Concurrent Acquisition of the Meaning of Sentence-Final Particles and Nouns Through Human-Robot Interaction

Natsuki Oka, Xia Wu, Chie Fukada, and Motoyuki Ozeki

Graduate School of Science and Technology, Kyoto Institute of Technology

Abstract. Sentence-final particles serve an important role in spoken Japanese, because they express the speaker's mental attitudes toward a proposition and/or an interlocutor. They are acquired at early ages and occur very frequently in everyday conversation. However, there has been little proposal for a computational model of the acquisition of sentence-final particles. In this paper, we report on a study in which a robot learns how to react to utterances that have a sentence-final particle and gives appropriate responses based on rewards given by an interlocutor, and at the same time, learns the meaning of nouns. Preliminary experimental result shows that the robot learns to react 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, and also learns the correct referents of nouns.

Keywords: language acquisition, function words, reinforcement learning.

1 Introduction

Sentence-final particles serve the important role of expressing the speaker's mental attitudes. They are acquired at early ages and occur very frequently in everyday conversation. Ohshima et al. proposed a robot that uses sentence-final particles in order to draw the hearer's attention [4]. Research on the computational model of language acquisition is rapidly increasing [1]. However, to the best of our knowledge, there has been no proposal for a computational model of the acquisition of sentence-final particles except for our earlier reports [3,5].

In the preceding reports, we dealt with the following two usages of sentence-final particles *yo* and *ne*, (although there are several other usages of *yo* and *ne*):

- The informing usage of *yo*: informing the listener of information that seems new to the listener [2].
- The agreement requesting usage of *ne*: requesting an agreement on information that seems to be shared between the speaker and the listener [2].

The purpose of the study was to get a robot to learn a series of appropriate physical reactions to a speaker's mental attitude expressed with a sentence-final particle. We used a robot, instead of a virtual agent, because the reactions to be learned included

M. Lee et al. (Eds.): ICONIP 2013, Part I, LNCS 8226, pp. 387–394, 2013.

the gaze direction, which is difficult for a virtual agent to express accurately. The robot learned appropriate reactions based on rewards given by its interlocutor.

In general, responses from a robot include the following:

1. physical reactions such as a nod, turning of its face in the direction of the referent of the utterance, etc.
2. utterances
3. inner information processing such as memorizing new information received, etc.

Among the three items listed, 1 and 2 are observable by an interlocutor; however, item 3 cannot be directly observed, which makes it difficult for him/her to give appropriate rewards in accordance with the robot's response, and inappropriate rewards make it difficult for the robot to learn appropriate responses. Our earlier study [3,5] only dealt with item 1. The robot thus only acquired outward behaviors, and did not learn inner information processing such as remembering the name of an object. While, this study deals with item 3 as well as item 1. Although it does not cover item 2, we believe that utterances can be learned as well as physical reactions because both are observable.

The remainder of this paper is organized as follows. We describe our preceding computational model for the acquisition of physical reactions to sentence-final particles in Section 2. We explain our new model for learning the invisible inner processing and demonstrate the leaning capability of the model in Section 3. Finally, we outline future work and conclude this paper in Section 4.

2 Acquisition of Appropriate Physical Reactions

2.1 Computational Model

In this study, the robot learned appropriate physical responses to the two usages of sentence-final particles *yo* and *ne* based on the rewards given by an interlocutor. We formulated the problem as a reinforcement learning (RL) process. State in RL consisted of the utterance of the interlocutor, the referents of the utterance, objects within the eyesight of the robot, and others. For simplicity, we assumed that the rewards are given every time without fail, and excluded delayed rewards, which simplified the action value update as follows:

$$Q(s,a) \leftarrow Q(s,a) + \alpha\big(r - Q(s,a)\big),$$

where $Q(s,a)$ is an action-value function, that is, the value of taking action a in state s, α is the learning rate, and r is the reward.

2.2 Experiments

We conducted three experiments: The robot learned (1) the informing usage of *yo*, specifically, the usage that relates the name of an object (Exp. 1), (2) the agreement requesting usage of *ne* (Exp. 2), and (3) the informing usage of *yo*, specifically, the usage that relates the existence of an object (Exp. 3).

Fig. 1 shows the experimental environment utilized. The general procedure used was as follows: The participant talks to the robot using a sentence-final particle, and he/she indicates the associated object with his/her hand by touching the object or pointing to it. The robot recognizes the word uttered using registered-word voice recognition, and identifies the object being referred to with Kinect by detecting the interlocutor's hand. It then reacts by randomly combining at most three of the following four elemental actions: nodding, turning its face toward the interlocutor's face, turning its face toward the object in question, or turning its face toward a different object. The participant gives a reward of 1 or -1 to the robot using a mouse, and the robot learns which action sequences result in the most reward using Q-learning.

