

# **Dialogue Systems Can Generate Appropriate Responses without the Use of Question Marks?**

– A Study of the Effects of “?” for Spoken Dialogue Systems –

**Tomoya Mizumoto, Takato Yamazaki, Katsumasa Yoshikawa,  
Masaya Ohagi, Toshiki Kawamoto, Toshinori Sato**

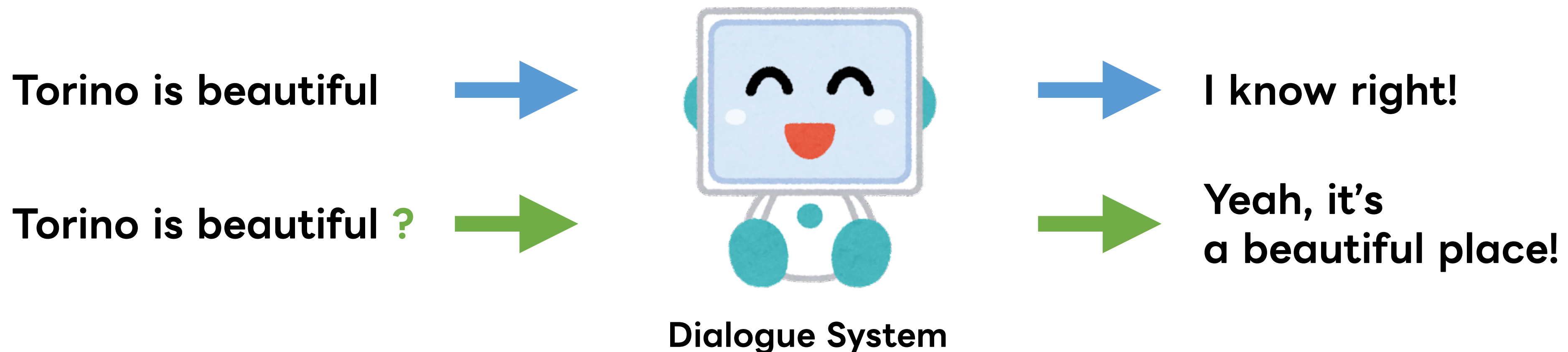
LREC-COLING 2024

**LY**

# Background (1/2)

---

- In spoken dialogue, **raising the intonation** at the end of a sentence can indicate a **question**
- When converted to text (e.g. ASR), **question marks** may not be present
  - Difficult to tell if a sentence is **declarative** or **interrogative**
  - NLP-based dialogue systems may fail to respond correctly to questions



