

## 指示詞と普通名詞が混在している中での意味獲得

Acquisition of the meaning of coexisting demonstrative pronouns and common nouns

荒木 修<sup>\*1</sup>      西垣 貴央<sup>\*1</sup>      植村 竜也<sup>\*1</sup>      中谷 仁<sup>\*1</sup>      尾関 基行<sup>\*1</sup>  
 ARAKI Osamu    NISHIGAKI Takahiro    UEMURA Tatsuya    NAKATANI Hitoshi    OZEKI Motoyuki

小島 隆次<sup>\*2</sup>      深田 智<sup>\*1</sup>      岡 夏樹<sup>\*1</sup>  
 KOJIMA Takatsugu    FUKADA Chie      OKA Natsuki

<sup>\*1</sup>京都工芸繊維大学  
 Kyoto Institute of Technology

<sup>\*2</sup>滋賀医科大学  
 Shiga University of Medical Science

Acquisition of the meaning of demonstrative pronouns such as *kore* (this) and *sore* (that) is difficult because the meaning of demonstratives depends on context, such as distance from the speaker and from the hearer, pointing gesture, and ownership. The paper proposes a model that simultaneously acquires the meaning of demonstratives, *kore* and *sore*, and common nouns. An experiment on meaning acquisition is conducted with a dining-table simulator, and the result demonstrates that the meaning of words of the both type can be successfully acquired.

## 1. はじめに

「こそあど」で始まる指示詞は、現場の事物や話題の中の事物を指し示す働きを持ち、日常生活において様々な場面で我々是用いている。

一方、指示詞の意味獲得について考えると、指示詞は発話単語のみで指示対象を特定できず、他の情報を考慮に入れて指示対象を決定するため、複雑である。指示詞の現場指示用法では、話し手と聞き手の関係において、話し手に近い物を「これ」、聞き手に近い物を「それ」、話し手と聞き手の両方に遠い物を「あれ」と表現する。さらに、指示対象の同定に指さしや視線を伴う場合が存在する。指示詞の意味獲得には、このような多様な情報を考慮する必要がある。

指示詞獲得の研究について、K. Goldらはロボットに発話とそのときの状態から決定木を用いた学習を行わせ、「I” “you” “he” “this” “that” “above” “below”の意味獲得を行っている [Gold 09]。意味獲得を行う単語については、特定の事物を指し示す普通名詞はあらかじめその意味が既知であるという前提で、人称代名詞と指示詞の獲得を行っている。

本研究では、計算機上のシミュレーションにより、シミュレーション上に表現されたロボットに、言葉の意味獲得を行わせる。また、意味獲得の言葉として指示詞の他に普通名詞も対象とし、同時並行で学習させる。指示詞と普通名詞を同時並行で学習させることは、特定の事物に対する名前付けと特定の事物に限定しない話し手と聞き手の関係から事物を参照するという異なる2種類の学習を同時に行わなければならない、簡単でない。

発話された各単語の意味を獲得する手法について、発話された各単語と単語が指し示す意味の候補との共起を数え上げ、得られた共起度数分布から各単語に意味づけを行う手法を用いる。またシミュレーションの場面として、食卓上で人とロボットが食べ物のやり取りを行う状況を想定する。

本稿の構成は次の通りである。まず2章で共起情報を用いた意味獲得について述べ、3章で実験で使用するシミュレーション

環境を説明する。4章で意味獲得の実験の説明と結果の分析を行い、5章でまとめと今後の展望を述べる。

## 2. 共起を用いた意味獲得

本稿で言う共起とは、発話された単語と現状内のある要素が同時に出現することを表す。例えば、単語「ちょうだい」「どうぞ」は知っている前提において、「りんごちょうだい」の発話に対して、<りんご>と<みかん>が眼前にある状況ならば、発話された単語「りんご」は物体である<りんご>と<みかん>に対して、それぞれ共起回数を1回ずつ増やす<sup>\*1</sup>。別の場面において、「りんごどうぞ」の発話に対して、<りんご>と<ぶどう>がある状態ならば、発話単語「りんご」は物体である<りんご>と<ぶどう>に対してそれぞれ共起回数を1回ずつ増やす。単語とそれに対応する事物とは共起することが多いと考えられるため、共起回数を数えることにより、単語の意味(指し示す事物)を推定することができる。

