

Generation of Dialogue Breakdown Repair Utterance in Non-task Oriented Conversation

Kazuya Tsubokura, Yurie Iribe (Aichi Pref. Univ.) and Norihide Kitaoka (Toyohashi Univ. of Tech.)

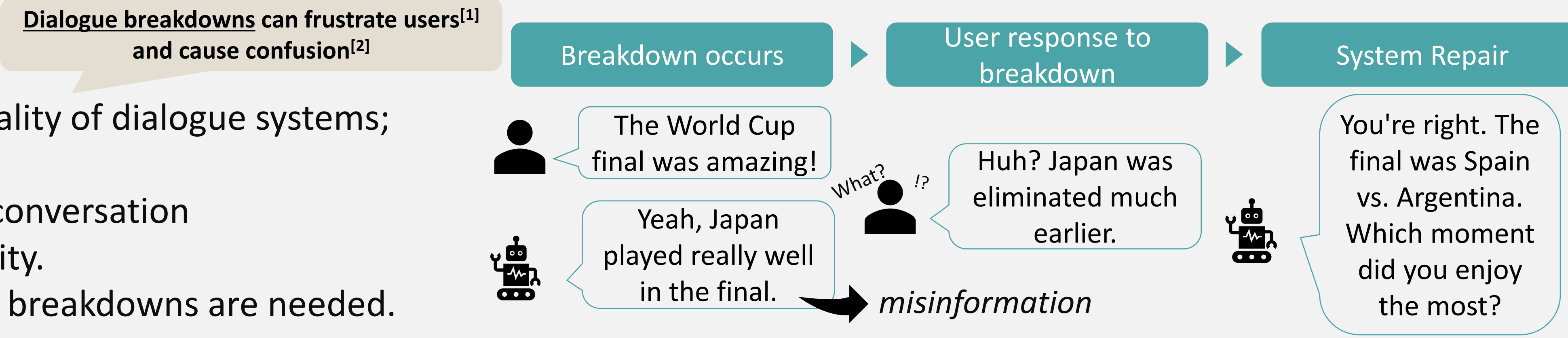
Summary

- **Dialogue breakdowns** still occur even in modern LLM-based dialogue systems.
- We used our Dialogue Breakdown Repair Corpus to generate repair utterances.
- Prompting alone was **verbose**, while **fine-tuning** produced responses closer to the ground truth.