## Background (2/2)

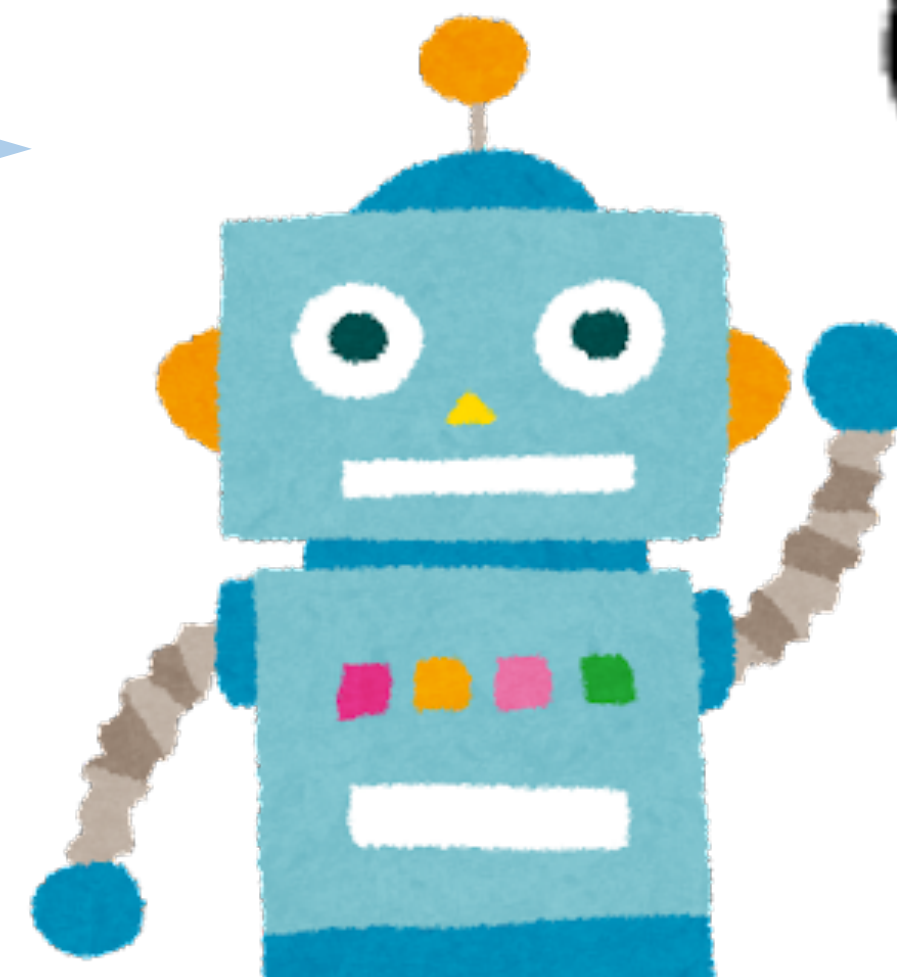
---

- Especially in **Japanese**, absence of question mark is more critical, because **subjects** can be omitted  
→ Easier to misinterpret without intonation in contrast to other languages

トリノに行きます?  
Going to Torino?

?  
No, I was asking you.

Nice!



Is he asking me?  
Or is he going?

# Research Question

---

Can dialogue systems...

respond correctly without question marks?

.ll

recognize if it is a question from **contexts**?

**Let's see how absence of question marks affect the system!**

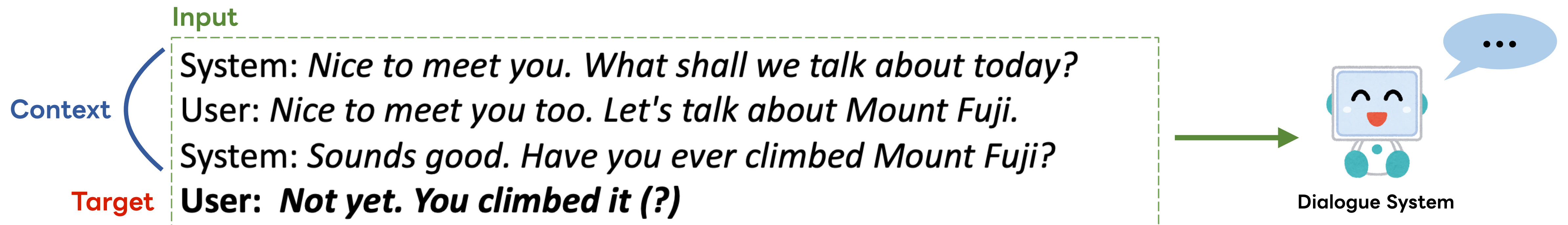
# Research Methodology

---

1. Collect **utterances** from dialogue datasets that **ends with a question mark** (and remove it)
2. **Generate responses** with dialogue systems
3. Compare them **with and without question marks**

## Note

- Used the **three utterances** preceding the question utterance as **context**



# Experiment Settings: Dataset

---

- We used three types of **Japanese** dialogue data
  - **NUCC**: casual conversations **between native Japanese speakers**
  - **DSL**: casual conversations **between human evaluators and dialogue systems**
  - **JED**: dialogues between two native Japanese speakers discussing an event while **expressing a range of emotions**