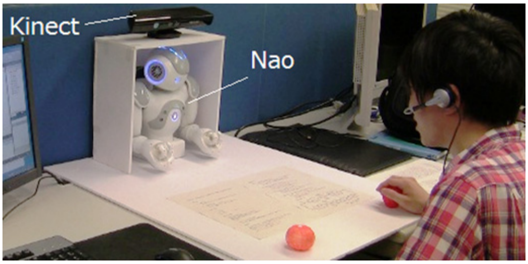

Fig. 1. Experimental environment for the earlier study. The participant talked to the robot using a sentence-final particle, and the robot learned physical reactions to it.

Fig. 2 gives examples of the sentence uttered in each experiment. While the participants were informed that the robot does not know the names of the objects on the table in Experiment 1, they were informed that the robot knows the names in Experiments 2 & 3. Please note that the participants talked about the object at which the robot looked in Experiments 1 & 2, but they talked about the object at which the robot did not look in Experiment 3.

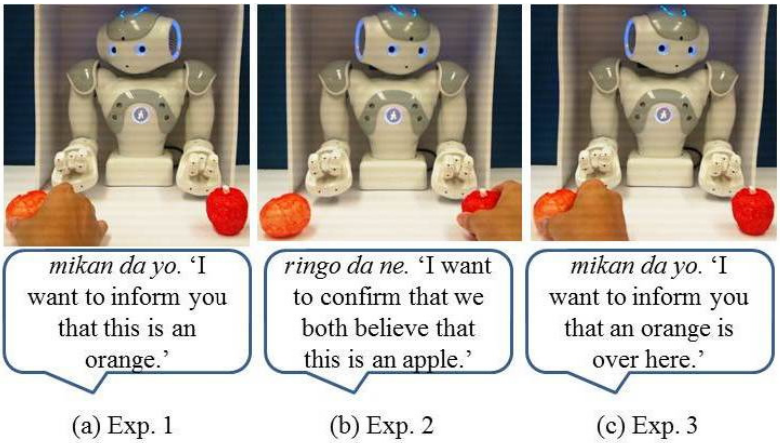

Fig. 2. Examples of the sentence uttered in each experiment

2.3 Results and Discussion

The main learning results are shown in Table 1 (see [3,5] for more details). The results indicate that the robot learned to react more or less correctly in response to sentence-final particles *yo* and *ne*. Several action sequences resulted in the opposite action values to those that were expected by us; these are indicated by colored cells in the table.

We found that there were individual differences in the evaluation of the robot's actions, and the aforementioned reversed evaluations were mostly observed for specific participants. This means that adaptation to an individual user is worthwhile.

Table 1. Main results of action-value learning. For each experiment, action sequences that have the top five action values are shown in descending order. Action sequences consisted of at most three elemental actions. "Face" represents the elemental action of turning the robot's face toward the interlocutor's face; "object" represents the action of turning toward the relevant object; and "other" signifies the action of turning toward another object.

Exp. 1: Instructing names			Exp. 2: Requesting agreement			Exp. 3: Informing existence		
1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd
nod	object	object	nod	face	-	face	nod	object
object	nod	object	nod	other	object	object	nod	other
nod	other	object	nod	nod	nod	other	object	nod
other	nod	object	face	face	nod	face	other	-
nod	other	other	face	nod	nod	face	other	nod

3 Learning Invisible Inner Processing

The robot described in the previous sections learned appropriate reactions such as turning toward an apple and nodding on hearing a sentence containing a sentence-final particle. However, it only acquires outward behaviors, and does not learn inward processing such as remembering the name of an object.

In this section, we explain our idea for learning inward processing as well as outward behaviors. Learning inner processing from rewards is much more difficult than learning visible behaviors. This is because accurate rewards are not always given for invisible inner processing. For example, it is probable that even though a reward was given when the robot nodded, it subsequently turns out that the robot does not actually remember the name.

3.1 Computational Model

In order to resolve the aforementioned issue, we employed the following policies: (1) the robot should learn from delayed rewards; and (2) the state space of learning, i.e., the number of states and actions, should be as small as possible. We employed the latter policy because it is difficult to obtain sufficient data for complicated learning when the data comes only from interaction with humans.

We thus set out a simple state space, shown in Fig. 3, in the first place. An important difference from the standard reinforcement learning (RL) is the alternate actions between a human and a robot. Human actions are represented by dashed arrows and robot's actions are expressed with solid arrows in Fig. 3, and both actions cause state transitions. While the robot acts according to the learned action values as in the standard RL, human actions are decided independently of the action values in the state space.

