

対話システムは疑問符なしで適切な応答生成できる？ -音声対話を目的とした、疑問文に対する応答生成の調査-

水本智也¹ 山崎天¹ 吉川克正¹ 大萩雅也¹ 川本稔己^{1,2} 佐藤敏紀¹

¹LINE 株式会社 ²東京工業大学

{tomoya.mizumoto, takato.yamazaki, katsumasa.yoshikawa,
masaya.ohagi, toshiki.kawamoto, toshinori.sato}@linecorp.com

概要

人が音声で対話する際、テキストベースの対話とは異なる現象が見られる。テキストベースの場合、疑問文は疑問符“?”を付けることが一般的であるが、音声対話では語尾を上げて疑問を表すことがある。音声認識エンジンには疑問文に対して“?”が付与されないものも多くあり、音声対話システムを作る際に障害となり得る。例えば、音声で「適切な応答生成できる？」と入力された際に、文末の“?”があるかないかで文意は異なり、システムがすべき適切な応答も変わる。本研究では、“?”の有無が対話システムに与える影響を調査し、またどのような発話がより対話に影響するか実例を元に考察する。

1 はじめに

スマートフォンやスマートスピーカーの普及に伴い、音声によってデバイス操作することが一般的になってきた。対話システム研究の分野においても自然言語処理の枠を越えて、ロボットやCGエージェントを使って、ユーザとマルチモーダルな音声対話を行うコンペティションが開催されている [1, 2, 3] が、コンペティションに参加した多くの対話システムは書き言葉コーパスを使って作られている。

音声で対話を行う際は、テキストベースの対話にはない難しさが生じる。その中の一つに音声認識されたテキストに句読点や疑問符(以下、“?”)などの記号(約物)が付与されていない問題がある。そのような問題に対処するために約物の自動付与の研究も行われている [4, 5, 6]。

入力文に約物がないと、入力文が平叙文か疑問文か区別できない問題が生じる。音声対話では語尾を上げて疑問を表現することがあり、テキストにすると疑問とわからないことがある。例えば、NLP2023



図1 疑問文に対する応答の評価の例。疑問に対する応答のため、二番目三番目は不適当な応答となる。

に参加するか聞く際、「NLP2023に参加する？」と語尾を上げて疑問として聞くことができるが、音声認識でテキストに起こされると「NLP2023に参加する」となる。疑問であった情報は落ち、テキストベース対話を想定した既存の対話システムは平叙文と認識し、応答生成に失敗する可能性がある。

そこで本研究では、対話コンペティションでも主流のテキストベース対話を想定したシステムが、音声認識で得られた、疑問符のない疑問発話にどれだけ頑健か調査する。“?”の有無が応答生成に影響しなければ、音声対話システムを作る際に既存の対話システムをそのまま使用できる。一方、“?”が応答生成に影響すれば、入力文に“?”を付与するなど、システムが疑問とわかる工夫をする必要がある。

“?”の影響調査のため、文末が“?”のものとして用意し、それらに対し対話システムで応答を生成し、その応答を手手で評価、分析した(図1)。評価・分析の結果として、“?”が対話システムの応答生成に与える影響は大きいことがわかり、またこの問題の解決のために音声情報などテキスト以外の情報が必要であることが示唆された。

2 調査方法

2.1 対象とする現象と比較方法

テキストに付与される約物は疑問符だけでなく、句読点や感嘆符、鉤括弧など多岐にわたる。本研究

では、問題の簡単化のため“?”だけを対象とする。また、音声対話において、発話が長くなると一発話中に複数文入ることがあり、句点や疑問符が入る箇所もそれに応じて増えるが、発話内に複数文ある場合でもテキストの最後の“?”のみを対象とする。例えば「沖縄好き NLP 行きます」のような発話があった際に、本来は「沖縄好き? NLP 行きます?」のように“好き”の後ろにも“?”が入るのが適切である。今回は「沖縄好き NLP 行きます?」のようにテキスト末尾に入る記号のみを対象とする。

文末の“?”の比較対象として、文末記号を記号なしにした場合と句点“。”にした2種類を用いる。記号なしは音声認識結果のテキストをそのまま使った場合を想定している。句点は、仮に約物の自動付与に失敗した場合に悪影響があるかを調査するために用いる。図1のように各事例に対して応答を生成し、疑問に対する応答であるかを評価する。

応答生成のために検証対象のユーザ発話だけでなく、文脈として過去の発話履歴も入力として使用する。これは、該当する発話が短いなど、一発話のみからの応答生成が難しい場合があるためである。また、一般的な音声対話においてユーザが疑問から始める場合は少なく、ある程度の文脈を使って応答を生成する方が現実的な設定であるため、本研究では文脈も使用して応答生成を行う。本研究では、対象のユーザ発話を含め4発話を入力とする。