# Experiment Settings: Dialogue System

---

- **Transformer** [Sugiyama+ 2020]
  - System of the winning team in Live Competition 3
  - Transformer-based system, fine-tuned on **JPersonaChatDialogue** data
- **HyperCLOVA** [Kim+ 2021]
  - in-house GPT-based LLM with **80B params**
  - A dialogue system was developed using **prompt programming**

Hiroaki Sugiyama, et al. "**Building a Hobby Conversation System Using a Transformer Encoder-Decoder Model**" in Proceedings of the 90th SIG-SLUD, JSAI, 2020, pp. 24. (in Japanese)

Kim Boseop et al. "**What Changes Can Large-scale Language Models Bring? Intensive Study on HyperCLOVA: Billions-scale Korean Generative Pretrained Transformers**". arXiv preprint arXiv:2109.04650.

# Experiment Settings: Evaluation Methods

---

- Evaluate responses generated by the dialogue system **manually**
  - Conduct binary evaluation of either **appropriate** or **not appropriate**
  - Evaluate each dialogue example with **three people**
  - The agreement exceeded **60%**, indicating **high level of agreement**

	NUCC	DSLCL	JED
Agreement	61.2	82.1	61.3



# Results (1/2)

	NUCC		DSLCL		JED	
	?	None	?	None	?	None
Transformer	97	65	96	72	98	89
HyperCLOVA	94	82	98	82	100	98

Degrade

(% Appropriate)

- The presence or absence of a “?” mark has a significant impact
  - With a “?” mark: appropriate responses were achieved about **90%**
  - Without a “?” mark: the rate of appropriate responses **drops about 20%**

## Results (2/2)

	NUCC		DSLCL		JED	
	?	None	?	None	?	None
Transformer	97	65	96	72	98	89
HyperCLOVA	94	82	98	82	100	98

(% Appropriate)

### Comparison between systems

- The performance degradation of not having a “?” mark is **smaller** in dialogue systems based on **HyperCLOVA (≠LLM)**
  - Large parameter models can infer if it is a question **using contexts**

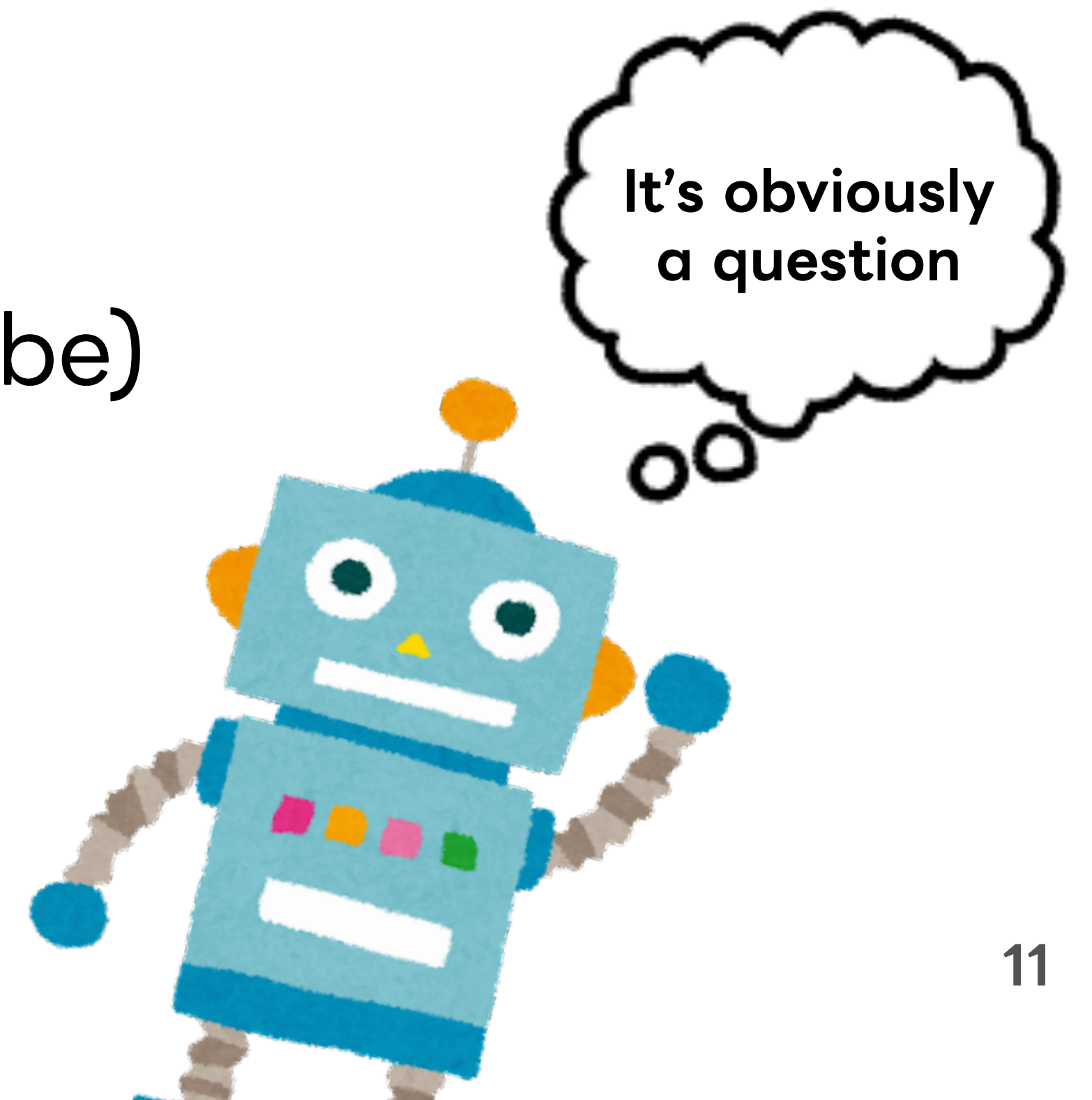
# Findings (1/3)