しかし、自由発話からの学習では発話単語の頻度に偏りがあるため、少数の発話しか行われぬ単語もある。こうした単語は共起回数も少ないため、出現頻度が多い単語が無関係の意味候補と偶然共起した回数との間で差がつかない。そこで、岡らは自由発話からの学習における、誤りを含み頻度に偏りがある多種類の少数データから、Fisherの直接法を用いることで共起頻度が有意に高い言葉と状況の対を精度よく見つける方法を提案した [Oka 08]。本稿でも、共起情報からの意味推定にFisherの直接法を用いる。

## 3. シミュレーション環境

## 3.1 概要

シミュレーションは食卓場面を想定し、話し手と聞き手がいる状態で、「りんごどうぞ」のように、話し手が聞き手に対して「[word] どうぞ」と発話(ただし吹き出しの文字により、発

\*1 本来は、単語「りんご」と共起しているのは、物体<りんご>、物体<みかん>だけでなく、果物が載っている<食卓>とも共起しているし、果物が食卓に<載っている>という関係や各果物の色や形等にも共起している。このように、共起を数える対象候補は極めて多種多様に存在するが、本稿では、簡単のため、食卓上の個々の物体との共起だけを考える。

連絡先: 荒木 修

京都工芸繊維大学 大学院工芸科学研究科 情報工学専攻  
 e-mail: osamu -at- ii.is.kit.ac.jp

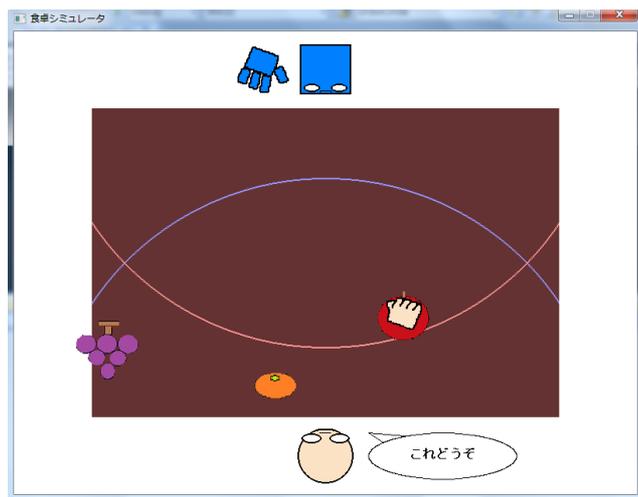


図 1: シミュレーションの画面

話を表現している) し、聞き手にその単語 [word] に対応する対象物体を食べさせるというやり取りを通して、[word] の意味を聞き手であるエージェント (以下、学習エージェントと呼ぶ) に獲得させる\*2。単語 [word] には食卓上に出現する物体名または指示詞の「これ」か「それ」が入る。話し手は、あらかじめ決められた動作を組み込んだエージェント (以下、話し手エージェントと呼ぶ) である。聞き手である学習エージェントは、共起に基づく学習を進めながら、学習中の知識を用いて話し手の発話に応じた動作を行う。

エピソード 1 回の流れは、以下のとおりである。

1. 食卓上で、話し手が聞き手もしくは両者の手が届く位置に、物体が配置される。
2. 今回の実行エピソードでやり取りを行う対象物体が食卓上の配置物体から選ばれる。
3. 話し手は、発話を行う。聞き手から対象物体が届かない場合は、話し手は届く位置に対象物体を持ち運ぶ。発話と持ち運び動作のタイミングは、持ち運び動作の始点と終点の間で発話を行うとしている。
4. 聞き手は、発話内容と食卓の状態から語意学習を行う。
5. 聞き手は、話し手の発話内容に対応すると判断した物体を選択し、取って食べる動作を行う。
6. 聞き手には、行った動作が正しい動作か間違った動作かのフィードバックが与えられる。

食卓上に出現する物体の種類や物体の食卓上での配置、およびやり取りを行う対象物体を変えてエピソードを繰り返すことにより、学習エージェントに普通名詞や指示詞の意味獲得を行わせる。実行中のシミュレーションの画面を図 1 に示す。話し手と聞き手のそれぞれについて図 1 では円弧で示されている手の届く範囲を決めてあり、手の届く範囲内でのみ、食卓上の物体を持ったり動かしたりできるようにしている。

\*2 場面として話し手と聞き手の他に、傍観者がいる場合も考えられる。このような設定での実験も実施したが、紙数の関係で、本論文では、話し手と聞き手のみの場面での実験について報告する。