2.2 使用する対話データ

特徴の異なるタイプの対話での“?”の影響を調査するため、三つのデータを使用する。一つ目は名大会話コーパス [7]¹⁾ (以下, NUCC), 二つ目は対話システムライブコンペティション5のLINEチームのオープントラック予選の対話履歴 (以下, Open), 三つ目は同じく対話システムライブコンペティション5のLINEチームのシチュエーショントラック予選の対話履歴 (以下, Situation) である。

NUCCは日本語母語話者の雑談対話を人手で書き起こしたものである。データの中には、複数人で対話をしているものもあるが、二人で対話しているもののみを使用する。音声対話したものを人手で書き起こしたデータであるため、音声認識誤りなどは含まれていないが、括弧などで場の雰囲気や笑い声なども情報として付記されている。本研究では、そのような情報は削除して純粋な発話のみを使用する。

1) <https://mmsrv.ninjal.ac.jp/nucc/>

Openは、コンペティション評価者とLINEのオープントラック対話システム [8] との雑談対話の履歴である。雑談のトピックはいくつか用意されており、アニメ、観光地、時事など多岐にわたる。評価者と対話システムは音声でやり取りし、評価者の発話は音声認識エンジンのGoogle Speech Recognitionを使ってテキスト化される。評価者の発話には音声認識誤りが含まれている可能性があり、約物も付与されていない。本研究のために、音声認識されたテキスト末尾に句点、疑問符を人手で付与した。

Situationは、コンペティション評価者とLINEのシチュエーショントラック対話システム [9] との、シチュエーションに沿った対話履歴である。本を失くしたことを謝るシチュエーションであり、また敬称無しで呼び合う仲という設定で比較的親しい間柄に近く、丁寧語(です、ますなど)があまり出ない対話となっている²⁾。Open同様に、音声認識されたテキスト末尾に句点、疑問符を人手で付与した。

2.1節で説明したように、ユーザ発話のテキスト末尾が“?”で終わっているもののみを対象とする。NUCCは全てが人による発話であるが、今回は“?”がある発話をユーザ発話とみなし、その前に発話しているユーザを対話システムとみなすことで調査用データを作成した。全てのデータセット、100対話例ずつ作成した。Openは“?”が含まれる発話が100に満たなかったため、以下の例のように“ますか”や“ですか”など疑問を表す助詞“か”で終わっている発話の“か”を削除することで対話例を増加させた。

- (1) 行きますか? → 行きます?

2.3 対話システム

本研究では、異なる性質を持つ対話システムで疑問符の有無の影響を調査するため、以下の二つのシステムを用いる。一つ目は、2022年に開催されたコンペティションにおいて優勝したHyperCLOVA [10] ベース³⁾の対話システムである。もう一つは、2020年に開催されたライブコンペティション [12] のオープントラックで優勝したTrasformerベースの対話システム [13, 14] を使用する。

HyperCLOVAベースの対話システムは、プロンプトプログラミングによって実現されている。NUCC

2) シチュエーションの詳細は [ホームページ](#) 参照。

3) HyperCLOVAはGPT-3 [11] と同様の性質を持つ日本語基盤モデルを含むシステムである。

表1 4人の評価結果の一致率。

データ	NUCC	Open	Situation
一致率	58.3	72.9	71.0

と Open のデータに対して応答を生成する際は、LINE のオープントラック対話システム [8] で使われた一般的なプロンプトを使用する。Situation のデータに対して応答を生成する際は、LINE のシチュエーショントラック対話システム [9] の謝罪フェーズのプロンプトを使用する。また、HyperCLOVA は 82 Billion パラメータのものを使用する。

Transformer ベースの対話システムは Twitter のリプライデータで Transformer [15] を事前学習しており、そこから対話ログを使って finetune している。本研究では、JPersonaChat [13] で finetune されたモデルを使用する⁴⁾。ハイパーパラメータは、先行研究実験時と同様のものを使用する。HyperCLOVA ベースの対話システムと同様、生成する直前の 4 発話をシステムに入力して応答を生成する。

