

# 間接的な応答と直接的な応答の対からなる 対話コーパスの構築

2021/07/28, 第249回自然言語処理研究会

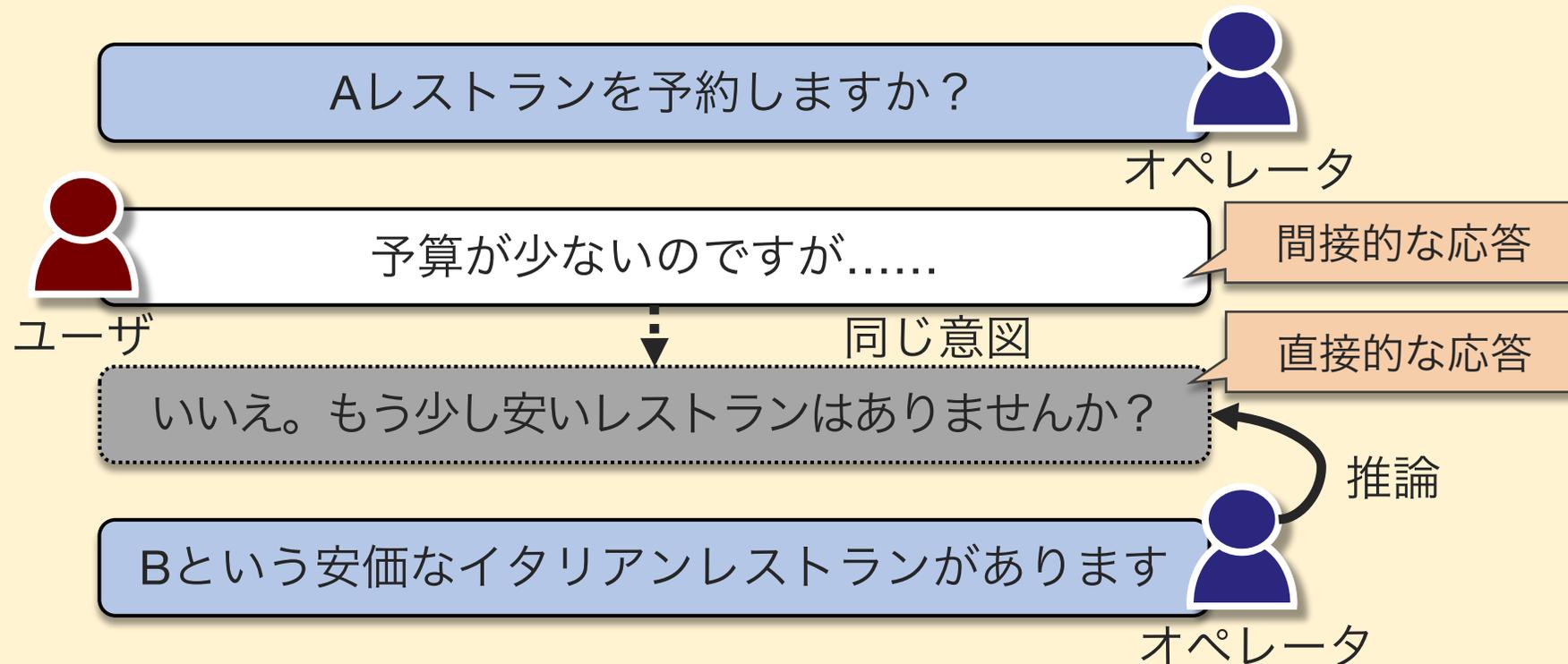
高山 隼矢<sup>†</sup>, 梶原 智之<sup>‡</sup>, 荒瀬 由紀<sup>†</sup>

