Question Types for Knowledge Acquisition in LLM-based Dialogue Systems: Experiments with Simulated and Human Users

Kazunori Komatani[1], Ryu Takeda[1], Mikio Nakano[1,2]
1: University of Osaka, 2: C4A Research Institute, Inc.

SANKEN THE UNIVERSITY OF OSAKA

Summary

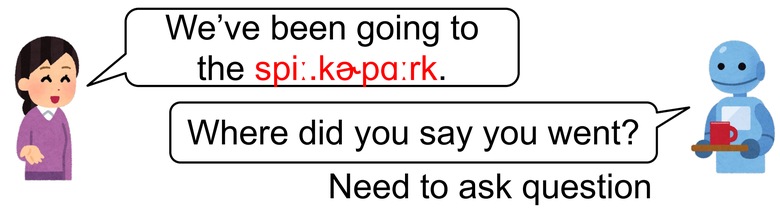
Objective
- To re-examine implicit, explicit, and wh-questions for knowledge acquisition in LLM-based dialogue

Method
- Evaluation with both an LLM-based user simulator and human users recruited via crowdsourcing

Findings
- User annoyance: Little difference across question types
- Correct response elicitation: Candidate-guided questions were more reliable than wh-questions

Background / Motivation

Knowledge acquisition during dialogue
- Unknown terms or user-specific knowledge
  - Not necessarily available in pre-trained LLMs



- **How to ask** is important
  - Not only **whether to ask** [Waki+ 2025] [Furuta+ 2026]
  - Studied in template-based dialogue [Komatani+ D&D 2022]
  - **Do the same tendencies hold in LLM-based dialogue?**

Experimental Setup

Task

- In our task, a user mentions one of 30 "unknown" food terms
- The system asks a follow-up question to acquire its country of origin

Users

Same task and comparable instructions

LLM-based user simulator
100 sessions, 1,000 dialogue blocks

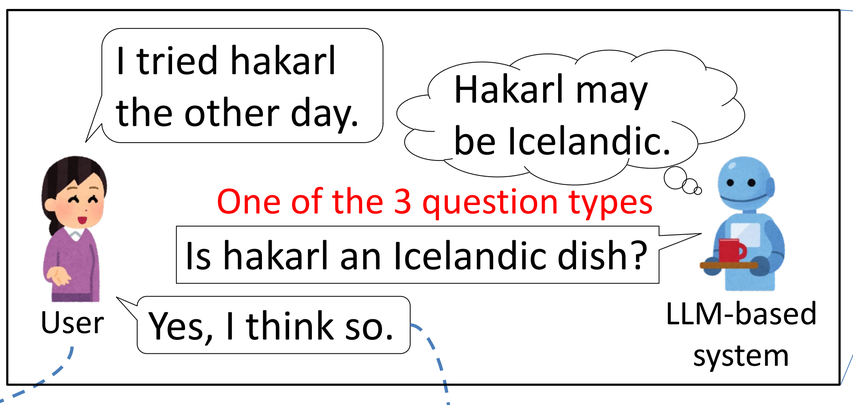
Human users
273 sessions, 1,135 dialogue blocks

Question types

Same as [Komatani+ D&D 2022]

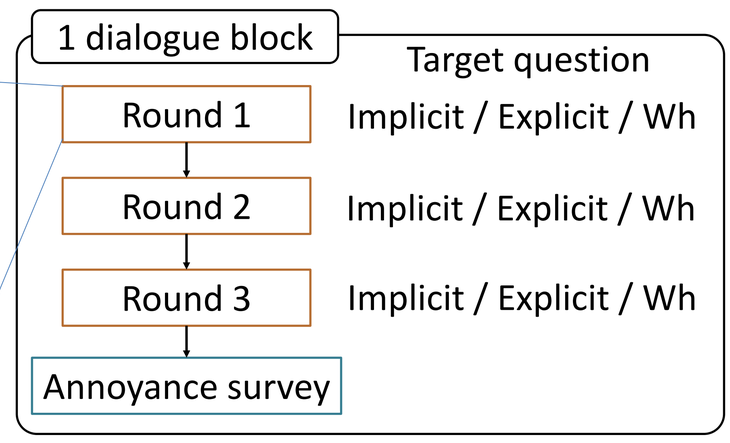
Implicit	Icelandic food is distinctive.
Explicit	Is hakarl an Icelandic dish?
Wh	What country is hakarl from?

For implicit and explicit questions, the candidate country was always correct.



System

- LLM-generated casual utterances
- Follow-up questions generated according to a state-transition network (STN) with DialBB
- Text chat in Japanese



x10 /session for user simulator
x5 /session for human users

Metrics

User annoyance
- Rated on a 1–5 scale after each dialogue block
- Estimated regression coefficients for the counts of the three question types in each block

$$annoyance = w_0 + w_1 Impl. + w_2 Expl. + w_3 Whq$$

Correct response elicitation
- Proportion of system questions followed by a correct user response
- Judged automatically by LLM
  - Manual check: 96/100 judgments agreed

Results

User annoyance

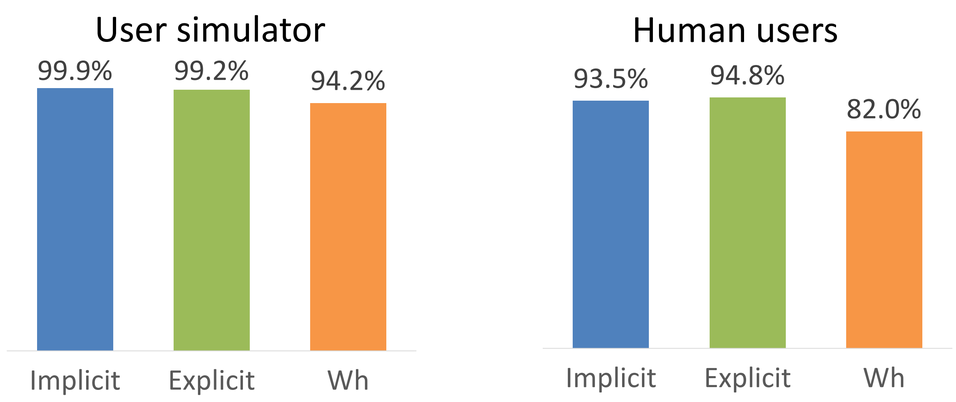
- Differences among question types were small
- Question type explained little variance in annoyance

	Implicit	Explicit	Wh	const.	R^2
User simulator	0.121	0.132	0.131	1.363	**0.015**
Human users	−0.113	−0.226	−0.144	2.953	**0.010**

Separate regressions for the two settings

Correct response elicitation

- Implicit and explicit questions performed comparably
- Wh-questions elicited correct responses less often



Same tendency in both settings: Implicit ≈ Explicit > Wh

Discussion

User annoyance
- Limited effect of question type
- Answer-finding effort may affect annoyance in the present setting
  - Wh-question: "What country is hakarl from?"
    → User may need to look up the answer before responding
- Unlike our previous template-based study [Komatani+ D&D 2022], explicit questions were not more annoying

Correct response elicitation
- Candidate-guided questions enable confirmation
- Wh-questions require open elicitation
- Granularity mismatch can occur

> System: What country is khao man gai from?
> User: It is popular in Southeast Asia.

Implications
- Select question types based on system confidence and purpose
  - **Explicit**: clear confirmation
  - **Implicit**: natural hypothesis strengthening
  - **Wh**: no plausible candidate

Limitations: food domain only; Japanese only; correct candidates only
Future work: Other domains/languages; incorrect candidates