Acquiring the Meaning of Sentence-Final Particles *yo* and *ne* Through Human-Robot Interaction

Natsuki Oka, Ryoma Ogami, 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 no proposal for a computational model of the acquisition of sentence-final particles. In this paper, we report on a study in which a robot learned how to react to utterances that have a sentence-final particle and gave appropriate responses based on rewards given by an interlocutor. Our experimental results show that the robot learned to react more or less 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. The next research target for the learning system is the acquisition of inner information processing such as word learning, using the learned actions as a guide.

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. Research on the computational model of language acquisition is rapidly increasing. However, to the best of our knowledge, there has been no proposal for a computational model of the acquisition of sentence-final particles.

The study reported on in this paper dealt with the following two usages of sentence-final particles *yo* and *ne* at the first onset, (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 [1].
- The agreement requesting usage of *ne*: requesting an agreement on information that seems to be shared between the speaker and the listener [1].

The purpose of the study was to get a robot to learn a series of appropriate responses to a speaker’s mental attitude expressed with a sentence-final particle. We used a robot, instead of a virtual agent, because the responses to be learned included the gaze direction, which is difficult for a virtual agent to express accurately. The robot learned appropriate responses 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. This study deals with items 1 and 3, and does not cover item 2, but 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 computational model for the acquisition of physical reactions to sentence-final particles and demonstrate the learning capability of our proposed model in Section 2. We explain our idea for learning the invisible inner processing 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 our study, the robot learned appropriate responses to the two usages of sentence-final particles *yo* and *ne* based on the rewards given by an interlocutor. The appropriate response is not necessarily a single action, but generally an action sequence that does not have the Markov property. In order to avoid tackling a non-Markovian process, we formulated the problem as a simple reinforcement learning (RL) process in which an action sequence is considered as an action.

State in RL consists of the utterance of the interlocutor, the referents of the utterance, objects within the eyesight of the robot, and others. For simplicity, we can assume that the rewards are given every time without fail, and exclude delayed rewards, which simplify 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).

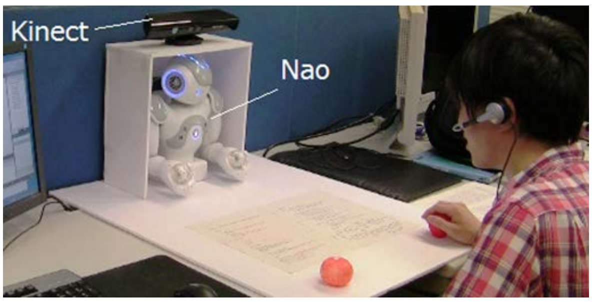
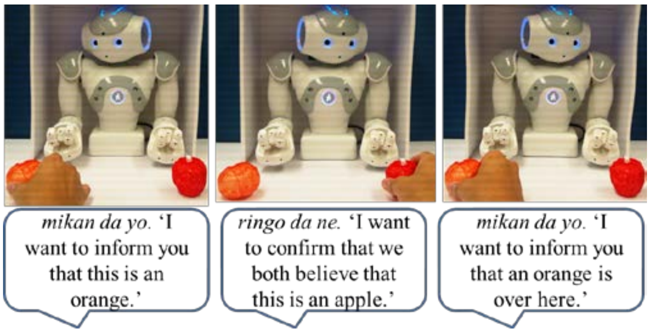


Fig. 1: Experimental environment for the acquisition experiments. The participant talks to the robot using a sentence-final particle, and the robot learns how to react to it.

Fig. 1 shows the experimental environment utilized. The general procedure used is 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 actions result in the most reward using Q-learning. Learning rate α was set to 0.1 in our experiments.



(a) Exp. 1 (b) Exp. 2 (c) Exp.3

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

Fig. 2 gives examples of the sentence uttered in each experiment. The experiments conducted were as follows:

Experiment 1: The participant was informed that the robot does not know the names of the objects on the table. The participant said to the robot "*mikan da yo*" (which means, "I want to inform you that this is an orange") when the robot was looking at the orange, as shown in Fig. 2(a); or "*ringo da yo*" (which means, "I want to inform you that this is an apple") when the robot was looking at the apple.

Experiment 2: The participant was informed that the robot knows the names of the objects on the table. The participant said to the robot "*ringo da ne*" (which means, "I want to confirm that we both believe that this is an apple") when the robot was looking at the apple, as shown in Fig. 2(b); or "*mikan da ne*" (which means, "I want to confirm that we both believe that this is an orange") when the robot was looking at the orange.

Experiment 3: The participant was informed that the robot knows the names of the objects on the table. The participant said to the robot "*mikan da yo*" (which means, "I want to inform you that an orange is over here") when the robot was looking at the apple, as shown in Fig. 2(c); or "*ringo da yo*" (which means, "I want to inform you that an apple is over here") when the robot was looking at the orange.

2.3 Results and Discussion

Each of eight participants (five male and three female students) uttered 10 sentences in each experiment, which generated 80 pieces of learning data in total for each scenario. Offline learning was executed. The main learning results are shown in Table 1 (see [2] for more details). For each experiment, action sequences that have the top five action values are shown in descending order, and the bottom five in ascending order. In ordinary reinforcement learning, action sequences with high action values tend to be produced by the robot, whereas those with low values are seldom produced.

