



Learning Dialogue Strategies for Interview Dialogue Systems That Can Engage in Small Talk Utterances



Tomoaki Nakamura¹, Takahiro Kobori¹, Mikio Nakano²

¹The University of Electro-Communications

²Honda Research Institute Japan Co., Ltd.

Background & Purpose of This Study

- ▶ **Interview dialogue systems**
 - ▶ Workforce can be reduced
 - ▶ Humans are more likely to disclose their information to dialogue systems than human interviewers [Lucas+ 14]
- ▶ **There are few studies on interview dialogue systems**
 - ▶ Mainly focused on collecting user information
 - ▶ The users are unwilling to use them repeatedly



**Develop interview dialogue systems
that users are willing to use repeatedly**

→ can be used for recording user's habit



Target of This Study

- ▶ Interview for diet recording
 - ▶ Japanese text-based chat system
 - ▶ The system asks what the user ate or drank
 - ▶ During the interview, the system sometimes engages in small talk
 - ▶ Example:

Asking question →

What did you eat for dinner?

System

I ate sushi.

User

Small talk →

That's great! Was it tasty?

System

Yes! That's very good!

User

Asking question →

Did you eat a lot?

System



Problems of the Previous Method

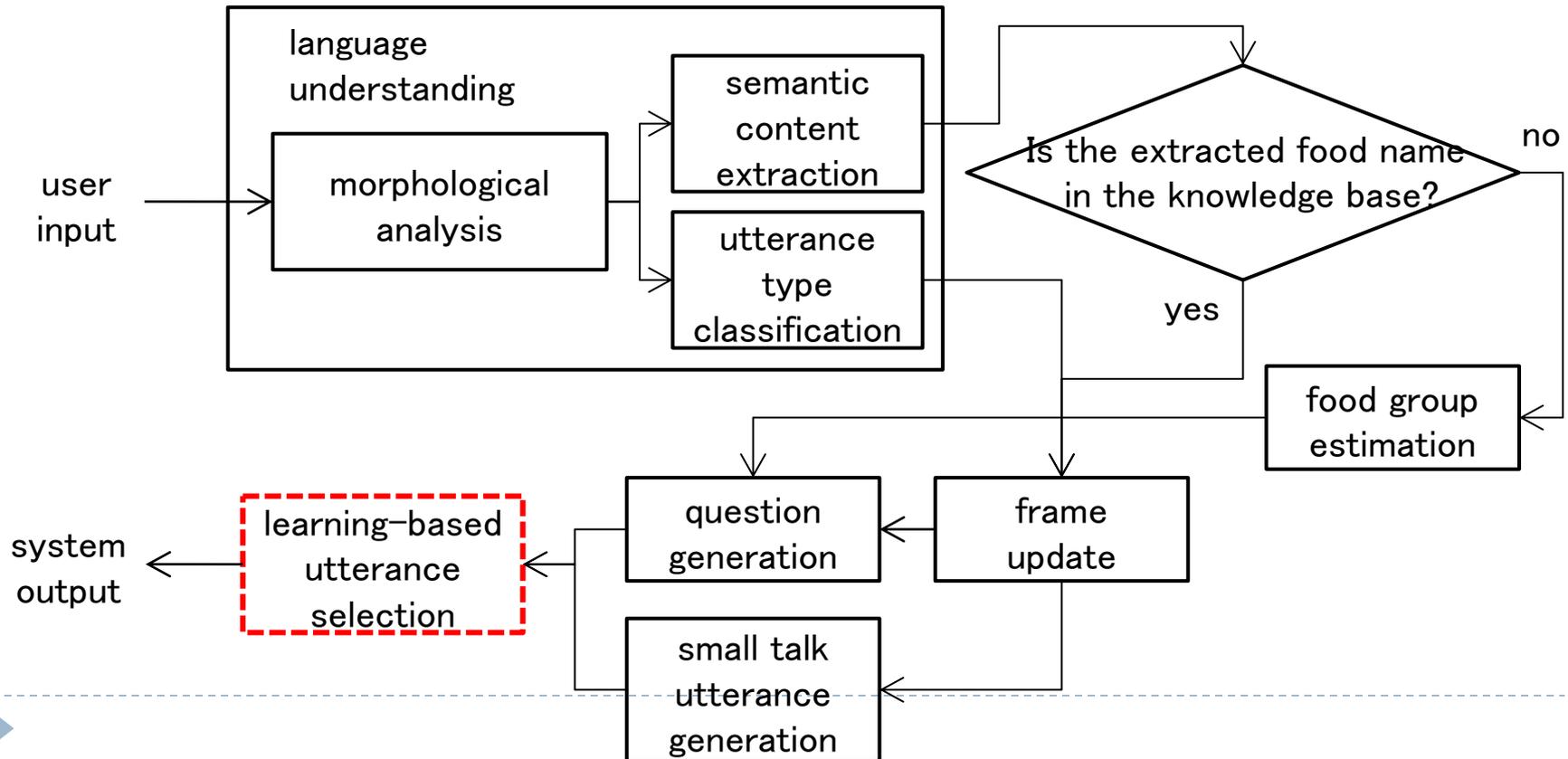
- ▶ We implemented a system by combining task-oriented and non-task-oriented dialogues [Kobori+ 16]
 - ▶ Mainly engage in an interview
 - ▶ Sometimes engage in small talk
 - ▶ Small talk improves users' impression
- ▶ Problems with the previous system
 - ▶ Candidates of small talk utterances are selected based on heuristics and one of them is randomly selected
 - ▶ The number of small talk after the question is fixed
 - ➔ It is difficult to always select appropriate utterances

