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***

LY Corporation

Kioi Tower 1-3 Kioicho, Chiyoda-ku, Tokyo 102-8282, Japan

{tomoya.mizumoto, takato.yamazaki, katsumsa.yoshikawa, masaya.ohagi}@lycorp.co.jp

Abstract

When individuals engage in spoken discourse, various phenomena can be observed that differ from those that are apparent in text-based conversation. While written communication commonly uses a question mark to denote a query, in spoken discourse, queries are frequently indicated by a rising intonation at the end of a sentence. However, numerous speech recognition engines do not append a question mark to recognized queries, presenting a challenge when creating a spoken dialogue system. Specifically, the absence of a question mark at the end of a sentence can impede the generation of appropriate responses to queries in spoken dialogue systems. Hence, we investigate the impact of question marks on dialogue systems, with the results showing that they have a significant impact. Moreover, we analyze specific examples in an effort to determine which types of utterances have the impact on dialogue systems.

Keywords: Dialogue System, Spoken Dialogue, Large Language Model, Evaluation

1. Introduction

Given the increasing ubiquity of smartphones and smart speakers such as Siri, Alexa, and Google Assistant, voice commands have become a routine means of operating devices. In the domain of dialogue system research, some competitions have been conducted to facilitate multimodal dialogues between a user and, android robots (Higashinaka et al., 2022a; Minato et al., 2022) or computer graphics agents (Higashinaka et al., 2022b), that go beyond the scope of natural language processing. Numerous dialogue systems that have been used in these competitions have been constructed using textual corpora, rendering them suboptimal for spoken dialogues.

In spoken dialogue, difficulties arise that are not present in text-based dialogues. One such issue is the lack of punctuation marks, such as question marks, in the text recognized by automatic speech recognition (ASR). Some reports have indicated that the absence of punctuation marks in transcribed speech can negatively impact human comprehension (Tündik et al., 2018) and performance for subsequent tasks, such as machine translation (Peitz et al., 2011), named entity recognition (Nguyen et al., 2020) and information extraction (Makhoul et al., 2005). Efforts have been made to address this issue through research on the automatic prediction of punctuation marks, but **the punctuation attachment accuracy is still around 80%, indicating that this is a difficult task** (Nozaki et al., 2022; Zhu et al., 2022; Li and Lin, 2020).

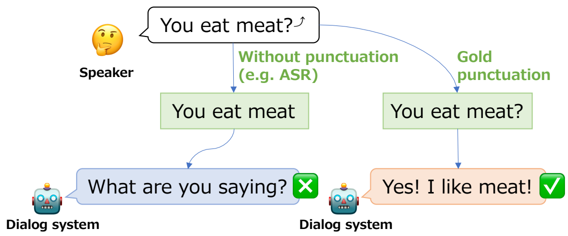

Figure 1: Example of our approach to investigating how a question mark affects the quality of the response generated by a dialogue system.

In spoken dialogue, a rising intonation at the end of a sentence can indicate a question, but when transcribing such a sentence into text, it becomes difficult to distinguish between a declarative and an interrogative sentence. For instance, when asking if someone is going to Torino, it is typically done through inversion, such as “Are you going to Torino?” in languages such as English, it is also possible to ask the question without inversion in the form of a declarative question with a rising intonation at the end, such as “You’re going to Torino?” When transcribing these utterances using ASR, question marks and periods might be missing, which can result in existing dialogue systems designed for text-based dialogues to generating inappropriate responses.

* The author’s affiliation has changed since the writing of this paper.

LREC-COLING 2024, pages 4858–4864

In this study, we investigate how accurate modern dialogue systems, which are trained on written text, in responding to questions uttered without a question mark. Specifically, we examine difference between the response generated by a dialogue system when presented with a speech input lacking a question mark, and that when presented with a correctly punctuated question (Figure 1). Moreover, in an effort to build practical speech dialogue systems, we analyze dialogue examples where question marks are more important based on examples where response generation has failed.