### 3.2 話し手エージェント

話し手エージェントは、あらかじめ組み込んだ動作を以下の通り行う。

- 対象物体が、話し手からは届くが聞き手からは届かない領域にある場合

対象物体をつかんで聞き手の届く領域に持っていきながら発話を行う。発話単語に対象物体名と「これ」のどちらか一方を等確率で使用する。

- 対象物体が、話し手からは届かず聞き手からは届く領域にある場合

発話単語に対象物体名を使用する。

- 対象物体が、話し手に届かない範囲にあり聞き手に届く範囲にある場合

話し手の届かない範囲に他の物体がなければ、発話単語に対象物体名と「それ」のどちらかを等確率で使用する。話し手の届かない範囲に他の物体がある場合、対象物体名を使用する。

### 3.3 学習エージェント

聞き手である学習エージェントは、発話単語に対して共起する情報を数え上げ、いままでのやり取りで数え上げた共起度数分布から、対象物体を判断する。

学習エージェントは、話し手の発話に対し以下の通りに動作する。

1. 共起の数え上げ
2. 意味の判定と動作
3. 動作に対するフィードバックを利用した学習

#### 3.3.1 共起の数え上げ

話し手の発話に含まれる単語に対して、現在の状況との共起の数え上げを行う。単語が指し示す意味として用いる候補は、以下のとおりである。

食卓上の物体

例えば、物体として<りんご>と<みかん>が出現する可能性のある場合で、<りんご>だけ食卓に置かれているとき、発話中の単語との共起回数は、

- <りんご> は、+1
- <みかん> は、0

となる。

物体の話し手からの距離

話し手からの距離を話し手から手が届くと位置手が届かない位置の 2 つに分割し、それぞれの位置に置いてある物体を、<話し手：近>と<話し手：遠>で表現する。例えば、りんごが食卓上の<話し手：近>の範囲に置かれているとき、発話中の単語との共起回数は、

- <話し手：近> は、+1
- <話し手：遠> は、0

となる。またこのとき、りんごを運ぶ動作が伴う場合の共起回数は、

- <話し手：近>は，+2
- <話し手：遠>は，0

と始点と終点の両方で共起を数える．

#### 物体の聞き手からの距離

聞き手からの距離を聞き手から手が届く位置と聞き手から手が届かない位置の2つに分割し，それぞれの位置に置いてある物体を，<聞き手：近>と<聞き手：遠>で表現する．例えば，みかんが食卓上の<聞き手：遠>の範囲に置かれているとき，発話中の単語との共起回数は，

- <聞き手：近>は，0
- <聞き手：遠>は，+1

となる．話し手が発話と同時に運ぶ動作を行う場合は，

- <聞き手：近>は，+1
- <聞き手：遠>は，+1

と始点と終点の両方の位置で共起を数える．

#### 話し手による物体の把持

話し手の発話時に，話し手が手に持っている物体を<話し手把持>で表現する．例えば，りんごとみかんが食卓上に置かれているとき，発話時に，話し手が手に何も持っていない場合の発話中の単語との共起回数は，

- <話し手把持>は，0

となり，発話時に，話し手が手にりんごを持っている場合は，

- <話し手把持>は，+1

となる．

#### 3.3.2 意味の判定と動作

発話に対する意味の判定から動作までは，以下の手順で行う．

1. 共起度数を記録した度数分布表から，ある単語が指し示す意味の候補とその単語に着目した四分表を作成する．
2. それぞれの四分表に対して，Fisherの直接法を用いて，作成した四分表以上に偏った四分表が得られる確率の総和を求め．その中からあらかじめ設定した閾値以下を意味の候補とする．閾値について，岡らはインタラクションデータからの意味学習において閾値の設定は0.2程度が望ましいと提言しており [Oka 08]，本設定でも閾値を0.2としている．閾値0.2以下の意味候補が存在しなかった場合，発話の意味が理解できないとして動作を行わない．
3. 推測された意味に当てはまる物体を食卓上から選択する．もし，複数の意味候補が得られていた場合，求めた総和確率の低い順に食卓上の物体と当てはまるか確かめる．
4. 推測した意味に当てはまる物体を選択できた場合，その物体を取って食べる．もし選択できなかった場合，動作を行わない．

#### 3.3.3 動作に対するフィードバックを利用した学習

意味を推定し動作を行った後，学習エージェントに動作が正しかったかどうか与えられ，その結果により以下の方法で共起度数を更新する<sup>\*3</sup>．

##### 正しい動作だった場合

発話単語と選択した物体との共起 (物体の種類・話し手からの距離・聞き手からの距離) について，もう一回数え上げる．話し手が物体を持って発話していたなら，<話し手把持>の共起度数をもう一回数え上げる．