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. 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					
Action sequences with the top five action values			Action sequences with the bottom five action values		
1st	2nd	3rd	1st	2nd	3rd
nod	object	object	other	face	face
object	nod	object	face	face	face
nod	other	object	object	object	other
other	nod	object	nod	object	other
nod	other	other	other	face	-

Exp. 2: Requesting agreement					
Action sequences with the top five action values			Action sequences with the bottom five action values		
1st	2nd	3rd	1st	2nd	3rd
nod	face	-	object	other	other
nod	other	object	object	face	face
nod	nod	nod	other	object	other
face	face	nod	other	object	object
face	nod	nod	other	object	-

Exp. 3: Informing existence					
Action sequences with the top five action values			Action sequences with the bottom five action values		
1st	2nd	3rd	1st	2nd	3rd
face	nod	object	nod	nod	other
object	nod	other	other	face	face
other	object	nod	nod	nod	nod
face	other	-	nod	nod	object
face	other	nod	nod	face	face

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.

In order to resolve the issue, we plan to employ 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 will employ 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. The most appropriate action in each state is learned as in ordinary reinforcement learning (RL). An important difference from the standard 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.

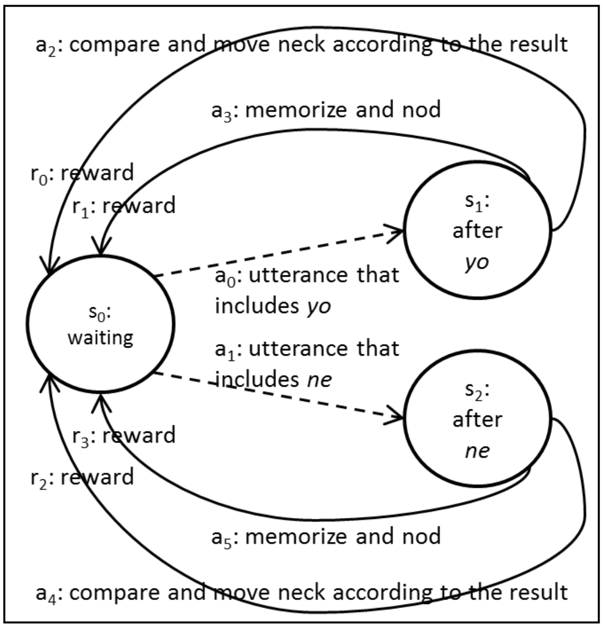


Fig. 3: State transition diagram. Dashed arrows represent human actions, and solid arrows depict robot's actions that include inner processing.

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 if there are no related pairs in memory.

While the robot acts according to the learned action values as in the standard RL, human actions are independent of the action values in the state space. The action values of human actions are not referred to by the human, but are used in the update of those of the robot's actions (See Fig. 4).

We use the learning algorithm shown in Fig. 4, which is a modified version of Sarsa(λ) [3]. The modification includes the following: (1) alternate actions between the human and the robot; and (2) use of a replacing trace [3] instead of an eligibility trace.

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
    Choose a' from s' using policy derived from Q
    $\delta \leftarrow r + \gamma Q(s', a') - Q(s,a)$
    $e(s,a) \leftarrow 1$
    For all s, a:
      $Q(s,a) \leftarrow Q(s,a) + \alpha\delta e(s,a)$
      $e(s,a) \leftarrow \gamma\lambda e(s,a)$
    $s \leftarrow s'$; $a \leftarrow a'$
  Robot's turn:
    Take action a
    Observe r, s'
    The human chooses a' from s'
    $\delta \leftarrow r + \gamma Q(s', a') - Q(s,a)$
    $e(s,a) \leftarrow 1$
    For all s, a:
      $Q(s,a) \leftarrow Q(s,a) + \alpha\delta e(s,a)$
      $e(s,a) \leftarrow \gamma\lambda e(s,a)$
    $s \leftarrow s'$; $a \leftarrow a'$
until the end of episode

Fig. 4: Learning algorithm for inward processing: A modified version of Sarsa(λ) [3].

4 Conclusions and Future Work

In this paper, we proposed a computational model for physical reaction acquisition. Our experimental results indicate that the robot learned to react more or less 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 has been shared.

We then outlined a learning algorithm for inward information processing as well as outward physical behaviors. We plan to conduct 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] Kinsui, S.: Latest information on linguistics: Japanese language study: Sentence-final particles *yo* and *ne*, *Gengo*, Vol. 22, No. 4, pp. 118-121, (1993) (in Japanese)

[2] Ogami, R, Matsuoka, H, Araki, O, Shibata, R., Takaoka, Y., Tsuchisaka, Y., Wu, X., Fukada, C., Ozeki, M, and Oka, N.: Meaning acquisition of function words through human-robot interaction: Viewing the meaning of sentence-final particles as appropriate actions for utterances, Human-Agent Interaction Symposium 2012, 2D-7, (2012) (in Japanese)

[3] Sutton, R. S., and Barto, A. G.: Reinforcement learning: An introduction, MIT Press, Cambridge, MA, (1998)