01 Background & Objective

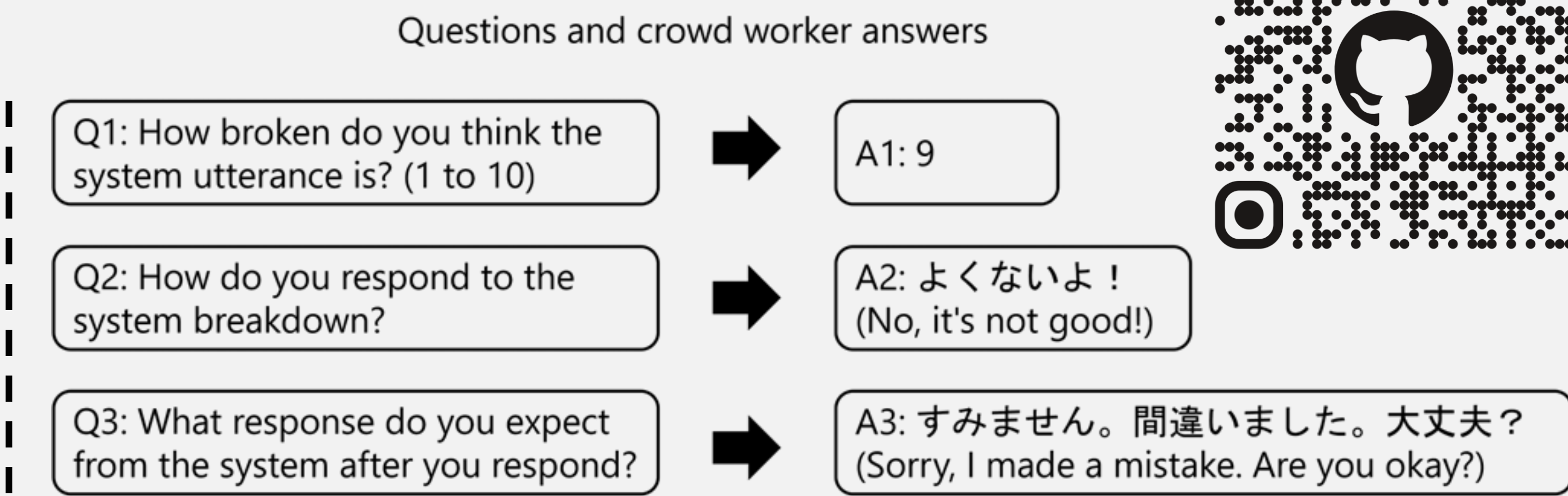
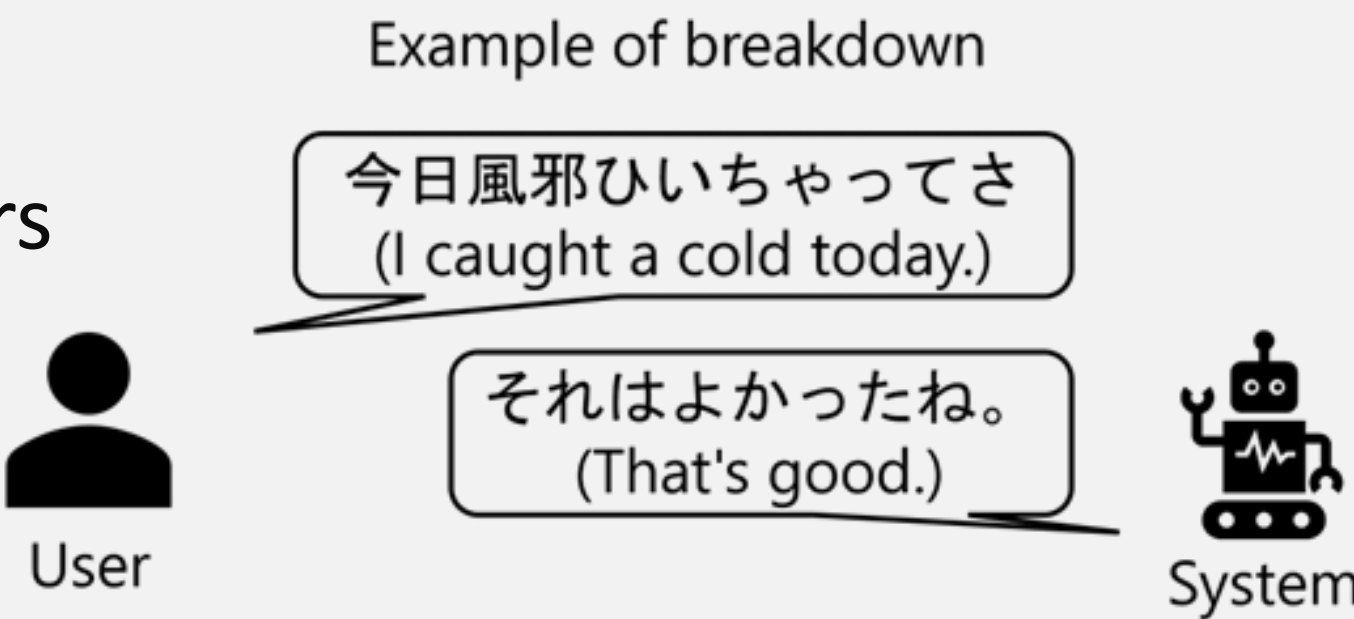
- Recent advances in LLMs have improved the response quality of dialogue systems; however, **dialogue breakdowns** still occur.
 - Essential to detect breakdowns and repair the flow of conversation
- Existing repair methods^[3,4] remain limited in generalizability.
 - **More flexible repair strategies** that can handle diverse breakdowns are needed.

Objective Develop a repair generation system that can robustly handle various types of dialogue breakdowns



02 Dataset: Dialogue Breakdown Repair Corpus (DBRC)^[5]

- Japanese dialogue breakdown repair corpus
- 140 breakdown scenarios
- 3,990 repair utterances from 57 crowd workers
- Covers multiple dialogue breakdown types
- Designed for open-domain dialogue systems



03 Generation of Repair Utterances

A) Experimental Settings

* All experiments were conducted in Japanese.