2.4 評価方法

“?” が対話システムに与える影響調査のため、システムが生成した応答を人手で評価した。評価は著者を含む社内の NLP エンジニア 8 名で行い、NUCC・Open・Situation それぞれ 4 名ずつで評価した。評価に際しては、対話システムが良い応答ができていないかという観点ではなく、疑問文に答えた応答になっているかで評価した。本研究の評価では、計算の簡単化のため“○”か“×”の二択で評価した。評価者には、対話システムが応答生成する際に使った文脈を含めて提示し、疑問文に答えられているかを評価してもらった。評価労力の削減のため、一つの対話事例ごとに二つのシステムと三つの文末記号の組み合わせとなる六つの応答結果をまとめて提示した。システムや文末記号による先入観やバイアスが入ることを防ぐため、どのシステム、どの文末記号によるものかは記載せず、各対話例ごとにランダムに並び替えた。評価ツールにはエクセルを用いて、Appendix の図 2 に示す形で評価した。

4 人の評価結果の一致率を調べるために Fleiss のカッパ係数を計算した。一致率を表 1 に示す。Open と Situation については 70 を越えており高い一致率である。NUCC に関しても Kappa 係数 58 となっており、4 人の結果でも比較的一致している。

4) <https://www.dropbox.com/s/e5ib6rhsbldup3v/japanese-dialog-transformer-1.6B-persona50k.pt>

表2 文末の記号ごとの、対話システムが質問と認識して応答を生成できた割合。“Tr.”は Transformer ベース，“HC”は HyperCLOVA ベースの対話システムを表す。

	NUCC			Open			Situation		
	?	。	なし	?	。	なし	?	。	なし
Tr.	94	59	55	92	60	63	85	47	55
HC	94	72	77	99	72	76	100	90	90

3 実験結果

文末記号を変えた際の対話システムの応答を人手評価した結果を表 2 に示す。それぞれ 4 名で評価しており、その内の 3 名以上が“○”を付けているものの割合を示している。システムやデータに関わらず、文末が“?”のものが高い精度を出しており、疑問発話に対する応答を生成する場合は“?”を付ける必要があることが分かる。

システム間で比較すると Transformer ベースの対話システムは、“?”がある場合とない場合で、およそ 30%の精度差があり“?”の果たす役割が大きいが、それでも大規模なパラメータを持つ汎用モデルによる対話システムにおいても“?”の重要度は高いことがわかる。Situation に対する HyperCLOVA の結果は差が小さいが、これはシチュエーションがかなり限定されており、プロンプトにもそれが反映されているからだと考える。しかし、それでも 10%の差はあり、プロンプトが作り込まれていても“?”は重要であることが示唆される。

記号なしと“。”とを比較するとシステムやデータに関わらず大きな差はない。このことから仮に“?”の自動付与に失敗して疑問文に対して“。”を付与してしまっても、音声認識結果をそのまま使う場合と同等の質の対話は維持できると言える。

4 考察

本節では、実際の対話例を元に分析する。対話システムによって生成された応答例を表 3 に示す。比較を容易にするため、最後のユーザ発話以外の文脈は同じで最後のユーザの疑問発話が助詞“か”で終わるものとそうでない場合での応答結果を比較する。疑問を表す助詞“か”で終わっている場合は、“?”が無くとも疑問系と識別できるため、表中の上の例のように全て適切な応答が生成できている。一

表3 対話システムの実際の応答生成例. システムカラムの“HC”はHyperCLOVA ベース, “Tr.”はTrasformer ベースを表し, “なし”, “。”, “?”は文末記号を表す. 右端に✓があるものは人手評価で適切な応答とされたものである.

入力	システム	生成結果	
Bot: なるほどー。やはり山頂まで登るのは大変でしたか?	HC-なし	登山したことないですね。	✓
User: はい大変でした。	HC-。	富士山に登ったことはないですね。	✓
Bot: 途中で断念される方も多いらしいですけど、無事登頂できたということですね。	HC-?	登山経験はあまりないですね。	✓
User: はいあの登ったことはありますか	Tr.-なし	無いですね。	✓
	Tr.-。	山梨に住んでいた時は登っていました。	✓
	Tr.-?	私は足が短いので登ったことはありません。	✓
Bot: なるほどー。やはり山頂まで登るのは大変でしたか?	HC-なし	すごいですね!登山経験は豊富なんですか?	
User: はい大変でした。	HC-。	富士山登ったことあるんですか!すごいですね!	
Bot: 途中で断念される方も多いらしいですけど、無事登頂できたということですね。	HC-?	登山経験はあまりないですね。	✓
User: はいあの登ったことはあります	Tr.-なし	いいですね。	
	Tr.-。	いいですね。	
	Tr.-?	ないですね。	✓

表4 “?”以外の場合で, HyperCLOVA の対話システムは疑問に答えられたが Transformer では失敗した対話例.