2. Research Methodology

2.1. Research Settings

In this section, we discuss the research methods used to investigate the impact of question marks on dialogue systems. To simplify the problem, we focused only on question marks, although punctuation marks such as commas, exclamation marks, and brackets are also included in text. When dialogue is lengthy, there might be multiple sentences within a single utterance, and the number of punctuation marks such as periods and question marks increases accordingly. However, even in cases where there are multiple sentences in a single utterance, we focus only on the final question mark in the text.

Regarding comparisons with question marks, we use two comparators: no punctuation marks and a closing period. The option without any punctuation marks is intended to simulate a scenario where the text is directly obtained from speech recognition output. The closing period is used to investigate whether there is any negative impact if automatic punctuation fails.

To generate responses, not only the target user utterance but also the past dialogue history is used as input. This is because it can be difficult to generate responses based solely on a brief utterance. In addition, in general spoken chit-chat dialogues, speakers seldom commence with a question, and thus generating a response that includes some context provides a more realistic setting. Therefore, in this study, we used context to generate responses, four historical utterances, including the target user utterance, used as input.

2.2. Dialogue Data

We used Japanese dialogue data because it is often unclear whether or not Japanese utterances are questions without question marks (e.g. [Verb] できます？" (you/I can [Verb]?)), as mentioned in Section 1. Although the data were obtained from Japanese dialogues, we believe that our findings are still useful for the broader research community. Languages other than Japanese, such as Malay and French, also allow for questions to be expressed through intonation, without altering the word order in declarative sentences.

To investigate the effect of question marks in different types of dialogues, three datasets were used: the Nagoya University Conversation Corpus (Fujimura et al., 2012) (NUCC), the dialogue history in the Dialogue System Live Competition 5 Open Track (Higashinaka et al., 2022b) (DSLC), and the JEmpatheticDialogues (JED) dataset (Sugiyama et al., 2021) . These three datasets all includes conversational data resulting from chit-chat dialogues.

NUCC is a collection of transcripts of casual conversations between native Japanese speakers, that have been manually transcribed by humans. The dataset includes conversations involving multiple participants, but we only used conversations between pairs of people.

DSLC is a collection of transcripts of chit-chat conversations between human evaluators and dialogue systems using a Large Language Model equivalent to GPT. Several topics were used for the competition, including anime, tourist spots, and current affairs. The evaluators and the dialogue system conducted verbal conversations, and the evaluator's speech was transcribed using the Google Speech Recognition. There is a possibility that speech recognition text contains errors and fails to include punctuation marks. Thus, we manually added periods and question marks as appropriate at the end of the speech recognition text.

JED is a collection of dialogues between two native Japanese speakers discussing an event while expressing a range of emotions, inspired by the EmpatheticDialogues (Rashkin et al., 2019). This dataset consists of text-based dialogues, rather than spoken dialogues, and was used to compare text-based and spoken dialogues.

As noted earlier, only utterances ending with a question mark were considered in our analysis. While the NUCC and JED datasets involve conversations between humans, for this study, we created research data by considering utterances with question marks as user utterances and the preceding utterances as the dialogue system. Each dataset consisted of 100 dialogue examples. Because there were less than 100 instances (only 66 instances) in the DSLC dataset where the human evaluators' utterances ended with a question mark, we augmented the sample by removing the particle "か (ka)" from utterances ending with interrogative particles such as "ますか (masuka)" or "ですか (desuka)" to provide more examples of utterances ending with a question mark, as shown in the following example:

(1) [Verb] できますか？ → [Verb] できます？

(Can you [Verb]?) → (You can [Verb]?)

2.3. Dialogue Systems

We use four dialogue systems to investigate the effect of question marks. First, we use the Transformer-based dialogue system (Sugiyama et al., 2021) that won the open track of the Dialogue System Live Competition 3 held in 2020 (Higashinaka et al., 2020). The other three dialogue systems were based on GPT (Brown et al., 2020).

The Transformer-based dialogue system pretrains a Transformer (Vaswani et al., 2017) using Twitter reply data, and then fine-tunes it using dialogue data. We used a model fine-tuned on JPersonaChat dialogue data (Sugiyama et al., 2021). The same hyperparameters used in the previous study are used in this study. Hereafter, we refer to this system as “TF”.