This study:

- introduces learning based utterance selection
 - evaluates the effectiveness of the proposed method
-

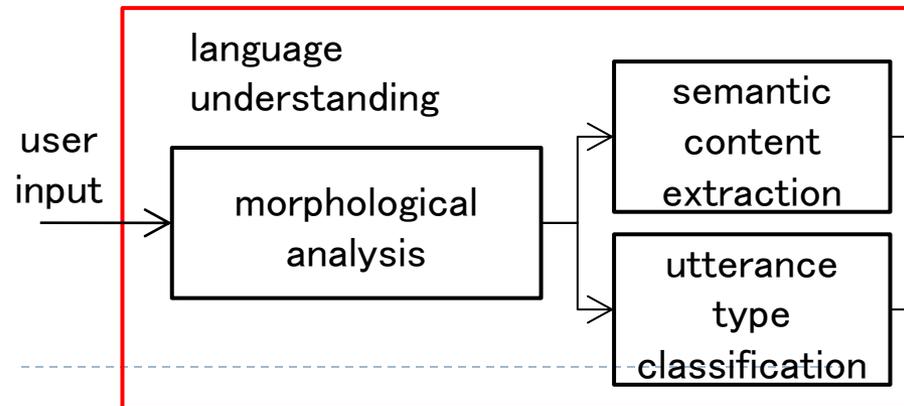
Architecture of the Proposed System

- ▶ This system is frame-based dialogue system
- ▶ Basically same as the previous system [Kobori+ 16]
 - ▶ We introduce learning-based utterance selection



