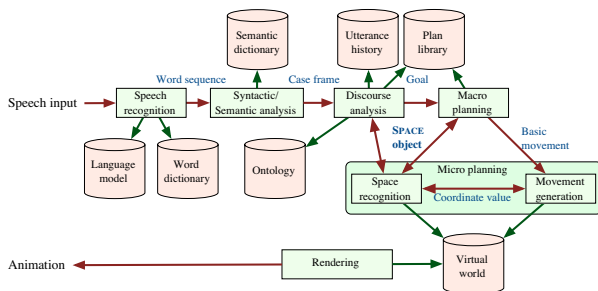


図2: K_2 のアーキテクチャ

談話解析では話者のゴールの抽出もおこない、これ

をプランナに渡しアニメーションを生成するためのプランを生成する。つまり、プランナは話者のゴールをアニメーションに変換すると考えることができるが、両者は異質なデータであり、直接的な変換は困難である。すなわち、話者のゴールは記号で表現されており、アニメーションのデータは数値の系列である。このギャップを解消するために、我々は図2に示すように、マクロ・プランニングとマイクロ・プランニングによる2段階のプランニングをおこなう。

マクロ・プランナは STRIPS [4] に始まる古典的なプランニング方式を採用している。この場合、基本プランオペレータとして基本動作を用いたプランを生成する。たとえば、「on(ボール#1, 机#2)」というゴールは「ボールの近くに行く」、「ボールをつかむ」、「ボールを机に置く」のような基本動作の系列を実行することにより達成できる。基本動作の集合は対象領域に依存するので、これを定義することは一般には困難である。ひとつのアプローチとして [10] を参照されたい。

マクロ・プランナはプランニングの途中でオブジェクトの位置や大きさなどの物理的特性を知る必要がある。たとえば、上述の例で「ボールの近くに行く」という基本動作において、具体的な位置を指定するためにはボールの座標を知る必要がある。この種の情報を記号的なマクロ・プランニングで扱うのは困難なので、すべてマイクロ・プランナで管理している。

マイクロ・プランナの結果は仮想空間のデータベースに反映され、これを基にアニメーションが生成される。

3 SPACE オブジェクト

本稿では、仮想空間中の位置を表現するのに SPACE オブジェクトと呼ぶ表現形式を提案する。SPACE オブジェクトは、記号的側面と数値的側面を合わせ持ち、マクロ・プランナとマイクロ・プランナの2つのプランナのインタフェースの役割を果たす。SPACE はその記号的側面からマクロ・プランナで用いる論理表現の引数として使うことができ、その数値的側面から同時に仮想空間中の座標を表わすことができる。特に SPACE の表わす位置は点座標を表わすのではなく、位置の漠然性を表現するために領域を表わす。正確な座標は状況によって変化するので、SPACE の持つ位置情報は、各点座標をその位置のもっともらしさへ写像するポテンシャル関数によって表現されている [11]。

我々は、SPACE が内包するポテンシャル関数 f を以下の2つの要請に沿って設計した。

- C1. すべての定義域で微分可能であること
- C2. 値域として $[0, 1]$ を取ること

最終的にアニメーションを生成するために点座標を決める場合は、SPACE のポテンシャル関数を用いて最急

降下法によって求める。微分可能性は最急降下法を用いることによる要請である。

ポテンシャル関数の値 1 はその点が一番もってもらしい、0 はその反対であることを表わす。[0, 1] に値域を限定することにより、論理演算子との対応関係を付けやすくなるというのが 2 番目の要請のねらいである。たとえば、2 つの SPACE の AND 操作は、それぞれのポテンシャル関数の値の積として計算できるし、NOT は $(1-f)$ で計算できる。この要請により、マクロ・プランナの世界で記号的に表現された SPACE 間の関係をマイクロ・プランナの世界ではポテンシャル関数の間の演算に変換することができる。

現在、以下のような位置関係を表わすポテンシャル関数を定義している：

- 「前」、「後」、「右」、「左」、「上」、「間」
- あるオブジェクトによって占められる場所
- あるオブジェクトの近傍

ポテンシャル関数のパラメタはオブジェクトの形状や大きさによって決まる。

本来、上記のような位置関係を表現するために考案した SPACE をより複雑な空間的制約を表現するのに用いることもできる。たとえば、「手が届く位置 (reachableByHand(Agent)) は、エージェントの腕のストレスなどをパラメタとして持つポテンシャル関数として定義できる。