The GPT-based dialogue systems used OpenAI’s GPT-3.5 model in addition to two in-house GPT models (Kim et al., 2021), one with approximately 40 billion parameters and the other with approximately 80 billion parameters [1]. The in-house GPT models were trained using Japanese language data. The GPT-3.5 model was accessed via the API provided by OpenAI. These dialogue systems were developed using prompt programming (see Appendix A for an example) and are respectively referred to as the “3.5”, the “40B”, and the “80B” systems.

2.4. Evaluation Method

To investigate the impact of question marks on dialogue systems, we manually evaluated the responses generated by the dialogue systems. The evaluation was conducted by four evaluators (all native Japanese speakers), with each response from a system and dataset being evaluated by three evaluators. The evaluation was based on whether the response answered the question asked, rather than whether it was a good response. In the interests of simplicity, we used a binary evaluation response of either good or bad. The evaluators were given the context within which the dialogue system generated the response, and were asked to evaluate whether the response answered the question. To prevent any biases or preconceptions based on the system or the punctuation mark, the responses were randomly presented to the evaluators in sets of 12 comprising of four systems and three punctuation marks per dialogue example, without any indication of which system or punctuation mark was used to generate the response.

To assess the agreement among the three evaluators, the Fleiss’ Kappa coefficient (Fleiss and Cohen, 1973) was calculated. The agreement exceeded 60% for all data, indicating a high level of agreement (Table 1).

Table 1: Inter-evaluator agreement among the three evaluators.

Data	NUCC	DSLC	JED
Agreement	61.2	82.1	61.3

3. Results

The results of human evaluation of the dialogue systems when changing the sentence-ending punctuation are shown in Table 2. Each response was evaluated by three evaluators, and the percentage of responses that received an evaluation of “good” from at least two of the three evaluators is shown.

Based on these results, it is clear that responses generated with a question mark at the end achieved greater accuracy. Therefore, to generate accurate responses to question utterances, it is necessary to include a question mark at the end of the question.

When comparing the data types, the JED dataset was least affected by question marks, followed by the DSLC and NUCC datasets. The JED dataset includes text-based dialogues, and often uses expressions such as “Would it be” that involve confirmation from the interlocutor. Therefore, it is considered that the impact of question marks on examples from this dataset was relatively small.

Comparing the systems, the “80B” and the “3.5” systems demonstrated better performance, suggesting that larger parameters allow for the generation of more appropriate responses using dialogue context, even without question marks.

Comparing dialogue without punctuation (“X”) and with a closing period (“.”), the results without punctuation were slightly better, but the difference was not significant. Therefore, even if a period is mistakenly added to a question utterance, the same quality of dialogue can be obtained as that using the ASR results.

4. Discussion

In this section, we analyze some actual dialogue examples. Many of the cases where a response can be generated to a question without a question mark are those where the sentence ends with a particle indicating a question (2a) or where there

[1] Due to the API usage fees, not all experiments were conducted with GPT-4. Testing on a portion of the NUCC revealed that GPT-4 performed better than GPT-3.5, however, it was observed that in about 20% of cases, it was unable to provide appropriate responses when there was no question mark.

Table 2: Results. The "?" symbol denotes inputs that end with a question mark, the "." symbols denotes inputs that end with a period, and "X" denotes inputs without any punctuation.

	NUCC			DSLC			JED		
	?	.	X	?	.	X	?	.	X
TF	97	61	65	96	67	72	98	84	89
40B	93	77	75	97	73	76	99	94	95
80B	94	83	82	98	79	82	100	94	98
3.5	92	72	76	98	86	91	99	91	94

Table 3: The accuracy of the comparison based on the presence or absence of interrogative particles.

	w/ interrog. particle			w/o interrog. particle		
	?	.	X	?	.	X
TF	0.97	0.94	0.94	1.0	0.35	0.41
40B	0.97	0.94	0.94	0.97	0.38	0.44
80B	1.0	0.91	1.0	1.0	0.56	0.56
3.5	1.0	0.97	0.97	0.94	0.65	0.76

Table 4: Examples where the "TF" system and the "40B" system failed to generate appropriate responses to utterances without punctuation. The ✓'s signify appropriate responses are determined by human evaluators.