##### 間違った動作だった場合

発話単語と選択した物体との共起 (物体の種類・話し手からの距離・聞き手からの距離) について，数え上げた共起度数を取り消す．

## 4. 実験

### 4.1 実験設定

前章で説明したシミュレーション環境を用いて，学習エージェントによる普通名詞と指示詞のそれぞれの獲得にどのような傾向があるかの実験を行う．実験の共通設定は，以下のとおりである．

- 食卓上での物体の同時出現数は1~3個の内から等確率で決定する
- 同じ種類の物体は最高1個までしか出現しない
- 食卓上に出現しうる物体は，りんご，みかん，ぶどうの3種類で，同時出現数が1か2の場合は，等確率で出現物体を決定する
- 食卓上での各物体の配置は，<話し手：近> <聞き手：遠>，<話し手：近> <聞き手：近>，<話し手：遠> <聞き手：近>の3領域のいずれかに等確率で配置する

条件設定として，学習エージェントが注目能力を持っている場合と持っていない場合を考え，実験を行う．注目能力を持っている場合，発話と並行して物体を渡してきた場合，渡してきた物体に対してのみ共起を数え上げ，学習結果とは関係なく話し手が渡してきた物体を学習エージェントが選択する．

### 4.2 実験結果と分析

シミュレーション中の単語「りんご」「これ」「それ」について各意味候補に対する共起頻度パターンの生起確率の推移をそれぞれ図2，図3，図4に示す．各グラフでは，計算した確率の総和が閾値0.2以下 (学習エージェントが発話に対する意味として動作する値) になった意味候補を表示している．

図2の単語「りんご」では，<りんご>の共起頻度パターンの生起確率が一番小さくなり，正しい意味候補が獲得できている．しかし，<話し手：遠>と<聞き手：近>についても共起頻度パターンの生起確率が0.2以下となっている部分がある．50エピソード実行後，学習エージェントが注意能力なしの場合の各名詞の間違って得られている意味候補については，単語「りんご」「みかん」「ぶどう」ともに<聞き手：近>となっている．今回の実験での物体の出現位置の確率を考えると<聞き手：近>と<聞き手：遠>では，<聞き手：近>は<聞き手：遠>の倍となり，<聞き手：近>と名詞が共起しやすい設定と