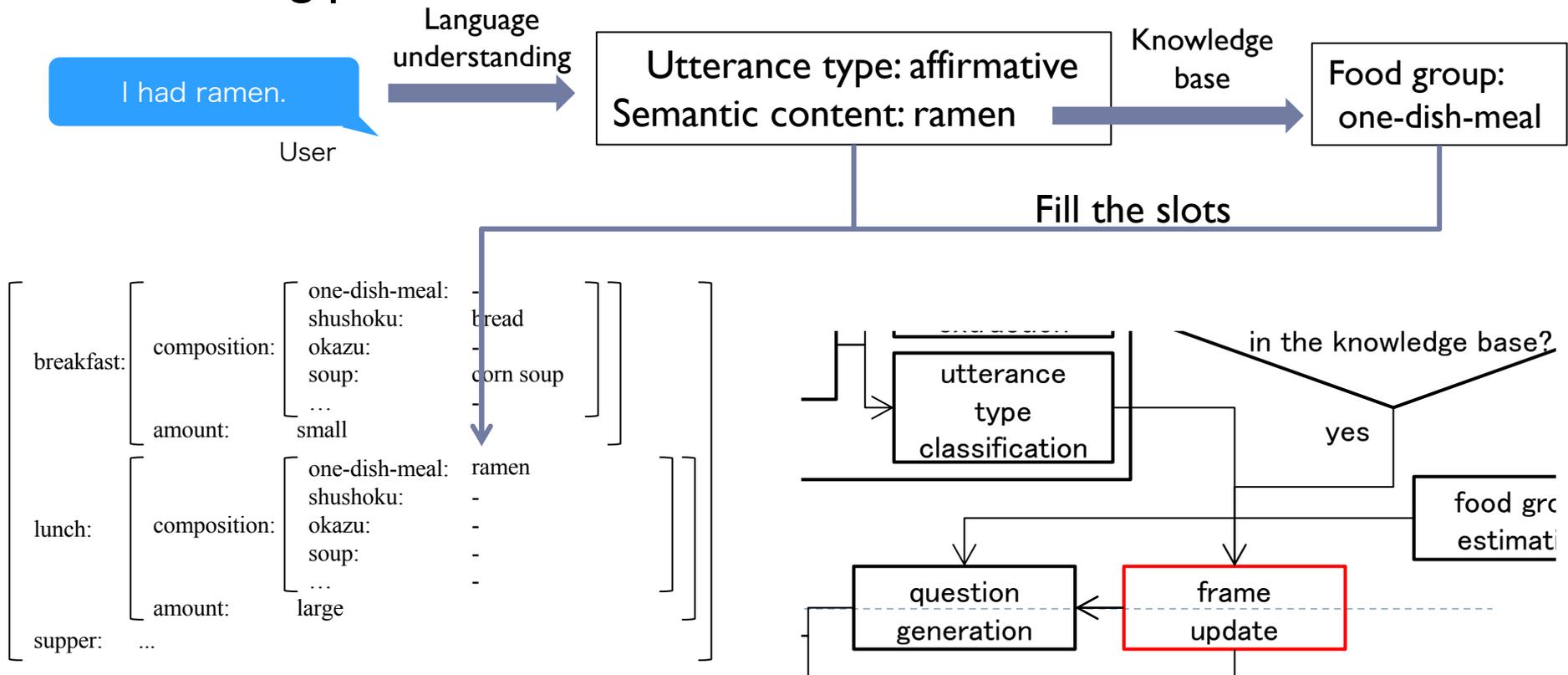
Language Understanding

- ▶ **Utterance type classification using logistic regression**
 - ▶ Three types: *greeting, affirmative utterance, negative utterance*
 - ▶ The system updates the frame when user utterance type is affirmative
 - ▶ Logistic regression with a BoW of user input as features
- ▶ **Semantic content extraction using CRF**
 - ▶ Five semantic contents are extracted from user's input: *food and drink, ingredients, food group, amount of food, time*
- ▶ **3,659 sentences are used for training data**



Slot Filling

- ▶ Based on the language understanding, the frame is updated
- ▶ The frame is composed of slots for Japanese ordinary composition of meal
- ▶ Slot filling procedure:



Knowledge Base

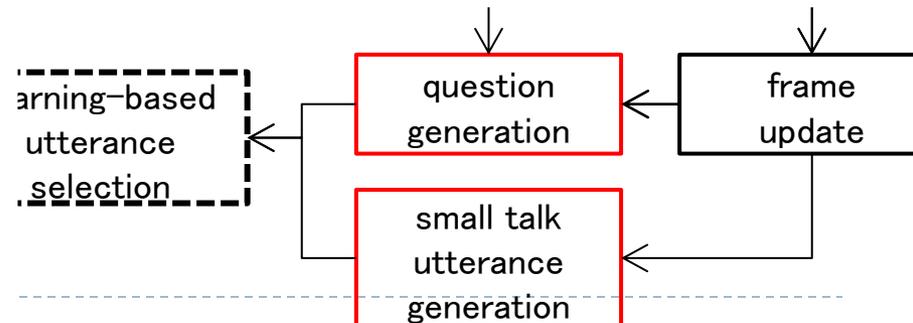
- ▶ The knowledge base is utilized to get the food group of a food or drink name
- ▶ This knowledge base consists of seven food groups and 2,134 instances
 - ▶ Food groups are based on Japanese meal