	Input (Context and user utterance)	
Bot	失恋した友達がかわいそうでさ。 (I'm sad for my newly single friend.)	
User	えー、落ち込んでた？ やっぱり。 (Oh, they were feeling down?)	
Bot	落ち込んでたから、気分転換に行った (They was sad, so we went out to cheer up.)	
User	友達は元気になった (Your friend cheers up)	
	Response of dialog system	
TF	それは良かったよ。 (That's good to hear.)	
40B	うん、それは良かった (Yeah, that's good to hear.")	
80B	うん、でもまだ少し暗いみたいだった (Yeah, but they still seemed a little down.)	✓
3.5	うん、元気になってくれたみたい (Yeah, they looks like it cheered up.)	✓

is a question-related word such as one from the "5W1H" (who, what, when, where, why, how) set (2b). Sentences for which it is difficult to generate appropriate responses without a question mark are those that even humans may perceive as declarative sentences without a question mark (2c). In Japanese, the subject is often omitted, making it difficult to distinguish whether a sentence is referring to "me" or "you." This is one reason why it is difficult to generate responses in Japanese.

(2) a. YouTube は見ますか
(Do you watch YouTube)

b. どんな場所です
(What kind of place it is)

c. アメリカに行ったことある
(I/you have been to the USA)

For further analysis, we investigated the impact of interrogative particles. As described in Section 2.2, the DSLC dataset consists of 34 instances of each condition, wherein the contexts are the same, and solely the presence or absence of a terminal interrogative particle differentiates them. Table 3 shows the performance comparison predicated on the impact of interrogative particles. In instances where interrogative particles are present, high performance is maintained irrespective of the punctuation at the sentence's end. Conversely, in the absence of interrogative particles, the performance diminishes unless the sentence concludes with a question mark. From this observation, it can be inferred that in cases where there is no interrogative particle, it becomes challenging to generate appropriate responses without a question mark.
Table 4 shows some examples of where dialogue systems could not provide appropriate responses in cases where there was no punctuation mark (see Appendix B for other examples). As mentioned earlier, it is difficult to determine whether the user's utterance is a question based solely on the utterance itself. In some cases, such as this example, some dialogue systems are able to generate an appropriate response by inferring a question from the context, even without a question mark. However, the systems are unable to do so in approximately 20% of cases. To solve this problem, it is not enough to use text-based information. Rather, it is necessary to explore methods that use non-linguistic information such as sound information or facial expressions, as Alnuhait et al. (2023) has demonstrated.

5. Conclusion and Future Works

In this study, we investigated the impact of a question mark at the end of a sentence on response generation in dialogue systems. The results showed that the presence of a question mark increased the proportion of appropriate responses by 20–30%, indicating the significant role that question marks play in generating appropriate responses in dialogue. Examples of failures when a question mark was absent showed that most utterances were not recognized as questions without a question mark. Thus, the development of multimodal response generation using not only text but also sound and facial expressions is needed to address these issues.

It remains to be investigated whether periods and question marks as well as commas, in multiple sentences affect the quality of responses. Another task for future research is to investigate whether the presence or otherwise of a question mark affects the quality of response generation in other languages.

6. Acknowledgements

We would like to express our gratitude to Kei Uchiumi, Koga Kobayashi, and Masato Umakoshi for their assistance in the evaluation during the preliminary experiment.

7. Bibliographical References

Deema Alnuhait, Qingyang Wu, and Zhou Yu. 2023. FaceChat: An Emotion-Aware Face-to-face Dialogue Framework. arXiv.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. In *Proceedings of Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901.

Joseph L Fleiss and Jacob Cohen. 1973. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and psychological measurement*, 33(3):613–619.

Itsuko Fujimura, Shoju Chiba, and Mieko Ohso. 2012. Lexical and Grammatical Features of Spoken and Written Japanese in Contrast: Exploring a lexical profiling approach to comparing spoken and Written corpora. In *Proceedings of the GSCP International Conference. Speech and Corpora*, pages 393–398.

Ryuichiro Higashinaka, Kotaro Funakoshi, Testuro Takahashi, Michimasa Inaba, Yuiko Tsunomori, Reina Akama, Mayumi Usami, Yoshiko Kawabata, Masahiro Mizukami, Masato Komuro, and Dolça Tellols. 2020. The 3th Dialogue System Live Competition (in Japanese). In *Proceeedings of the Special Interest Group on Spoken Language Understanding and Dialogue Processing*, pages 96–103.

