

# ヒューマン-ロボット・インタラクションを通じた終助詞「よ」「ね」の意味獲得

## Meaning acquisition of sentence-final particles *yo* and *ne* through human-robot interaction

大上 涼麻  
Ogami Ryoma

呉 霞  
Wu Xia

深田 智  
Fukada Chie

尾関 基行  
Ozeki Motoyuki

岡 夏樹  
Oka Natsuki

京都工芸繊維大学 大学院工芸科学研究科  
Graduate School of Science and Technology, Kyoto Institute of Technology

Sentence-final particles serve an important role in (spoken) Japanese, because they express the speaker's mental attitudes toward the proposition and/or the interlocutor. They are acquired at early ages and occur very frequently in everyday conversation. There has been, however, little proposal for the computational model of the acquisition of sentence-final particles. The purpose of this study is to get a robot to learn how to act upon the utterance with a sentence-final particle. The robot learns appropriate responses based on the rewards given by the interlocutor. The experimental results show that the robot learns to behave correctly in response to *yo*, which expresses the speaker's intention to communicate new information, and to *ne*, which denotes the speaker's desire to confirm that some information is shared. Using the learned actions as a lead, the acquisition of inner information processing such as word learning is the next research target.

### 1. はじめに

終助詞は日常会話で頻繁に用いられ、話者の心的態度を表すという重要な働きを持つ。また、子どもによる獲得時期も早いことが知られている。しかし、その意味獲得の計算モデルはあまり研究されていない。ロボットに言語を獲得させようとする研究は近年盛んであるが、終助詞の獲得は、著者の知る範囲では試みられていない。

本研究では以下の2つの終助詞の用法を学習対象としてとらげる:

- 「よ」の教示用法(聞き手が知らないと思われる情報を聞き手に告げ知らせる用法)[金水 93]
- 「ね」の同意要求用法(話し手・聞き手ともに共有している目される情報について、聞き手に同意を求める用法)[金水 93]

本研究の目的は、終助詞によって表現された話者の心的態度に応じた適切な反応を、ロボットに学習させることである。ロボットは、人とのインタラクションを通して適切な反応を強化学習により獲得していく。ロボットの反応は、次の3種類を含む:

1. 体の動き(例:発話対象物の方を見る、うなづく)
2. 発話
3. 内部情報処理(例:教えてもらった新情報を記憶する)

このうち、1と2は外部から観測できるが、3は外部から直接は観測できないため、獲得が難しいと考えられる。

本論文では、2節で体の動きの獲得についての我々のこれまでの研究成果を報告し、3節において、体の動きと内部情報処理を同時に獲得させる試み(構想)を紹介する。発話の獲得は本論文では扱わない。

### 2. 発話に対する適切な行動の獲得

我々は、終助詞「よ」の教示用法と終助詞「ね」の同意要求用法に注目し、それらを含む発話に対してロボットが適切に行動できるようになることを目指した[大上 12]。

具体的には、図1に示すように、ロボットが前方に置かれたリングかミカンを見ている状態で、ロボットの正面に座った実験参

加者が「りんごだよ」「みかんだね」などと話しかける。実験参加者は、発話対象物がどれであるかを、手で発話対象物に触れる(か、または、指さす)ことにより示す。ロボットは Kinect で手の位置を検出することにより発話対象物がどれであるかを知る。話しかけられたロボットは、発話者の顔を見る/発話対象物を見る/それ以外の物を見る/うなづく、の4種類の行動をいくつか(最大で3つ)組み合わせて動作する。