\*3 今回は，共起度数による学習のみを扱うため，フィードバックも共起回数として表現する．

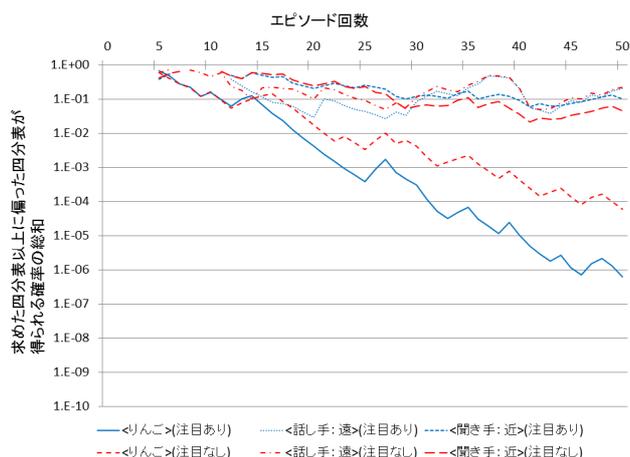


図 2: 単語「りんご」についての各意味候補に対する共起頻度パターンの生起確率の推移

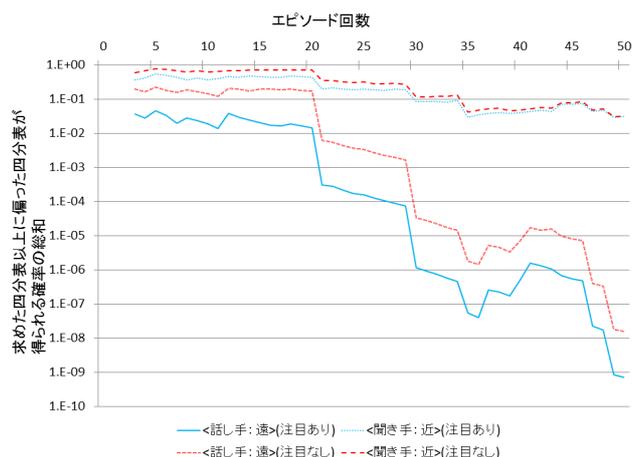


図 4: 単語「それ」についての各意味候補に対する共起頻度パターンの生起確率の推移

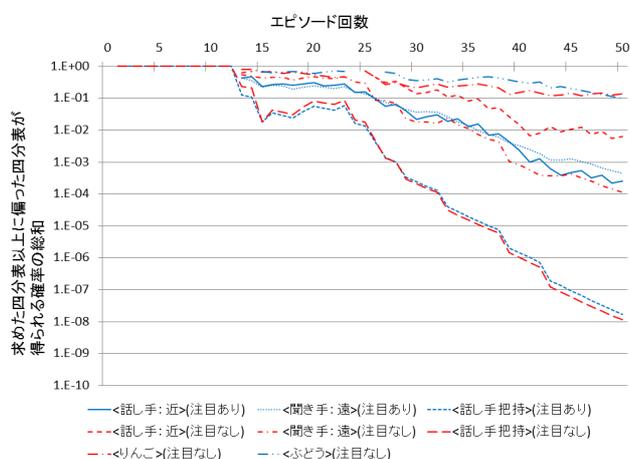


図 3: 単語「これ」についての各意味候補に対する共起頻度パターンの生起確率の推移

なっているため、名詞について<聞き手:近>の意味候補が得られている。

図 3 の単語「これ」では、共起頻度パターンの生起確率は注目能力がある場合は小さい順に<話し手把持>、<話し手:近>、<聞き手:遠>、注目能力がない場合は小さい順に<話し手把持>、<聞き手:遠>、<話し手:近>となっている。<話し手把持>の共起頻度パターンの生起確率が一番小さい原因は、単語「これ」の発話に<話し手把持>が必ず伴うためである。また、単語「これ」では物体を運ぶ動作があるため、共起カウントに必ず「<話し手:近>が+2、<聞き手:近>と<聞き手:遠>がそれぞれ+1」が含まれ、共起頻度パターンの生起確率は小さい順に<話し手把持>、<話し手:近>となり、<聞き手:遠>は共起頻度パターンの生起確率が低い状態では現れないはずである。原因として、エピソード数 200 回後の注意能力がない状態での共起頻度パターンの生起確率を見ると、小さい順に<話し手把持>、<話し手:近>、<聞き手:遠>と<話し手:近>と<聞き手:遠>が逆転しているため、確率的な変動の範囲内の差だと考えられる。

図 4 の単語「それ」では、共起頻度パターンの生起確率は小さい順に<話し手:遠>、<聞き手:近>となっている。<

話し手:遠>の共起頻度パターンの生起確率は小さいのは、単語「それ」が発話される時、<話し手:遠>に必ず一つのみ物体が存在しているためである。

注目能力がある場合とない場合の比較では、注目能力がある場合の方が正しい意味候補は共起頻度パターンの生起確率がより小さくなり、正しくない意味候補は共起頻度パターンの生起確率が大きくなる傾向があり、意味獲得の精度がよくなっている。注目能力がある場合は、物体を渡しながらの発話では正しい物体のみを参照するため共起のカウントにノイズが入らないためである。

以上をまとめると、意味として適切でない意味候補も共起頻度パターンの生起確率が 0.2 以下となり意味として扱われた。この点にさらに判定法を加え、適切でない意味候補を省く必要があると考えられる。

## 5. 結論

単語の意味獲得において、獲得方法として発話された単語と物体との共起を数え上げる方法を用い、指示詞と普通名詞の同時意味獲得の計算機シミュレーションを行った。その結果、意味候補を特定の数に絞った実験設定においては、指示詞と普通名詞の意味が獲得できることが分かった。

今後の展望として、計算機上のシミュレーションのみで行うのではなく現実世界でロボットを用いて、今回のシミュレーションと同じことが可能かどうかを試したい。方法として、話し手に人間を用いる場合とあらかじめ正しい動作を組み込んだロボットを用いる場合が考えられ、やり取りを通じて聞き手のロボットに言葉の意味獲得をさせることを考えている。

## 参考文献

- [Gold 09] K. Gold, M. Doniec, C. Crick, and B. Scassellati, "Robotic vocabulary building using extension inference and implicit contrast", *Artificial Intelligence*, 173(1): 145-165(2009)
- [Oka 08] 岡夏樹, 増子雄哉, 林口円, 伊丹英樹, 川上茂雄, "Fisher の直接法を用いたインタラクションデータからの意味学習", *知能と情報 (日本知能情報ファジィ学会誌)*, 20(4): 461-472(2008).