4 プランニングの例

本節では、1 の例を用いて SPACE オブジェクトが言語処理とアニメーション生成の仲介をする様子を説明する。

図 1 の状況で話者が「机の前にボールがあるよね」という発話をしたとしよう。談話処理は以下のように「ボール」が指すオブジェクトを同定する。

- (1) `space#1 := new inFrontOf(desk#1, viewpoint#1, MIRROR)`
- (2) `list#1 := space#1.findObjects()`
- (3) `ball#1 := list#1.getFirstMatch(kindOf(BALL))`

ステップ (1) では、`inFrontOf` クラスのインスタンスを生成する。`inFrontOf` の構成子は、参照物、視点、参照軸タイプを引数としてとる [6]。参照枠に関する研究は数多く報告されているが [6, 9, 7]、ここでは Herskovits's の枠組に従う。実際に発話を正確に理解するには参照枠の決定が必要となるが、現在のシステムでは簡単なヒューリスティクスを用いて参照枠を決定している。

`inFrontOf` クラスのポテンシャル関数 f は式 (1) のように定義する。ただし、各パラメタは机の形状から図 3

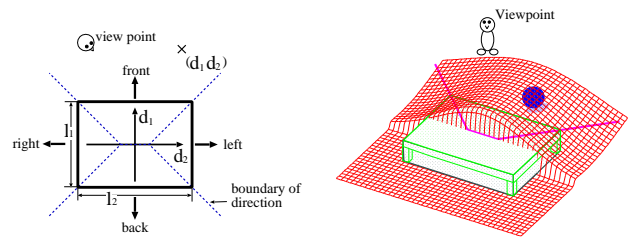


図 3: f のパラメタ

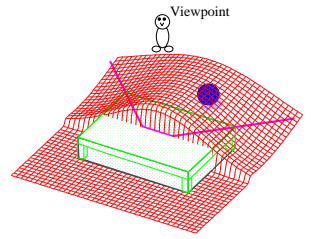


図 4: f のポテンシャル場

のように決定する。最初のガウス関数は“front”方向軸に対する両側の広がり、2 番目のシグモイド関数は“back”方向のポテンシャルの減衰を表わす。この関数は 3 で述べたポテンシャル関数に対する要請を満たしている。この f によって表現されるポテンシャル場の様子を図 4 に示す

$$f(d_1, d_2) = \exp \frac{-a_2^2}{b^2(a_1^2 - \frac{l_2}{2} + \frac{l_1}{2})^2} \times \frac{1}{1 + \exp^{-ad_1}} \quad (1)$$

- d_1 : “front-back” 軸の値
- d_2 : “left-right” 軸の値
- l_1, l_2 : 各軸方向のオブジェクトの最大長
- a, b : 係数

ステップ (2) では、メソッド `matchObjects()` を呼び出すことにより `space#1` の位置にあるオブジェクト候補のリストを得る。オブジェクトの候補はポテンシャルの値によってソートされる。

さらに、ステップ (3) では、オブジェクト候補のリストからオブジェクトの型の制約 (BALL) を満たすものを取り出す。

次の発話「それを机の上に置いて」をエージェントが受けとると、談話解析で「それ」が最初の発話で解決された `ball#1` であることを解釈し、さらに話者のゴールを抽出しこれをマクロ・プランナに渡す。マクロ・プランナはこのゴールを満たすために以下のようにプランを生成する。

- (1) `walk(inFrontOf(ball#1, viewpoint#1, MIRROR) AND reachableByHand(ball#1) AND NOT(occupied(ball#1)))`
- (2) `grasp(ball#1)`
- (3) `put(ball#1, on(desk#1, viewpoint#1, MIRROR))`¹

`walk`, `grasp`, `put` などは基本動作として定義されており、これらはマイクロ・プランナにひとつずつ渡される。

`walk` 動作は目的地として SPACE オブジェクトを引数にとる。この例では、3 つの SPACE の連言が引数として渡されている。つまり、歩いて行く目的地は「ボー