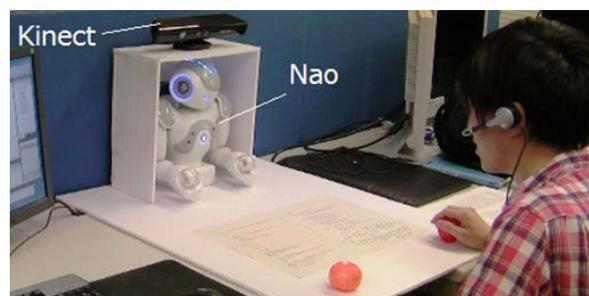


図1: 行動獲得の評価実験の様子。実験参加者が「りんごだよ」などと話しかけ、ロボットはそれに対する適切な行動を獲得する。

ロボットの行動に対して、実験参加者はマウス操作で報酬を与える。ロボットは強化学習(Q学習)により、報酬が多く得られる行動をとるようになっていく。

次の3種類の設定で評価実験を行った。

実験1(名前の教示):ロボットは物体の名前を知らないという設定で実験を行う。ロボットが見ている方の物体に手を添え「××だよ」と名前を教える。

実験2(同意の要求):ロボットは物体の名前を知っているという設定で実験を行う。ロボットが見ている方の物体に手を添え「××だね」と同意を要求する。

実験3(存在の教示):ロボットは物体の名前を知っているという設定で実験を行う。ロボットが見ていない方の物体に手を添え「××だよ」とそこに××があることを知らせる。

図2にはそれぞれの実験で想定している場面における発話例を示した。

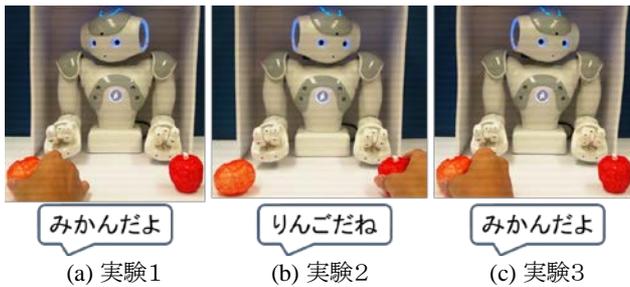


図2: 各実験で想定している場面における発話例。(a)では、ミカンを見ているロボットに対して、「みかんだよ」と名前を教えている。(b)では、リンゴを見ているロボットに対し、「りんごだね」と同意を要求している。(c)では、リンゴを見ているロボットに対し、「みかんだよ」とミカンがあることを教えている。

8名の実験参加者が各実験において10回ずつ発話をし、ロボットの反応に対して報酬を与えた。こうして得られた各実験80個ずつのデータに基づいてQ学習を行った。表1に学習結果の一部を示す。表1では、各実験において行動価値が最も高くなった行動パターン5個を行動価値の高い順に、また、行動価値が最も低くなった行動パターン5個を行動価値の低い順に、それぞれ示した。これ以外の結果については[大上12]を参照されたい。

表1: 実験ごとの行動価値の学習結果(一部)。「顔」は発話者の顔を見る行動を、「対象物」は発話対象物を見る行動を、「その他」は発話対象物以外の物体を見る行動を、それぞれ示す。行動パターンは最大3つの行動(1番目、2番目、3番目)からなる。著者らが想定した評価とは逆の評価が得られた行動パターンには色をつけて示した。

実験1(名前の教示)					
高評価の行動パターン			低評価の行動パターン		
1番目	2番目	3番目	1番目	2番目	3番目
うなづく	対象物	対象物	その他	顔	顔
対象物	うなづく	対象物	顔	顔	顔
うなづく	その他	対象物	対象物	対象物	その他
その他	うなづく	対象物	うなづく	対象物	その他
うなづく	その他	その他	その他	顔	-

実験2(同意の要求)					
高評価の行動パターン			低評価の行動パターン		
1番目	2番目	3番目	1番目	2番目	3番目
うなづく	顔	-	対象物	その他	その他
うなづく	その他	対象物	対象物	顔	顔
うなづく	うなづく	うなづく	その他	対象物	その他
顔	顔	うなづく	その他	対象物	対象物
顔	うなづく	うなづく	その他	対象物	-

実験3(存在の教示)					
高評価の行動パターン			低評価の行動パターン		
1番目	2番目	3番目	1番目	2番目	3番目
顔	うなづく	対象物	うなづく	うなづく	その他
対象物	うなづく	その他	その他	顔	顔
その他	対象物	うなづく	うなづく	うなづく	うなづく
顔	その他	-	うなづく	うなづく	対象物
顔	その他	うなづく	うなづく	顔	顔