Food group	Example instances	#
<i>shushoku</i> (side dish mainly containing carbohydrates)	steamed rice, bread, cereal	152
<i>okazu</i> (main or side dish containing few carbohydrates)	hamburg steak, fried shrimp, grilled fish	668
soup	corn soup, miso soup	70
one-dish meal	sandwich, noodle soup, pasta, rice bowl	695
drink	orange juice, coffee	343
dessert	cake, pancake, jelly	134
confectionery	chocolate, donut	72
total		2,134



Generating Utterance Candidates

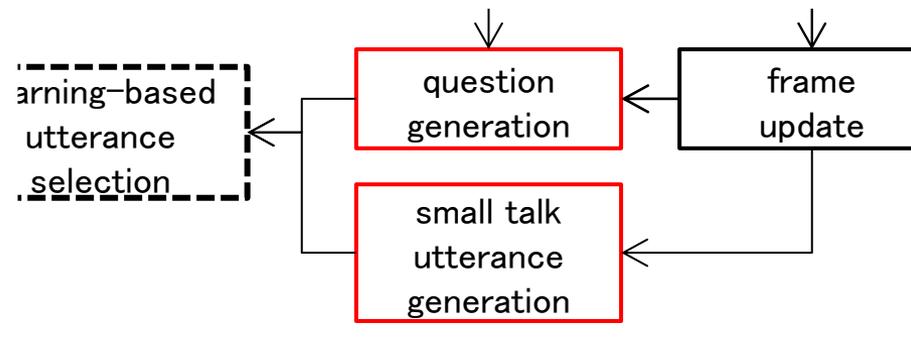
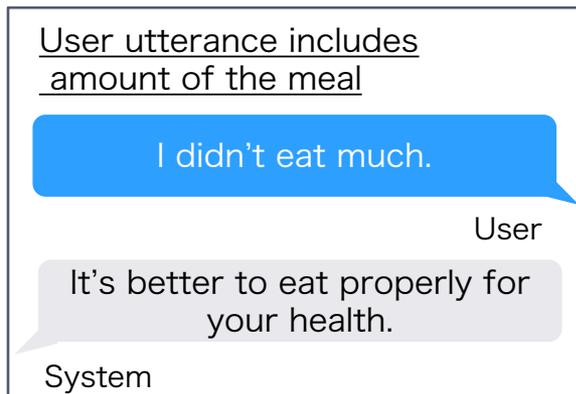
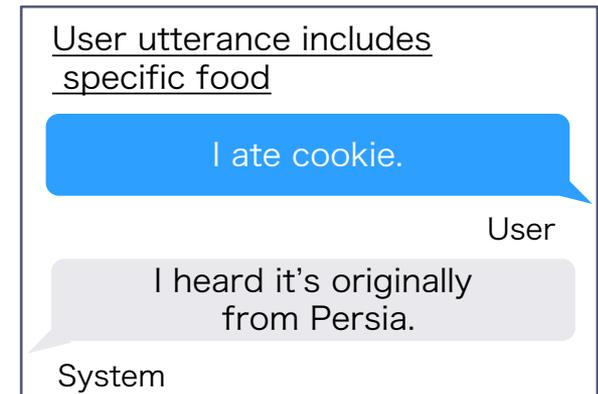
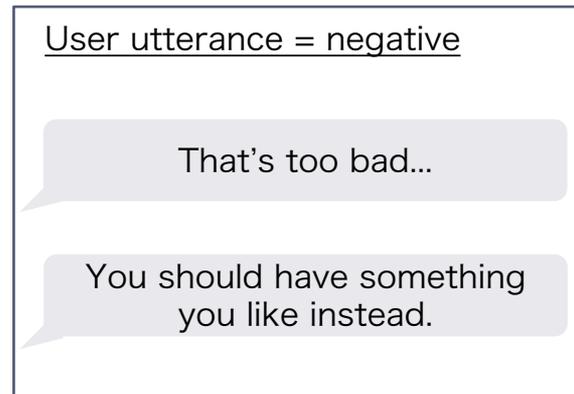
- ▶ Utterance candidates are generated based on language understanding results and current frame
- ▶ Question candidates
 - ▶ Generate questions to fill vacant slots by referencing the frame
 - ▶ Generate questions in various forms
 - ▶ Example:
 - ▶ What did you have for breakfast? } Question for asking food
 - ▶ Did you have another food for breakfast? }
 - ▶ How much did you have breakfast? } Question for asking amount of food
 - ▶ Did you have a lot? }



