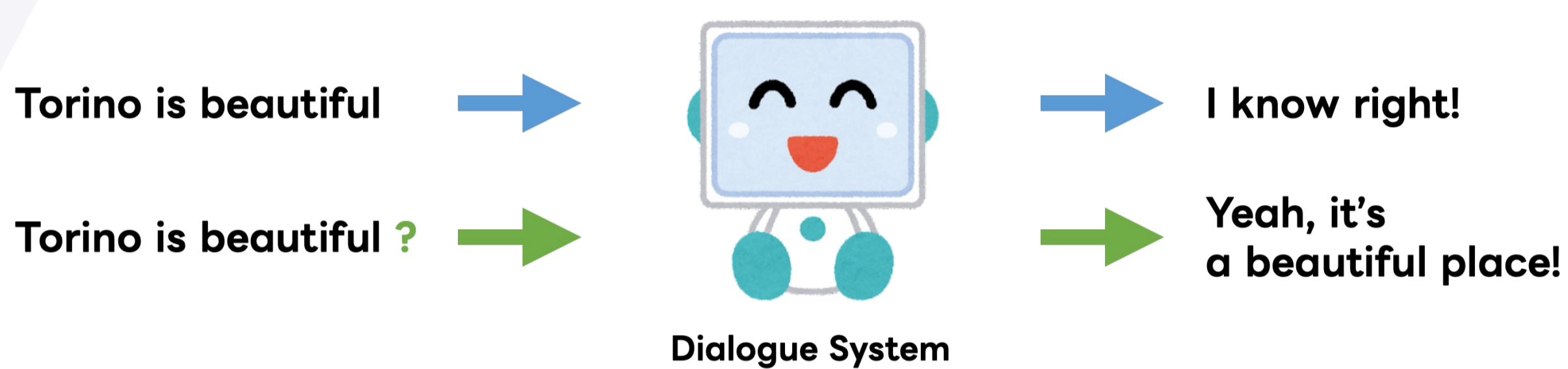


Summary

Research Purpose:

Clarifying the importance of question marks at the end of sentences in dialogue systems

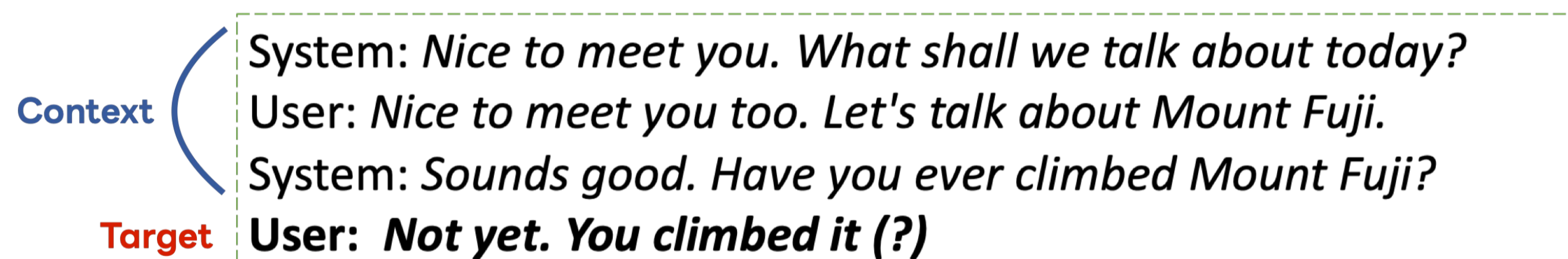
- In spoken dialogue, **raising the intonation** at the end of a sentence can indicate a question
- When converted to text, there is no difference between interrogative and declarative sentences
- NLP dialogue systems may fail to respond correctly to questions
- Investigate whether the presence of a question mark affects the appropriateness of the response



Research Methodology

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

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



Dataset

We used three types of Japanese dialogue dataset:

- NUCC**: casual conversations between native Japanese speakers [Fujimura+, 2012]
- DSL**: chit-chat conversations between human evaluators and dialogue systems
- JED**: dialogues between two native Japanese speakers discussing an event while expressing a range of emotions [Sugiyama+, 2021]