表1から分かるように、いずれの実験条件においてもロボットは概ね適切な行動をとれるようになった。著者らが想定した評価とは逆の評価が得られた行動パターン(表中の色つきのセル)もあったが、それらは特定の実験参加者に偏ってみられる傾向があることが分かった。このことは、本実験のような条件下のロボットの行動に対する評価には個人差があることを示しており、学習機能を持たせてユーザごとに個人適応する価値があると考えられる。

### 3. 内部処理と行動の同時獲得の計算モデル

前節のロボットは、「りんごだよ」と名前を教える発話に対して、リンゴの方を見てうなづくといった適切な行動がとれるようになるが、表面的な行動を獲得しただけで、実際に名前を覚えるという内部処理は獲得できていない。そこで、本節では、行動とともに内部情報処理も獲得させることを目指し、その実現に向けての構想を記す。

内部処理を報酬から獲得することは、次の理由により、行動獲得よりも難しい。内部処理は対話相手から直接は観測できないため、適切な報酬を与えられるとは限らない。たとえば、うなづくいたので報酬を与えたが、後になって実際は名前を覚えていなかったことが判明することも考えられる。

この問題に対応するため、本研究では、(1) [大上12]で用いたモデルを、遅れのある報酬を扱えるように(すなわち、少し前の行動を遡って評価できるように)拡張するとともに、(2) 図3に示すように、学習すべき状態と行動の種類をできるだけ減らして単純な課題設定にする—という方針をとる。(2)の方針を採用した理由は、人とのインタラクションを通して学習するという設定では、多くの学習データが利用できることは期待できないため、複雑な設定の学習は困難であると予想されるからである。

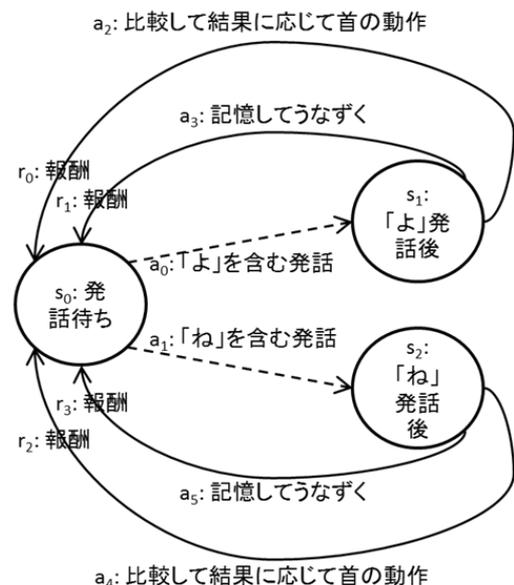


図3: 学習対象とする状態と行動。丸は状態を表す。破線の矢印は人の行動(ここでは発話)を示し、実線の矢印はロボットの行動(内部処理も含む)を示す。

通常の強化学習と同様に、各状態(図3中の丸が状態に対応)における適切な行動(将来の報酬も含めて最大の報酬がもらえる行動)を学習するが、人とロボットが交互に行動(ここでは発話や内部処理も含めて行動と呼ぶ)することを想定しているため、それぞれの行動によって状態遷移が起こる。図3では、人の行

動(発話)を破線で示し、ロボットの行動(内部処理も含む)を実線で示した。ロボットは学習結果である行動価値(ある状態におけるある行動の結果将来期待できる報酬の合計値)に応じて行動するが、人はここで学習される行動価値とは独立に行動を決定するとする。したがって、ここで学習される「人の行動の行動価値」は、人の行動の決定には使われないが、ロボットの行動の行動価値の学習には用いられる。行動価値の学習には、replacing trace を使用した Sarsa( $\lambda$ ) [Sutton 98]を用いる。

なお、ロボットの行動の1つとして図中に記した「記憶してうなずく」は、人の発話に含まれる「りんご」などの単語と目の前のリンゴの画像をペアにして長期記憶に格納し、うなずくことを想定している。また、「比較して結果に応じて首の動作」と図中に記した行動は、マイクでとらえた音声から切り出した単語と、カメラでとらえた画像から切り出した物体の画像を、ロボットの長期記憶中の単語-画像ペアと比較し、一致したらうなずき、不一致なら首を振り、該当する記憶がなければ首は動かさないことを想定している。