Generating Utterance Candidates

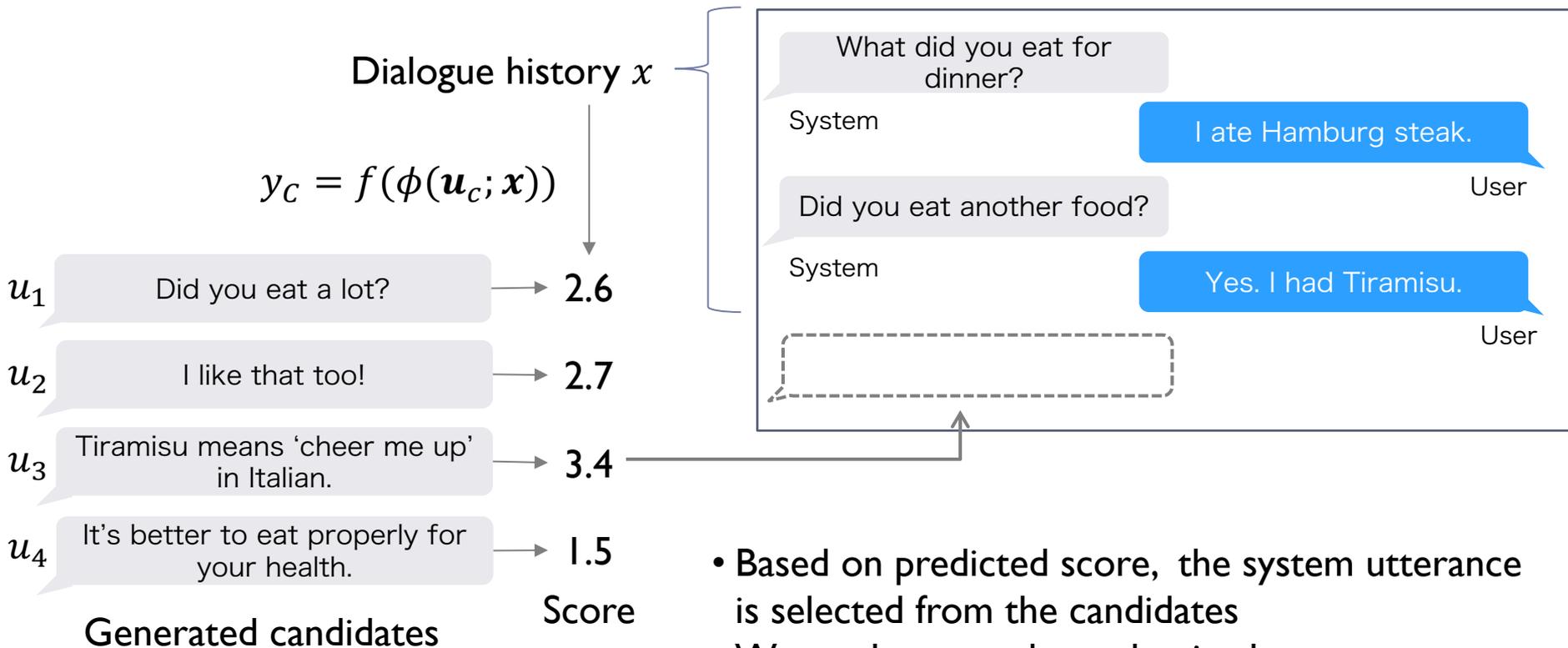
▶ Small talk utterance candidate

- ▶ The candidates are generated from predefined 407 patterns of small talk utterances depending on the user utterance