One of the robot's actions in Fig. 3, "memorize and nod," is the act of memorizing a pair of a word, such as apple, which is a segment of speech, and an image of an object in front of its eyes, and nodding. "Compare and move neck according to the result" is the act of nodding if the currently presented word-image pair is the same as the pair in memory, shaking its head if the current pair disagrees with the stored pair, or no neck motion otherwise.

We use the learning algorithm shown in Fig. 4, which is a modified version of Sarsa(λ) [6]. The modification includes the following: (1) alternate actions between the human and the robot; and (2) use of a replacing trace [6] instead of an eligibility trace.

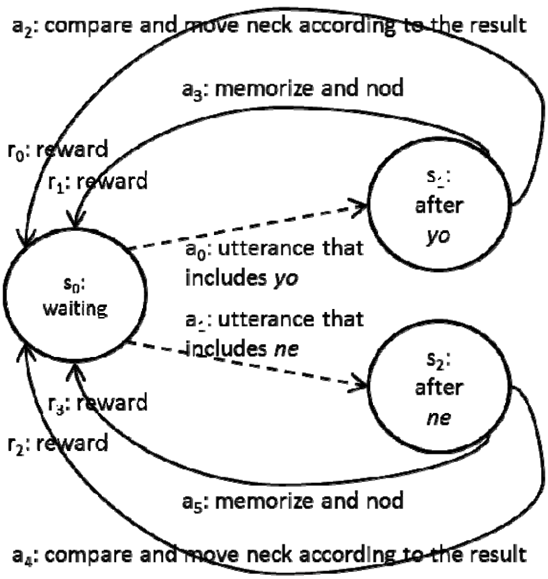

Fig. 3. State transition diagram. Dashed arrows represent human actions, and solid arrows depict robot's actions that include inner processing.

Initialize $Q(s,a) = 0$; $e(s,a) = 0$, for all s, a
Initialize $s = s_0$
Repeat
  Human's turn:
    The human takes action a
    Observe r, s'; $r = 0$ for human actions
    $s \leftarrow s'$
  Robot's turn:
    Choose a from s using policy derived from Q (ε-greedy)
    Take action a
    Observe r, s'
    $\delta \leftarrow r + \gamma Q(s',a') - Q(s,a)$; $Q(s',a') = 0$ for human actions
    $e(s,a) \leftarrow 1$
    For all s, a:
      $Q(s,a) \leftarrow Q(s,a) + \alpha\delta e(s,a)$
      $e(s,a) \leftarrow \gamma\lambda e(s,a)$
    $s \leftarrow s'$
until the end of episode

Fig. 4. Learning algorithm for inward processing: a modified version of Sarsa(λ) [6]

3.2 Preliminary Experiment

A preliminary experiment with five participants was conducted. The setting of the experiment was similar to that described in Section 2.2 except for the number of objects on the table; not two but one. In this experiment, three kinds of objects, an orange, an apple, and a banana, were provided. A participant selected one of the three objects, put it in front of the robot, and talked about it.

At first, the robot did not know the meaning of sentence-final particles *yo* and *ne*, nor the names of the three objects. The participant tried to teach the two sentence-final particles and the three nouns to the robot. Learning rate α was set to 0.1, and probability for random action selection ε was set to 0.1. In order to simplify the analysis of the experimental results as a first step, discount rate γ was set to 0, that is, the robot learned only from immediate rewards, in this preliminary experiment.

Table 2 shows a typical example of the progress of learning. In this case, the robot concurrently and successfully learned the two sentence-final particles, which are shown in the action value columns, and the three nouns, which are indicated in the contingency table column. The action value columns demonstrate that appropriate inner processings have positive values, and an inappropriate processing has a negative value. Another inappropriate processing has not been tried. An incorrect pair 'object: apple'-'name: orange' is stored in the contingency table, but it is not serious because the correct pair 'object: apple'-'name: apple' has more counts.

We, however, found that not every participant succeeded in teaching. There was a participant who gave punishment every time when the robot nodded after hearing sentence-final particle yo, and the robot failed to learn the meaning of yo. In order to deal with this kind of unusual but consistent rewards, we plan to revise the fixed relation between inward processing and outward behavior to flexible one.

Table 2. Typical example of the progress of learning in the preliminary experiment. “A”,“O“, and “B” respectively stand for apple, orange, and banana. “M” and “C” represent the inward processing of memorizing and comparison, respectively. The first data row indicates that the participant put an apple in front of the robot and said “*ringo da yo*” (which means, “I want to inform you that this is an apple”), then the robot memorized the name of the object and nodded, and then the participant gave a reward and the robot updated the action value. The names of the objects in the robot’s memory are represented as a 3×3 contingency table.