Ryuichiro Higashinaka, Takashi Minato, Kurima Sakai, Tomo Funayama, Hiromitsu Nishizaki, and Takayuki Nagai. 2022a. Dialogue Robot Competition for the Development of an Android Robot With Hospitality. In *Proceedings of IEEE Global Conference on Consumer Electronics*.

Ryuichiro Higashinaka, Tetsuro Takahshi, Sota Horiuchi, Michimasa Inaba, Shiki Sato, Kotaro Funakoshi, Masato Komuro, Hiroyuki Nishikawa, Mayumi Usami, Takashi Minato, Kurima Sakai, and Tomo Funayama. 2022b. The 5th Dialogue System Live Competition (in Japanese). In *Proceeedings of the Special Interest Group on Spoken Language Understanding and Dialogue Processing*, pages 93–100.

Boseop Kim, HyoungSeok Kim, Sang-Woo Lee, Gichang Lee, Donghyun Kwak, Jeon Dong Hyeon, Sunghyun Park, Sungju Kim, Seonhoon Kim, Dongpil Seo, Heungsub Lee, Minyoung Jeong, Sungjae Lee, Minsub Kim, Suk Hyun Ko, Seokhun Kim, Taeyong Park, Jinuk Kim, Soyoung Kang, Na-Hyeon Ryu, Kang Min Yoo, Minsuk Chang, Soobin Suh, Sookyo In, Jinseong Park, Kyungduk Kim, Hiun Kim, Jisu Jeong, Yong Goo Yeo, Donghoon Ham, Dongju Park, Min Young Lee, Jaewook Kang, Inho Kang, Jung-Woo Ha, Woomyoung Park, and Nako Sung. 2021. What Changes Can Large-scale Language Models Bring? Intensive Study on HyperCLOVA: Billions-scale Korean Generative Pretrained Transformers. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 3405–3424.

Xinxing Li and Edward Lin. 2020. A 43 Language Multilingual Punctuation Prediction Neural Network Model. In *Proceedings of Interspeech*, pages 1067–1071.

John Makhoul, Alex Baron, Ivan Bulyko, Long Nguyen, Lance Ramshaw, David Stallard, Richard Schwartz, and Bing Xiang. 2005. The Effects of Speech Recognition and Punctuation on Information Extraction Performance. In *Proceedigs of Interspeech*.

Takashi Minato, Ryuichiro Higashinaka, Kurima Sakai, Tomo Funayama, Hiromitsu Nishizaki, and Takayuki Nagai. 2022. Overview of Dialogue Robot Competition 2022. In *Proceedings of the Dialogue Robot Competition 2022*.

Thai Binh Nguyen, Quang Minh Nguyen, Thi Thu Hien Nguyen, Quoc Truong Do, and Chi Mai Luong. 2020. Improving Vietnamese Named Entity Recognition from Speech Using Word Capitalization and Punctuation Recovery Models. In *Proceedings of Interspeech*, pages 4263–4267.

Jumon Nozaki, Tatsuya Kawahara, Kenkichi Ishizuka, and Taiichi Hashimoto. 2022. End-to-end Speech-to-Punctuated-Text Recognition. In *Proceedings of Interspeech*, pages 1811–1815.

Stephan Peitz, Markus Freitag, Arne Mauser, and Hermann Ney. 2011. Modeling Punctuation Prediction as Machine Translation. In *Proceedigs of International Workshop on Spoken Language Translation*.

Hannah Rashkin, Eric Michael Smith, Margaret Li, and Y-Lan Boureau. 2019. Towards Empathetic Open-domain Conversation Models: a New Benchmark and Dataset. In *Proceedigs of the Annual Meeting of the Association for Computational Linguistics*.

Hiroaki Sugiyama, Masahiro Mizukami, Tsunehiro Arimoto, Hiromi Narimatsu, Yuya Chiba, Hideharu Nakajima, and Toyomi Meguro. 2021. Empirical Analysis of Training Strategies of Transformer-based Japanese Chit-chat Systems.