入力		システム応答	
Bot	そんなの聞かなくていいです。	HC-なし	組み立てられたけどまだ完成じゃない ✓
User	じゃあね、あの棚使ってる?	HC-。	組み立てられました。 ✓
Bot	あの棚、使ってますよ。使ってます。	Tr.-なし	組み立て、大変ですよね。
User	組み立てられた	Tr.-。	あらら、大変だったんですね。

方, “あります”のように平叙文の形で文が終わっている場合は, 現状の対話システムでは疑問文と認識して応答を生成できずに, “ユーザが行ったことがある”と捉えて応答生成していることがわかる. 表3の例にも示すように“?”が無くても疑問に対して応答生成できるものの多くは, 先述した例のように疑問を表す助詞で文が終わっているものや (2a), いわゆる 5WH のような疑問詞がある場合 (2b) であった. それに対して, “?”がないと応答生成が難しいものは“?”がないと人間でも平叙文として捉えてしまう可能性があるもの (2c) である.

- (2) a. YouTube は見ますか
- b. どんな場所です
- c. オランダに行ったことある

議論を深めるために, 文末が“?”でない場合に疑問に答えられなかった例を中心に分析する. 表4にHyperCLOVA ベースでは疑問に答えた応答を生成できたが, Transformer ベースでは失敗した例を示す. 発話単体では疑問と判断するのは難しいが, HyperCLOVA ベースのシステムは文脈から疑問であ

ると判断できたと考える. この例のように文脈から疑問として判断し応答生成できる例もあるが, “?”がない場合に失敗しているものは20%程度ある. 参考に Appendix の表5に“?”がない場合に疑問に答えた応答が生成できない他の発話例を示す. これらの対話例は, 疑問符の自動付与もテキストのみから行うのは難しい. また, 日本語対象ではないが, 音声入力への約物の自動付与の先行研究でも, その精度は8割程度である [4, 5, 6]. この問題を解決していくには, 言語情報だけで応答生成することは難しく, テキスト処理の枠を越えて, 音声や顔の表情も同時に使った応答生成をする必要があると考える.

5 おわりに

本研究では, 対話システムの応答生成において, 文末の“?”が生成結果に与える影響について調査した. その結果, “?”がある場合とない場合を比較すると, “?”がある場合に適当な応答を返す割合が20~30%ほど高く, 対話の応答生成において“?”が果たす役割が大きいことがわかった.

疑問符がない場合に失敗する例を見ると, 疑問符がないと疑問とわからないものがほとんどであった. 文脈を含めてもテキストのみで疑問と判断するのが難しいものもあり, 今回は対象としていないが音声対話には音声認識誤りの問題もある. これらの問題を解決するためには, テキスト処理を越えて音声や顔の表情を使った応答生成をする必要がある.

複数文ある際の句点や疑問符, また読点の有無が応答の質に影響を与えるかは課題として残っている. また, 日本語以外での言語で疑問符の有無が応答生成に影響するかの検証も今後の課題としたい.