人とロボットが交互に行動して状態が遷移していくモデルとしたため、学習アルゴリズムは通常の強化学習と少し異なるものとなった。アルゴリズムの提示だけでは分かりにくいと考えられるため、図4において、学習の進行の様子を例示しながらアルゴリズムを説明する。図4中の  $s_0$ ,  $a_0$  等は、図3中の記号と対応している。

#### 4. おわりに

本論文では、「りんごだよ」「みかんだね」などと話しかけられたロボットが、終助詞「よ/ね」により表現された発話意図(新情報を知らせる/共有情報について同意を求める)に対応する適切な行動(例:発話対象物の方を見る、うなずく)を、報酬に基づいて獲得できることを実験的に示した。つづいて、目に見える行動だけでなく、必要な内部情報処理(例:教えてもらった新情報を記憶する)も同時に獲得することを目指し、そのための計算モデルと学習アルゴリズムを提案した。

今後、この提案したアルゴリズムについて、いくつかの実験設定や学習機能に関する設定を変えながら評価実験を行い、どのような場合に、どの程度の量のインタラクションにより、終助詞の意味と名詞の意味を同時に獲得することができるかを調べる予定である。

#### 参考文献

- [金水 93] 金水敏: <言語学の最新情報>—日本語学:終助詞ヨ・ネ, 月刊言語, Vol. 22, No. 4, pp. 118-121, 1993.
- [大上 12] 大上涼麻, 松岡啓, 荒木修, 柴田諒子, 高岡勇紀, 土坂恭斗, 呉霞, 深田智, 尾関基行, 岡夏樹: 人とロボットのインタラクションを通じた機能語の意味獲得:「終助詞の意味」を「発話に対する適切な行動」であると捉える立場から, HAIシンポジウム 2012, 2D-7, 2012.
- [Sutton 98] Sutton, R. S. and Barto, A. G.: Reinforcement Learning: An Introduction, MIT Press, Cambridge, MA, 1998.

Q 値をすべて0に初期化しておく。すべての  $s, a$  に対し、eligibility trace  $e(s, a) = 0$  と初期化する。 $s_0$  から始めるとする。

1 番目のステップ(例):

- $s_0$  で人が行動  $a_0$  をとり  $s_1$  に遷移したとしよう。報酬は無し(人の行動への報酬は、いつも無し)。
- $s_1$  でロボットがとる行動を選択。Q 値に応じて行動  $a_3$  を選択(ルーレット選択を使用)するとしよう。
- このステップでは報酬がないので

$$\delta \leftarrow 0 + \gamma Q(s_1, a_3) - Q(s_0, a_0)$$

- $e(s_0, a_0) \leftarrow 1$ ; replacing trace を使用する

For all  $s, a$ :

- $Q(s, a) \leftarrow Q(s, a) + \alpha \delta e(s, a)$ ; 人は Q 値とは無関係に行動を決めるが、ロボットの行動の Q 値の更新のためには、人の行動の Q 値が必要なので、人の行動の Q 値も更新しておくことに注意。ただし、Q 値が0に初期化されているので、このステップでは Q 値は更新されない。
- $e(s, a) \leftarrow \gamma \lambda e(s, a)$

2 番目のステップ(例):

- (先ほどのステップで選択した通り)  $s_1$  でロボットが行動  $a_3$  をとり  $s_0$  に遷移する。報酬は  $r_1$ 。
- $s_0$  で人が行動  $a_1$  を選択したとしよう。(人は Q 値とは無関係に行動を決める。なお、人が次の行動を決めないと  $\delta$  が計算できない。)

$$\delta \leftarrow r_1 + \gamma Q(s_0, a_1) - Q(s_1, a_3)$$

- $e(s_1, a_3) \leftarrow 1$ ; replacing trace を使用する

For all  $s, a$ :

- $Q(s, a) \leftarrow Q(s, a) + \alpha \delta e(s, a)$ ;  $Q(s_1, a_3)$  が少し増える。 $Q(s_0, a_0)$  も trace の大きさに応じて少し増えることに注意。
- $e(s, a) \leftarrow \gamma \lambda e(s, a)$

以降のステップも同様に続く

図4: 使用する学習アルゴリズム。ここでは、分かりやすくするため、学習の進行の様子を例示することによりアルゴリズムを説明した。