Máté Ákos Tündik, György Szaszák, Gábor Gosztolya, and András Beke. 2018. User-centric Evaluation of Automatic Punctuation in ASR Closed Captioning. In *Proceedings of Interspeech*, pages 2628–2631.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention Is All You Need. In *Proceedings of Advances in Neural Information Processing Systems*, pages 5998–6008.

Yaoming Zhu, Liwei Wu, Shanbo Cheng, and Mingxuan Wang. 2022. Unified Multimodal Punctuation Restoration Framework for Mixed-Modality Corpus. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 7272–7276.

A. Prompt Examples

See Figure 2.

```
===  # Example of a shot
あなたは A です。A と B が雑談をします。
(You are A. A and B engage in small talk.)
以下は A と B の対話例です。
(Here are some examples of conversations between A and
  B.)

A: 初めまして、A といいます! さて、何のお話をしましょうか。
(A: Nice to meet you, I'm A! So, what shall we talk
  about?)
B: こんにちは! 最近はオリンピックばかり見ていますが、A くんは注
  目している競技はありますか?
(B: Hello! Recently, I've been watching a lot of the
  Olympics. Is there any specific event you are paying
  attention to?)
A: オリンピック面白いですよね。私は今回スケートボードを初めて見
  たのですが、とても激しくて見応えがありました。
(A: Yeah, the Olympics are really exciting! I watched
  skateboarding for the first time this year, and it was
  absolutely thrilling. )
B: スケボーですか! 私も見ましたが、引き込まれましたね。他には、
  サッカーが今面白いですね。
(B: Oh, skateboarding! I watched it too, and it was
  mesmerizing. As for other sports, I've been really
  enjoying soccer.)
A: サッカー今面白いんですか?私は見れてないですね。今どんな感じ
  なんですか?
(A: Is soccer exciting right now? I haven't been able to
  watch it. How is it going?)
B: 先日日本とスペインの試合がありまして、延長戦で点数が決められ
  て日本が負けてしまったんです。惜しかったです。
(B: The other day, there was a match between Japan and
  Spain, and it ended with a score in extra time, and
  Japan ended up losing. It was a close game.)
A: 日本負けてしまったんですね。それは悔しいですね...。スペイン
  はサッカーが強いですから、日本はよく健闘しましたね。
(Japan lost, huh? That's disappointing... Spain is known
  for their strong soccer skills, so Japan really put up
  a good fight.)
###
===
# We repeat shots similar to that shown above.
...
===
まず、A の発話からはじめてください。
(First, let's start with A's utterance.)
```

Figure 2: An example of prompts used in GPT-based dialogue systems. Note that the text following the # symbol is a comment and is not actually included in the prompt.

B. Dialogue Examples

See Tables 5 and 6.

Table 5: Examples where the “TF” system and the “3.5” system failed to generate appropriate responses to utterances without punctuation.

	Input (Context and user utterance)	
Bot	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?)	
Bot	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)	
	Response of dialog system	
TF	組み立て大変ですよね (It’s difficult to assemble, isn’t it)	
40B	はい、組み立てられました (Yes, it has been assembled.)	✔
80B	組み立てられたけど、大変でした (I was able to assemble, but it was difficult.)	✔
3.5	それは良かったですね (Yes, that’s good to hear)	

Table 6: Examples where all systems except the “3.5” system failed to generate appropriate responses to utterances without punctuation.

	Input (Context and user utterance)	
Bot	今日はなんのお話をしましょうか? (What would you like to talk about today?)	
User	ハイオランダって言ったことある? (Have you ever mentioned HiNetherlands)	
Bot	すみません。わからなかったのでもう一度言ってもらえますか? (Sorry, I didn’t understand. Could you please say it again?)	
User	オランダに行ったことある (I/you have ever been to the Netherlands)	
	Response of dialog system	
TF	へー、そうなんですね (Oh, I see.)	
40B	オランダに行ったことあるんですね! (You’ve been to the Netherlands! Great!)	
80B	オランダに行ったことあるんですね! (You’ve been to the Netherlands! Great!)	
3.5	オランダには行ったことがあります。 (I have been to the Netherlands.)	✔