Dialogue Systems

- Transformer**: System of the winning team in Live Competition 3 [Sugiyama+ 2021]
 - Transformer-based model fine-tuned on JPersonaChatDialogue
- HyperCLOVA**: in-house GPT models with **80B** parameters [Kim+ 2021]
 - Developed using prompt programming

Evaluation Method

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

	NUCC	DSL	JED
Agreement	61.2	82.1	61.3

Results

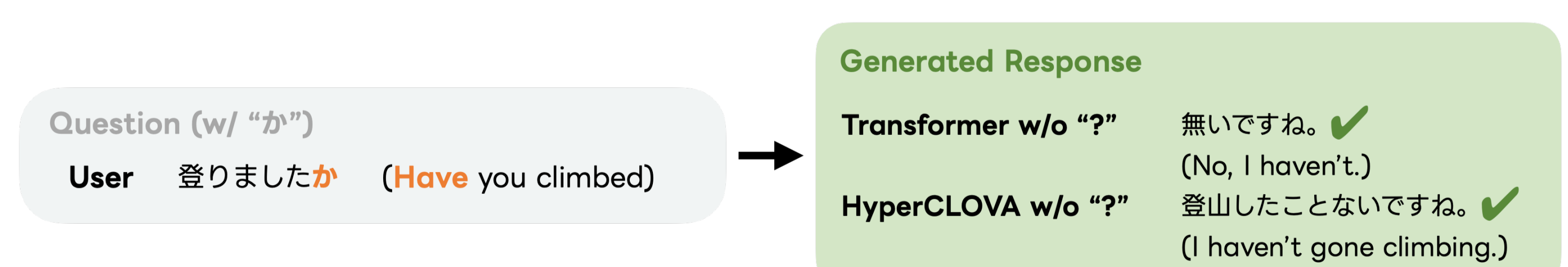
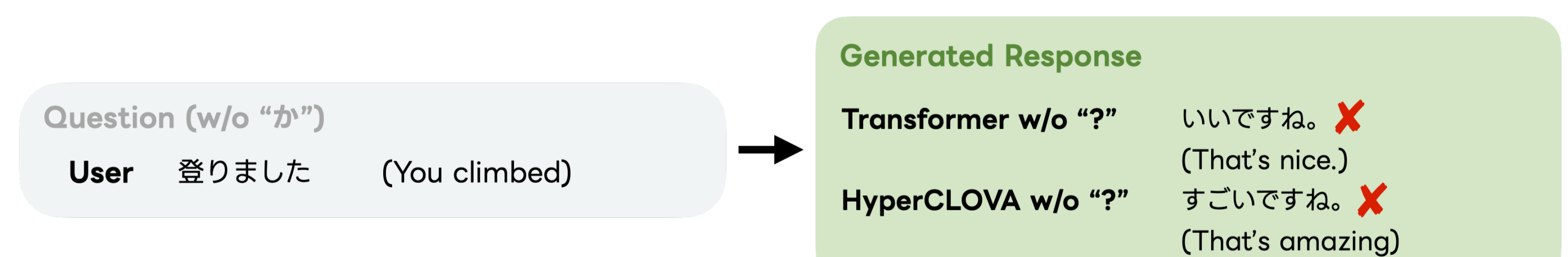
- The presence/absence of a “?” has a **significant impact**
 - In cases with a “?” mark, appropriate responses were achieved about **90%**
 - In cases without a “?” mark, the rate of appropriate responses drops about **20%**
- The performance degradation due to the lack of “?” was less in HyperCLOVA than Transformer
 - Models with **large parameters are capable of inferring the user intention** from the context

Percentage of appropriate responses

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

Discussion

- The impact of a “?” mark was small 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)
- When without the interrogative particle “か” (ka), system often fails to respond appropriately, even with the context



- In some cases, HyperCLOVA was able to appropriately respond without “?”, by inferring from the context