謝辞

本研究の評価を手伝ってくださった LINE NLP チームの内海慶氏, 小林滉河氏, 馬越雅人氏に感謝する。

参考文献

- [1] Ryuichiro Higashinaka, Takashi Minato, Kurima Sakai, Tomo Funayama, Hiromitsu Nishizaki, and Takayuki Nagai. Dialogue robot competition for the development of an android robot with hospitality. In **Proc. of 2022 IEEE 11th Global Conference on Consumer Electronics**, 2022.
- [2] Takashi Minato, Ryuichiro Higashinaka, Kurima Sakai, Tomo Funayama, Hiromitsu Nishizaki, and Takayuki Nagai. Overview of Dialogue Robot Competition 2022. In **Proc. of the Dialogue Robot Competition 2022**, 2022.
- [3] 東中竜一郎, 高橋哲朗, 堀内颯太, 稲葉通将, 佐藤志貴, 船越孝太郎, 小室允人, 西川寛之, 宇佐美まゆみ, 港隆史, 境くりま, 船山智. 対話システムライブコンペティション 5. 第 96 回人工知能学会 言語・音声理解と対話処理研究会, pp. 93–100, 2022.
- [4] Jumon Nozaki, Tatsuya Kawahara, Kenkichi Ishizuka, and Taiichi Hashimoto. End-to-end Speech-to-Punctuated-Text Recognition. In **Proc. of Interspeech**, pp. 1811–1815, 2022.
- [5] Yaoming Zhu, Liwei Wu, Shanbo Cheng, and Mingxuan Wang. Unified Multimodal Punctuation Restoration Framework for Mixed-Modality Corpus. In **Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing**, pp. 7272–7276, 2022.
- [6] Xinxing Li and Edward Lin. A 43 Language Multilingual Punctuation Prediction Neural Network Model. In **Proc. of Interspeech**, pp. 1067–1071, 2020.
- [7] Itsuko Fujimura, Shoji Chiba, and Mieko Ohso. Lexical and Grammatical Features of Spoken and Written Japanese in Contrast: Exploring a lexical profiling approach to comparing spoken and Written corpora. In **Proc. of the VIIth GSCP International Conference. Speech and Corpora**, pp. 393–398, 2012.
- [8] 山崎天, 川本稔己, 大萩雅也, 水本智也, 小林滉河, 吉川克正, 佐藤敏紀. HyperCLOVA を用いた音声雑談対話システム. 第 96 回人工知能学会 言語・音声理解と対話処理研究会, pp. 113–118, 2022.
- [9] 吉川克正, 川本稔己, 山崎天, 水本智也, 小林滉河, 大萩雅也, 佐藤敏紀. シチュエーションに合わせたシナリオ誘導と HyperCLOVA を利用した応答生成によるハイブリッド対話システム. 第 96 回人工知能学会 言語・音声理解と対話処理研究会, pp. 124–129, 2022.
- [10] 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. What Changes Can Large-scale Language Models Bring? Intensive Study on HyperCLOVA: Billions-scale Korean Generative Pretrained Transformers. In **Proc. of the 2021 Conference on Empirical Methods in Natural Language Processing**, pp. 3405–3424, 2021.
- [11] Tom B Brown, Benjamin Mann, et al. Language Models are Few-Shot Learners. In **Proc. of Advances in Neural Information Processing Systems**, pp. 1877–1901, 2020.
- [12] 東中竜一郎, 船越孝太郎, 高橋哲朗, 稲葉通将, 角森唯子, 赤間怜奈, 宇佐美まゆみ, 川端良子, 水上雅博, 小室允人, Dolça Tellols. 対話システムライブコンペティション 3. 第 90 回人工知能学会 言語・音声理解と対話処理研究会, pp. 96–103, 2020.
- [13] Hiroaki Sugiyama, Masahiro Mizukami, Tsunehiro Arimoto, Hiromi Narimatsu, Yuya Chiba, Hideharu Nakajima, and Toyomi Meguro. Empirical analysis of training strategies of transformer-based japanese chat systems, 2021.
- [14] 杉山弘晃, 成松宏美, 水上雅博, 有本庸浩, 千葉祐弥, 目黒豊美, 中嶋秀治. Transformer encoder-decoder モデルによる趣味雑談システムの構築. 第 90 回人工知能学会 言語・音声理解と対話処理研究会, pp. 104–109, 2020.
- [15] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention Is All You Need. In **Proc. of Advances in Neural Information Processing Systems**, pp. 5998–6008, 2017.

A 参考情報

	A	B	C	D
1	ID	文脈	ロボット応答	評価
2	6856		精神年齢ですよ。まだ10代だからかなあ。	
3	0a09		人生やがな。	
4	5bb4	Bot: でもまあおっさんだけど幼いやろうねえ、まだ。なんかそんな気がする。 User: うーん、そうかもね。	わたし12月生まれなんで、まだ20代前半です。	
5	89bff	Bot: まだまだですわ。 User: 何がまだまだなんですか？	もっと頑張らなあかんやんけ。	
6	5987		幼さはまだまだです。あなたはお仕事は何を？	
7	c99e		だって俺らもう40歳やで。	

図2 評価に使ったエクセルのサンプル。

表5 HyperCLOVA ベースの対話システムで、文末“?”の場合は適切な応答ができたもので、“。”と記号なしの場合両方で応答に失敗したユーザ発話の例。

NUCC	Open	Situation
適当なテキスト見つかったなんか検索した	富士山には登ったことがあります	私は電話してないあなたかけてみた
新しい先生そしたらA先生って知ってる	ところでチェコには行ったことがあります	よかった専門書もその中に入ってる
でもあの売上に協力しますそれとフッ素の入った歯磨きある	富士山以外です	じゃあ大丈夫静か一人で行く
あほんとじゃ何豚はだめってこと	攻殻機動隊について知っています	まだ忘れ物とか届いてないって
中学のときも太っと思っただ	美味しい日本のカレー屋さん知ってる	いよいよあの鉄道に落とし物の届けをしたの
うんチョコレート食べるー	ところで日本のカレーって好きです	自殺には連絡した
じゃあねあの棚使ってる	嫌いなものは特別嫌いっていうのはないですあります	-
ほんでやねターキーがどこにあんねん	ギターは持っています	-
コロッケ使ったことない	カレーは好きです	-
うんやってもらったでも自分でやっとな子っっていなかった	そうですね四つ葉さんはズームで飲み会とかします	-