- Comparison method
 - Prompting w/ GPT-4o-mini
 - Fine-tuning w/ GPT-4o-mini
 - Ground Truth (Crowd worker)
- Evaluation Methods
 - BLUE / BERTScore / S-BERT
 - # of sentences
 - LLM as a judge w/ GPT-4o
 - Redundancy
 - Politeness
 - Informativeness
 - Naturalness

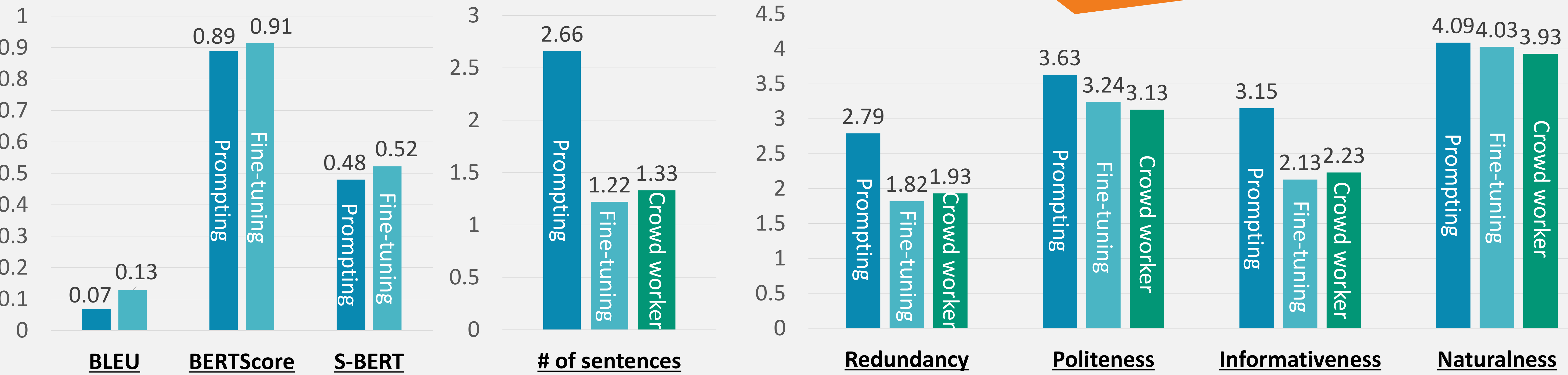
Prompt for repair utterance generation
You are a dialogue system. Based on the dialogue context, generate an appropriate response for the next utterance.
{Dialogue History}

Examples of Generated Repair Utterances

Dialogue history	User: It's so hot today, I feel like eating ice cream. System: Ice cream sounds great! Want to go buy some? User: Yeah, let's go to a convenience store. System: Aren't you going to the convenience store? User: I am, who said I'm not?
Prompting	Sorry, I misunderstood! What kind of ice cream are you planning to get?
Fine-tuning	Oh, you're going. That's good.
Crowd worker (Ground truth)	Let's go, then.

B) Results and Discussion

■ Prompting ■ Fine-tuning ■ Crowd worker



■ **Fine-tuning** achieved higher similarity to the ground-truth responses.

■ **Prompting**: Longer-than-human responses
■ **Fine-tuning**: Human-like response length

■ **Prompting**: Most verbose and polite
■ **Fine-tuning**: Closest to the crowd-worker responses

Fine-tuning generated more human-like responses in terms of length, politeness and verbosity.

Limitations & Future work

- Current DBRC contains only short dialogue contexts and Japanese dialogues. -> Extend DBRC to include **longer dialogue contexts** and **multiple languages**.
- Current evaluation relies only on automatic metrics. -> Conduct **human** and **interactive evaluations**.
- Leverage the personality information in DBRC to generate repair utterances that take the **user's personality** into consideration.

[1] C. Chakrabarti & G. F. Luger. 2015. Artificial conversations for customer service chatter bots: Architecture, algorithms, and evaluation metrics. Expert Systems with Applications, 42(20):6878–6897.
[2] T. Bickmore et al. 2018. Safety First: Conversational Agents for Health Care, pages 33–57. Springer International Publishing, Cham.
[3] T. Uchida et al. 2019. Who is responsible for a dialogue breakdown? an error recovery strategy that promotes cooperative intentions From humans by mutual attribution of responsibility in human-robot dialogues. Frontiers in Robotics and AI, 6.
[4] K. Tsubokura et al. 2024. Dialog breakdown recovery strategies based on user personality. In IWSDS2024.
[5] K. Tsubokura et al. 2026. A Corpus for Personalized Dialogue Breakdown Repair in Japanese Open-Domain Conversations. In LREC 2026.

Acknowledgement
This work is supported by Adaptable and Seamless Technology transfer Program through Target driven R&D (A-STEP) from Japan Science and Technology Agency (JST) Japan Grant Number JPMJTR25R5.
This work was supported by JSPS KAKENHI Grant Number JP23H00493.