<sup>†</sup>大阪大学大学院情報科学研究科

<sup>‡</sup>愛媛大学大学院理工学研究科

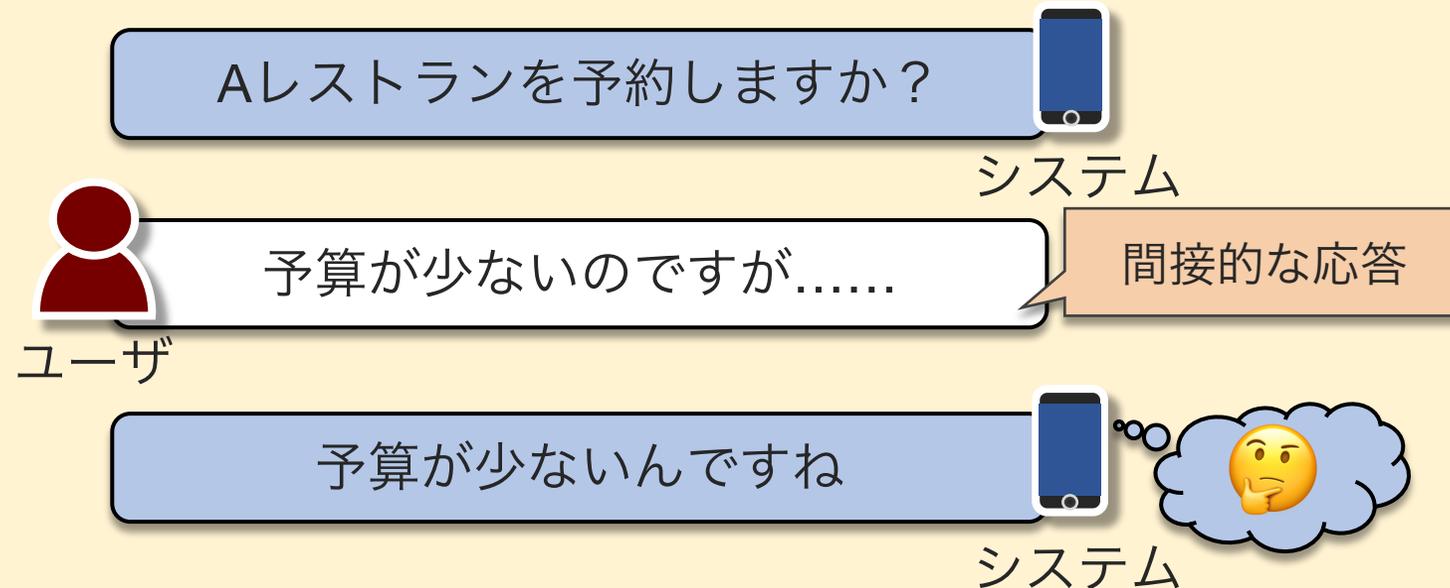
# 対話における間接的な応答

- 人間は自分の意図や相手への要求を間接的に表現することがある
- 人間は対話の履歴等を基に**同じ意図を表現するより直接的な応答を推論する**



# 課題

- 対話システムも間接的な応答の意図を正確に推定できなければ不適切な出力を返す可能性がある
- **間接的な応答と直接的な応答の関係性を取り扱った実用的なコーパスは存在しない**



# 本研究の目的と貢献

**[目的]** 間接的な応答を理解する対話システムの開発

## **[貢献]**

1. 間接的・直接的応答の対からなる**7万件規模**の対話コーパス **DIRECT** (Direct and Indirect **RE**sponses in **C**onversational **T**exts) を構築
2. 間接的応答から直接的応答への言い換えモデルを構築し、ユーザ発話を事前により直接的な発話に言い換えることで、**対話応答生成の性能を向上**

# DIRECTコーパスの構築と分析

# パラレルコーパスの構築

実際の対話に近いコーパスを構築するために、  
既存のタスク指向対話コーパスである MultiWoZ 2.1 を拡張する

## MultiWoZ 2.1

- ユーザとオペレータが交互に発話を行う形式の対話データ
- ユーザはホテルの予約やタクシーの手配など、達成したい目的を持って発話を行う
- オペレータはユーザとの対話を通してユーザが達成したい目的を代わりに達成する
- 対話行為や対話状態など豊富なアノテーションが付属