<table>
<tr><th rowspan="3">object</th><th rowspan="3">noun used</th><th rowspan="3">particle used</th><th rowspan="3">inner process</th><th rowspan="3">action</th><th rowspan="3">reward</th><th colspan="4">action value</th><th colspan="4" rowspan="3">contingency table</th></tr>
<tr><th colspan="2">yo</th><th colspan="2">ne</th></tr>
<tr><th>M</th><th>C</th><th>M</th><th>C</th></tr>
<tr><td rowspan="4">A</td><td rowspan="4">A</td><td rowspan="4">yo</td><td rowspan="4">M</td><td rowspan="4">nodding</td><td rowspan="4">+1</td><td rowspan="4">0.10</td><td rowspan="4"></td><td rowspan="4"></td><td rowspan="4"></td><td>obj \ n</td><td>A</td><td>O</td><td>B</td></tr>
<tr><td>A</td><td>1</td><td>0</td><td>0</td></tr>
<tr><td>O</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>B</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td rowspan="4">A</td><td rowspan="4">O</td><td rowspan="4">ne</td><td rowspan="4">M</td><td rowspan="4">nodding</td><td rowspan="4">−1</td><td rowspan="4"></td><td rowspan="4"></td><td rowspan="4">−0.10</td><td rowspan="4"></td><td>obj \ n</td><td>A</td><td>O</td><td>B</td></tr>
<tr><td>A</td><td>1</td><td>0</td><td>0</td></tr>
<tr><td>O</td><td>1</td><td>0</td><td>0</td></tr>
<tr><td>B</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td rowspan="4">A</td><td rowspan="4">A</td><td rowspan="4">yo</td><td rowspan="4">M</td><td rowspan="4">nodding</td><td rowspan="4">+1</td><td rowspan="4">0.19</td><td rowspan="4"></td><td rowspan="4"></td><td rowspan="4"></td><td>obj \ n</td><td>A</td><td>O</td><td>B</td></tr>
<tr><td>A</td><td>2</td><td>0</td><td>0</td></tr>
<tr><td>O</td><td>1</td><td>0</td><td>0</td></tr>
<tr><td>B</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>A</td><td>O</td><td>ne</td><td>C</td><td>shaking</td><td>+1</td><td></td><td></td><td></td><td>0.10</td><td colspan="4"></td></tr>
<tr><td rowspan="4">B</td><td rowspan="4">B</td><td rowspan="4">ne</td><td rowspan="4">M</td><td rowspan="4">nodding</td><td rowspan="4">−1</td><td rowspan="4"></td><td rowspan="4"></td><td rowspan="4">−0.19</td><td rowspan="4"></td><td>obj \ n</td><td>A</td><td>O</td><td>B</td></tr>
<tr><td>A</td><td>2</td><td>0</td><td>0</td></tr>
<tr><td>O</td><td>1</td><td>0</td><td>0</td></tr>
<tr><td>B</td><td>0</td><td>0</td><td>1</td></tr>
</table>

4 Conclusions and Future Work

In this paper, we outlined a computational model for learning appropriate physical reactions to utterances that have a sentence-final particle. Our experimental results indicated that the robot learned to react more or less correctly in response to yo and ne.

We then proposed a learning algorithm for inward information processing as well as outward physical behaviors. The result of the preliminary experiment seems promising, and we plan to conduct thorough experiments to test whether the meaning of both the sentence-final particles and nouns can be learned at the same time. We also plan to investigate the relation between the complexity of the state space and the amount of interaction necessary for learning.

Acknowledgement. This work was supported by JSPS KAKENHI 21500137 and 25330260.

References

1. Alishahi, A.: Computational Modeling of Human Language Acquisition, Morgan & Claypool (2010)
2. Kinsui, S.: Latest information on Linguistics: Japanese language study: Sentence-final particles *yo* and *ne*. Gengo 22(4), 118–121 (1993) (in Japanese)
3. Ogami, R., Matsuoka, H., Araki, O., Shibata, R., Takaoka, Y., Tsuchisaka, Y., Wu, X., Fukada, C., Ozeki, M., Oka, N.: Meaning acquisition of function words through human-robot interaction: Viewing the meaning of sentence-final particles as appropriate actions for utterances. Paper Presented at Human-Agent Interaction Symposium 2012, 2D-7 (2012) (in Japanese)
4. Ohshima, N., Ohyama, Y., Odahara, Y., De Silva, P.R.S., Okada, M.: Talking-ally: Intended persuasiveness by the utilizing hearership and addressivity. In: Proceedings of the Fourth International Conference on Social Robotics, pp. 317–326 (2012)
5. Oka, N., Ogami, R., Wu, X., Fukada, C., Ozeki, M.: Acquiring the meaning of sentence-final particles *yo* and *ne* through human-robot interaction. Paper Presented at the First International Conference on Human-Agent Interaction, Sapporo, Japan (2013)
6. Sutton, R.S., Barto, A.G.: Reinforcement learning: An introduction. MIT Press, Cambridge (1998)