Estimating Appropriateness of Candidates

- ▶ Appropriateness score y_c of each candidate u_c is estimated based on regression of features w.r.t. u_c and dialogue history x

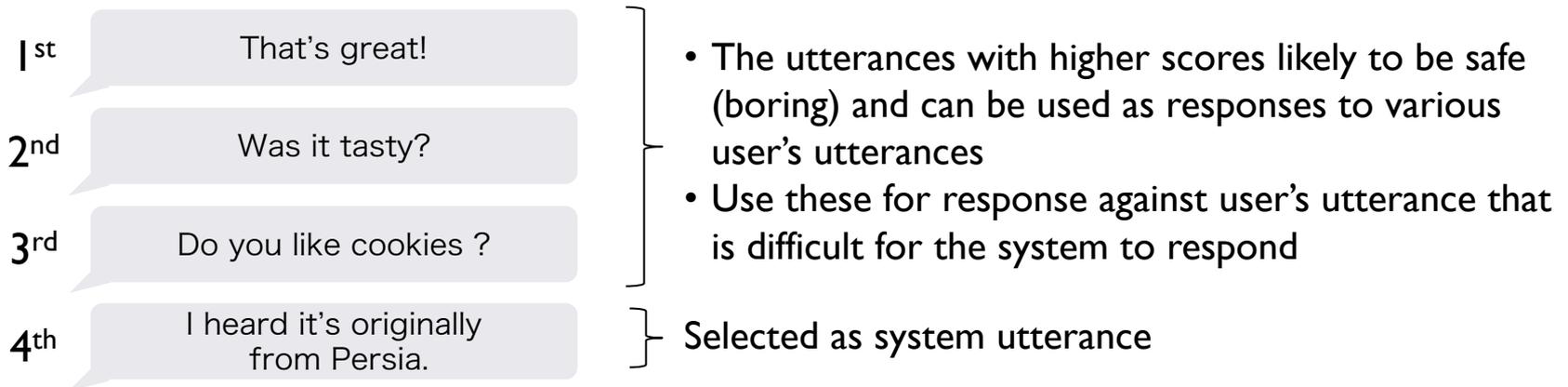


- Based on predicted score, the system utterance is selected from the candidates
- We used extremely randomized tree regressor (ETR) as regression function f and 4,523 utterances are used for training

Utterance Selection

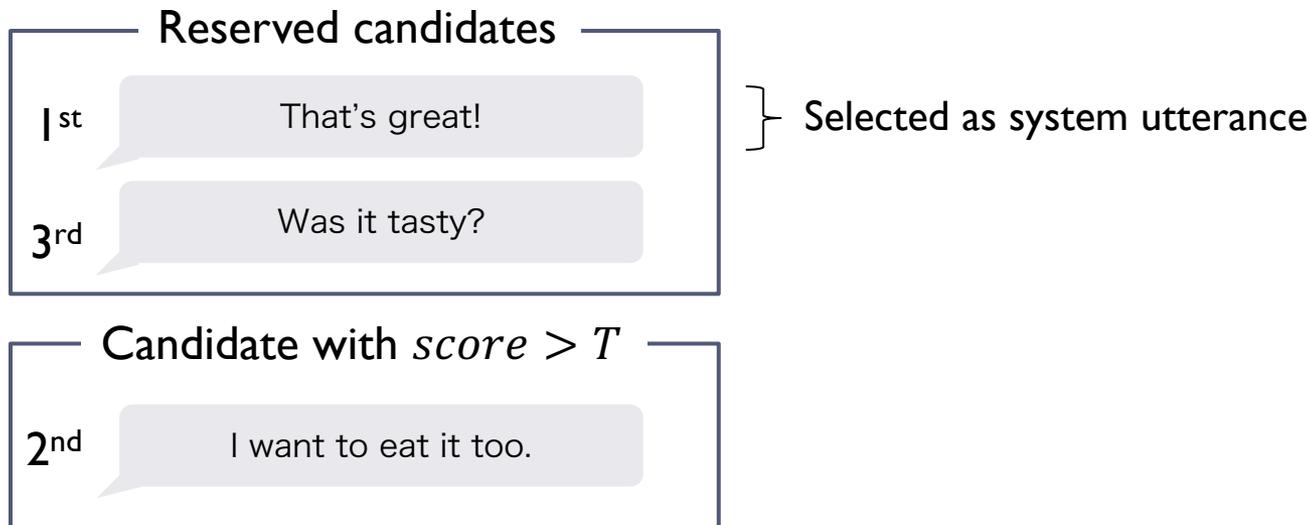
- ▶ A system utterance is selected from the candidates according to the following four rules