# クラウドソーシングを用いた応答対の収集

MultiWoZ コーパスにおける各ユーザ応答に対し，その言い換えとなる直接的な応答・間接的な応答を Amazon Mechanical Turk を用いて収集

The diagram illustrates the process of collecting response pairs for the MultiWoZ corpus. On the left, a vertical flowchart shows the sequence of information provided to workers: '過去の対話履歴' (Past dialogue history), '言い換え対象の応答' (Response to be paraphrased), and '未来の対話履歴' (Future dialogue history). Below this, two boxes indicate the input forms for '直接的な応答' (Direct paraphrase) and '間接的な応答' (Indirect paraphrase). On the right, a screenshot of the Amazon Mechanical Turk interface shows the 'Dialogue Context' section with a conversation between a user and an operator. The user's response 'I'm a vegetarian' is highlighted in red. Below the context, the 'Your Answer' section shows two input fields: 'Type-1 (Direct paraphrase):' with the text 'Yes, I need to find a vegetarian restaurant' and 'Type-2 (Indirect paraphrase):' with the text 'I do not eat meat or fish.' A 'Submit' button is visible at the bottom of the interface.

**過去の対話履歴**

**言い換え対象の応答**

**未来の対話履歴**

**直接的な応答の入力フォーム**

**間接的な応答の入力フォーム**

**Dialogue Context**

**USER:** I would like to have dinner in Cambridge  
**OPERATOR:** Do you have a preference for restaurants?

**USER(TGT):** I'm a vegetarian

**OPERATOR:** OK, there are one vegetarian restaurant near the hotel.  
Would you like to book?

**USER:** Yes, please.

**Your Answer**

Type-1 (**Direct** paraphrase):

Yes, I need to find a vegetarian restaurant

Type-2 (**Indirect** paraphrase):

I do not eat meat or fish.

Submit

# 収集結果・データ例

合計 71,498 件の言い換え対を得た。データ例を以下に示す

発話者	発話
	(省略)
OPERATOR	Would you like to pick a different type of food?
USER	I need a place to stay in the north
OPERATOR	What price range are you comfortable with?
USER (言い換え対象)	Something in the moderate price range would be good.
<b>USER (間接的な応答)</b>	<b>I don't want to overspend but remember its also vacation</b>
<b>USER (直接的な応答)</b>	<b>Can you choose something that is not too expensive and not to cheap.</b>
OPERATOR	Do you have a preference as to what area of town you dine in?

間接的な応答は単に自己開示的な内容だが、  
直接的な応答は価格帯に対して具体的に言及している

# コーパスの人手評価

コーパスの品質評価のため、7,372 件のテストデータについて人手評価を実施  
応答対を、どちらが直接的な応答かわからないように Response A, B と  
して対話履歴と共にアノテータに提示し、以下の観点で評価してもらう

**Intention-accuracy:** それぞれの発話が元の発話と同じ意図のものかどうか  
Yes/No の二値で判定

**Directness-accuracy:** どちらがより直接的か  
"Response A", "Neutral", "Response B" の三値で判定

尺度	Accuracy [%]
Intention (間接的応答)	94.99
Intention (直接的応答)	99.72
Directness	81.44

# コーパスの統計的分析

間接的・直接的応答それぞれについて、語彙サイズと平均文長を調査

尺度	# of words
語彙サイズ (間接的応答)	6,273
語彙サイズ (直接的応答)	4,664
文長 (間接的応答)	15.59
文長 (直接的応答)	12.38

- 語彙サイズは間接的応答の方が大きい  
→ 同じ意図であっても間接的応答の方がフレーズの多様性が大きい
- 文長は間接的応答の方が大きい (有意水準 0.1% で有意差あり)  
→ 平均的には間接的応答の方が冗長である