¹実際には机の上にボールを置くスペースがあるかなどの条件も必要となる。

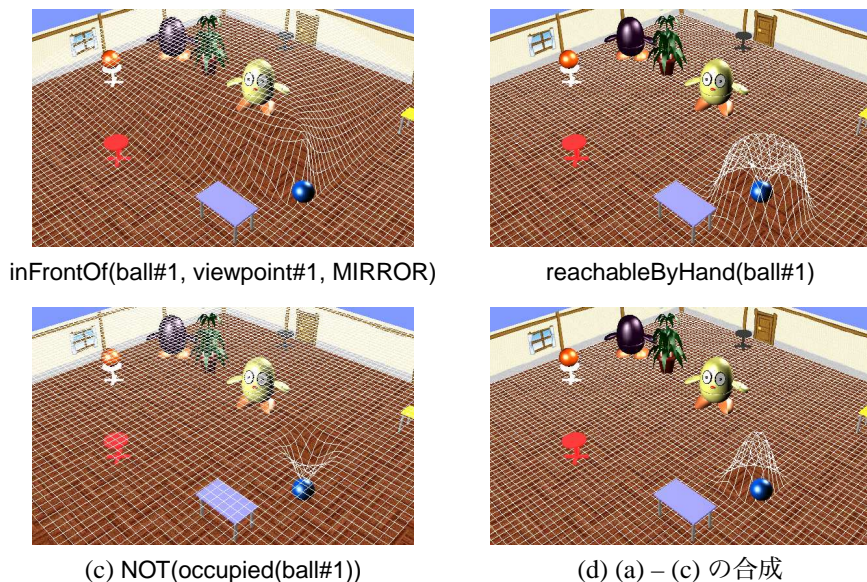


図 5: ポテンシャル関数の合成

ルの手前であり、かつ、ボールに手が届く位置であり、かつ、ボールが占めない位置」である。3で述べたように、マイクロ・プランナの観点からはこの位置は3つのポテンシャル関数の値の積で計算できる。図5は3つのポテンシャル関数((a), (b), (c))と、それから計算された目的地を表わすポテンシャル関数(d)の様子を示している。この結果、エージェントは(d)のポテンシャル場においてもっとも値の大きい点座標を目指して歩くことになる。

その後、基本動作 **grasp** によってボールをつかむが、歩いて行った位置からは手が届くこと (**reachableByHand(ball#1)**) が論理的に保証されているので、この動作は成功するはずである。

ボールを机に置く際にもプランナは **walk** 動作の時と同様な方法でボールが置ける位置を計算する。このように **SPACE** を使うことによって、マクロ・プランナでは記号表現による論理的な関係を扱い、マイクロ・プランナではポテンシャル関数を用いて実際の点座標を計算することが可能となる。

5 おわりに

本稿では、エージェントとの対話における空間表現の漠然性を扱うために、記号的な側面と数値的な側面を合わせ持つ位置の表現形式を提案した。また、これをプロトタイプ・システムに実装しその有効性を確認した。研究課題としては空間表現の語彙の増強、参照枠の決定アルゴリズムの洗練などがあげられる。

参考文献

- [1] N. I. Badler, M. S. Palmer, and R. Bindiganavale. Animation control for realtime visual humans. *Communication of the ACM*, 42(8):65–73, 1999.
- [2] R. Bindiganavale, W. Schuler, J. Allbeck, N. Badler, A. Joshi, and M. Palmer. Dynamically altering agent behaviors using natural language instructions. In *Autonomous Agents 2000*, pages 293–300, 2000.
- [3] J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, editors. *Embodied Conversational Agents*. The MIT Press, 2000.
- [4] R. E. Fikes. STRIPS: A new approach to the application of theorem problem solving. *Artificial Intelligence*, 2:189–208, 1971.
- [5] B. J. Grosz, A. K. Joshi, and P. Weinstein. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):203–226, 1995.
- [6] A. Herskovits. *Language and Spatial Cognition. An Interdisciplinary Study of the Prepositions in English*. Cambridge University Press, 1986.
- [7] W.J.M. Levelt. *Speaking: From Intention to Articulation*. The MIT Press, 1989.
- [8] P. Olivier, T. Maeda, and J. Tsujii. Automatic depiction of spatial descriptions. In *AAAI 94*, pages 1405–1410, 1994.
- [9] G. Retsz-Schmidt. Various views on spatial prepositions. *AI Magazine*, 9(2):95–105, 1988.
- [10] T. Tokunaga, M. Okumura, S. Saitô, and H. Tanaka. Constructing a lexicon of action. In *the 3rd International Conference on Language Resources and Evaluation (LREC)*, pages 172–175, 2002.
- [11] A. Yamada, T. Nishida, and S. Doshita. Figuring out most plausible interpretation from spatial description. In *the 12th International Conference on Computational Linguistics (COLING)*, pages 764–769, 1988.
- [12] 田中穂積 and 徳永健伸. ロボットとの会話-人工知能からのアプローチ-. *情報処理*, 44(12):1247–1252, 2003.