1. Select from the candidates whose scores are greater than T
➔ Prevent unnatural response
2. Not select utterances that are already used
➔ Prevent unnatural repetition
3. Top three candidates are reserved, and the forth candidate is selected



Utterance Selection

- ▶ One candidate is selected as system utterance by the following four rules from the candidates
 4. If # of candidates is less than four, select the one with the highest score from candidates including reserved candidates



This prevents dialogue breakdown by using safe (boring) utterances



Experimental Setup

- ▶ We conducted a user study to investigate the effectiveness of our proposed method
- ▶ Compare three methods
 - ▶ NO-STU: Interview without small talk
 - ▶ I-STU (baseline)[Kobori+ 16]:
 - ▶ Generates one small talk after each user utterance
 - ▶ The system utterance is selected from candidates randomly
 - ▶ ETR (proposed):
 - ▶ System utterances are selected using ETR by the proposed method
 - ▶ Threshold value: $T = 2.4$



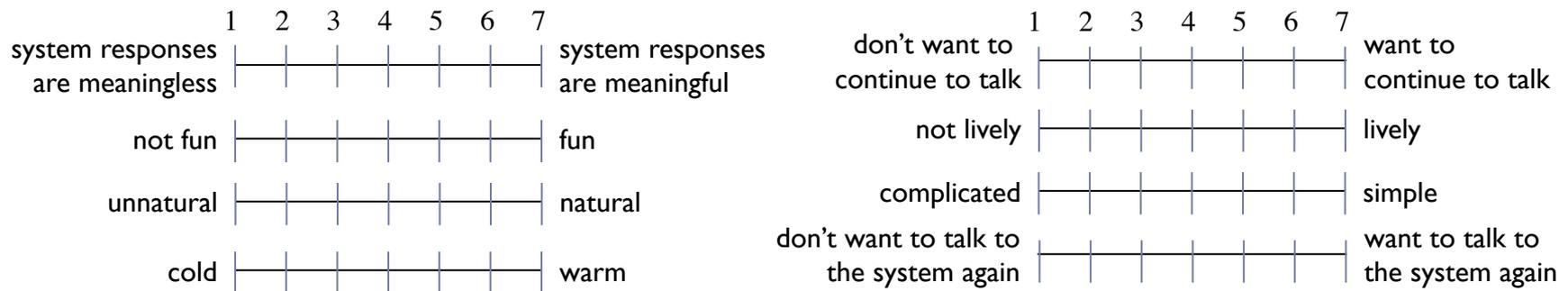
Experimental Setup: Subjects

- ▶ We asked subjects to use assigned system four times in a week
 - ▶ To find how the impression against the same system changes depending on the number of times of its use
- ▶ 40 people interacted with each system
 - ▶ We excluded the participants who could not use the system appropriately because of the system errors
 - ▶ # of participants used in this investigation
 - ▶ NO-STU:17
 - ▶ I-STU:20
 - ▶ ETR:20



Experimental Setup: Evaluation Measure

- ▶ After the dialogue, we asked the participants to evaluate the dialogue using a seven-point Likert scale for 8 survey items