---

- Systems were able to answer appropriately when the **sentence ended** with a **particle indicating a question**, such as **"か"** (ka), or with **interrogative words** like the 5W1H

「YouTubeは見ます**か**」 (**Do you** watch YouTube)

「**どんな**場所です」 (**What** kind of place)

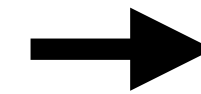


# Findings (2/3)

- Comparing with and without **the interrogative particle "か" (ka)**, system often fails to respond appropriately, even with the same context

Question (w/o "か")

User 登りました (You climbed)



## Generated Response

Transformer w/o "?" いいですね。✗  
(That's nice.)

HyperCLOVA w/o "?" すごいですね。✗  
(That's amazing)

Question (w/ "か")

User 登りましたか (Have you climbed)



## Generated Response

Transformer w/o "?" 無いですね。✓  
(No, I haven't.)

HyperCLOVA w/o "?" 登山したことないですね。✓  
(I haven't gone climbing.)

## Findings (3/3)

---

- In some cases, **HyperCLOVA** was able to generate an appropriate response without “?”, by **inferring a question from the context**

### Context

**System** そんなの聞かなくていいです。 (You don't have to ask that.)

**User** じゃあね。あの棚使ってる? (Are you using that shelf over there?)

**System** あの棚、使ってますよ。 (I'm using that shelf.)

**User** 組み立てられた?  
(I/you managed to assemble it?)

### Generated Response

**HyperCLOVA w/o “?”** 組み立てられたけど大変でした。 ✓  
(I was able to assemble, but it was difficult.)

**Transformer w/o “?”** 組み立て大変ですよ。 ✗  
(It's difficult to assemble, isn't it.)

# Conclusion

---

We investigated how the **absence of question mark** affects the response of NLP-based dialogue systems

- Dialogue systems **often misinterprets** user's utterance without question mark, and fails to respond appropriately
- **Larger models were more capable** of inferring if the user utterance is a question from contexts
- Even with contexts, there were cases which failed to respond appropriately
  - Need to explore methods that use **non-linguistic cue** (e.g. audio information, facial expressions)