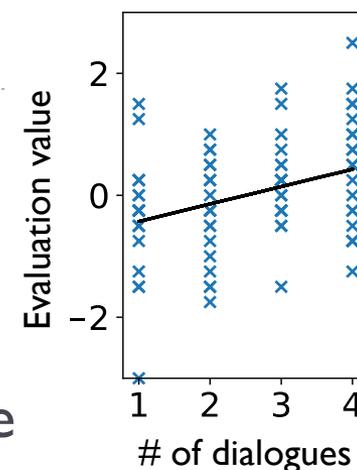


- ▶ Each subject used and evaluated the assigned system four times a week
-



Experimental Results

- ▶ We conducted a regression analysis to find how the evaluation changed
 - ▶ Explanatory variable: the number of dialogues
 - ▶ Dependent variable: the normalized evaluation value
- ▶ Significant items (*: $p < 0.05$, **: $p < 0.01$)

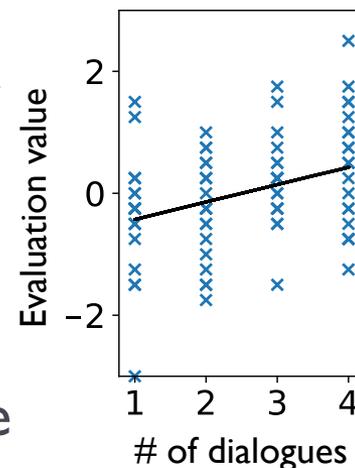


	Meaningful	Fun	Natural	Warm	Continue to talk	Lively	Simple	Want to talk again
NO-STU	-	-	-	-	-	-	-	Negative*
I-STU (baseline)	-	-	-	-	-	-	-	-
ETR (proposed)	Positive**	Positive**	Negative*	-	Negative*	Positive*	-	-



Experimental Results

- ▶ We conducted a regression analysis to find how the evaluation changed
 - ▶ Explanatory variable: the number of dialogues
 - ▶ Dependent variable: the normalized evaluation value
- ▶ Significant items (*: $p < 0.05$, **: $p < 0.01$)

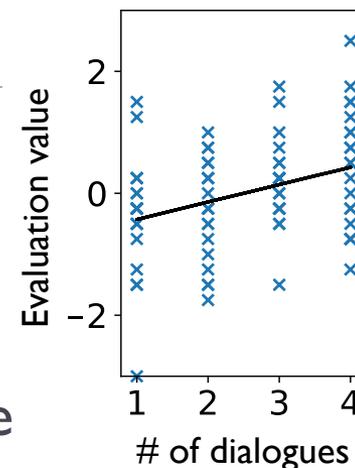


	Meaningful	Fun	Natural	Warm	Continue to talk	Lively	Simple	Want to talk again
NO-STU	-	-	-	-	-	-	-	Negative*
I-STU (baseline)	-	-	-	-	-	-	-	-
ETR (proposed)	Positive**	Positive**	Negative*	-	Negative*	Positive*	-	-

- Can select meaningful responses and the users feel small talk is fun
- Generate unnatural utterances in the latter part because the # of candidates were decreasing in progress of the dialogue

Experimental Results

- ▶ We conducted a regression analysis to find how the evaluation changed
 - ▶ Explanatory variable: the number of dialogues
 - ▶ Dependent variable: the normalized evaluation value
- ▶ Significant items (*: $p < 0.05$, **: $p < 0.01$)



	Meaningful	Fun	Natural	Warm	Continue to talk	Lively	Simple	Want to talk again
NO-STU	-	-	-	-	-	-	-	Negative*
I-STU (baseline)	-	-	-	-	-	-	-	-
ETR (proposed)	Positive**	Positive**	Negative*	-	Negative*	Positive*	-	-

- Users were bored and did not want to talk with the system without small talk again
- This indicates that small talk might prevent this negative effect

Conclusion

- ▶ We have developed an interview dialogue system that generates small talk utterances
- ▶ We utilized system utterance selection based on the appropriateness score estimated by the regression model
- ▶ We conducted a user study to investigate the effectiveness of our proposed method
 - ▶ As the number of dialogues increases, “meaningful”, “fun” and “lively” increased, and “natural” and “continue to talk” decreased
 - ▶ “want to talk again” did not have correlation with the number of dialogues, which is different from our expectation
- ▶ We plan to investigate the results in detail