# 語彙に関する分析

間接的/直接的それぞれの応答について  
trigram の出現頻度 Top-20 を比較

## 直接的応答

"book" や "find" など**要求を直接表現する  
単語**や "the reference number" のように  
**言及対象を具体的に示すフレーズ**が目立つ

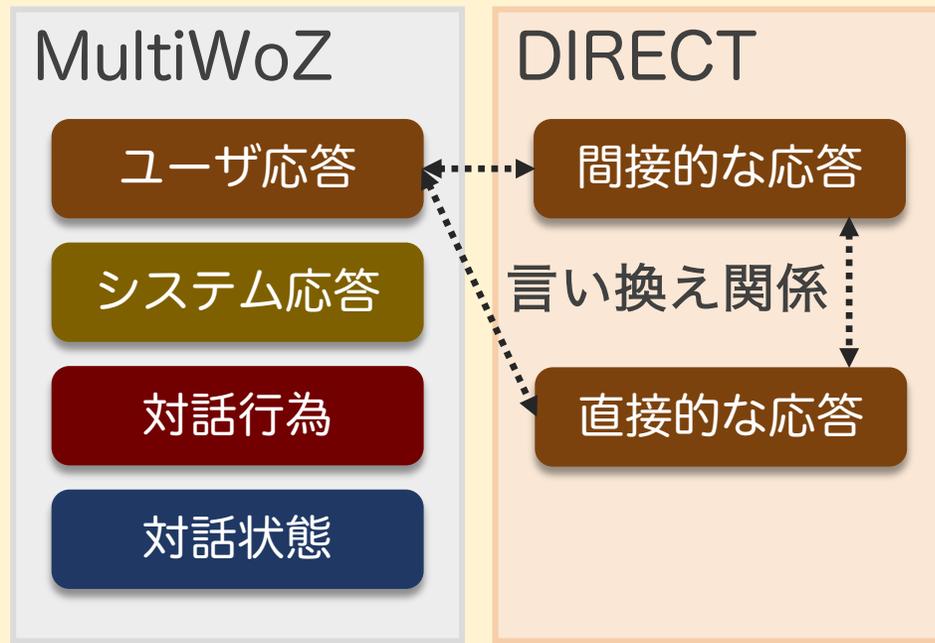
## 間接的応答

"i think that" や "is there any" など、  
直接的応答には出現しないようなフレーズ  
が含まれる

Indirect responses		Direct responses	
trigram	freq	trigram	freq
i want to	2387	<b>find me a</b>	2617
i need to	2223	i want to	2442
would like to	1969	<b>can you find</b>	1971
i would like	1903	<b>please find me</b>	1924
is there any	1762	all i needed	1807
that would be	1588	thanks for the	1643
you help me	1584	in the centre	1628
thanks a lot	1407	i need to	1623
in the centre	1402	for the help	1539
i think that	1312	<b>you find me</b>	1531
i think i	1270	that's all i	1403
a place to	1246	can you get	1376
you have been	1242	give me the	1358
would be swell	1056	i need a	1206
such a great	1042	you get me	1200
i think you	1027	i would like	1145
you have done	1011	please give me	1097
a great help	981	get me a	1094
have been such	979	<b>book it for</b>	1086
been such a	979	<b>the reference number</b>	1060

# DIRECTコーパスの特徴まとめ

言い換え可能な間接的・直接的応答の対からなる7万件規模の対話コーパス  
<https://github.com/junya-takayama/DIRECT> にて公開中



## 複数の応用タスク

- 直接的な応答の言い換え (後述)
- 間接的な応答の言い換え
- 発話の直接性推定

MultiWoZ との組み合わせで、  
既存の対話タスクへの適用も可能

- 対話応答生成 (後述)
- 対話行為推定
- 対話状態追跡

# 間接的な応答を理解する 対話応答生成モデルの構築

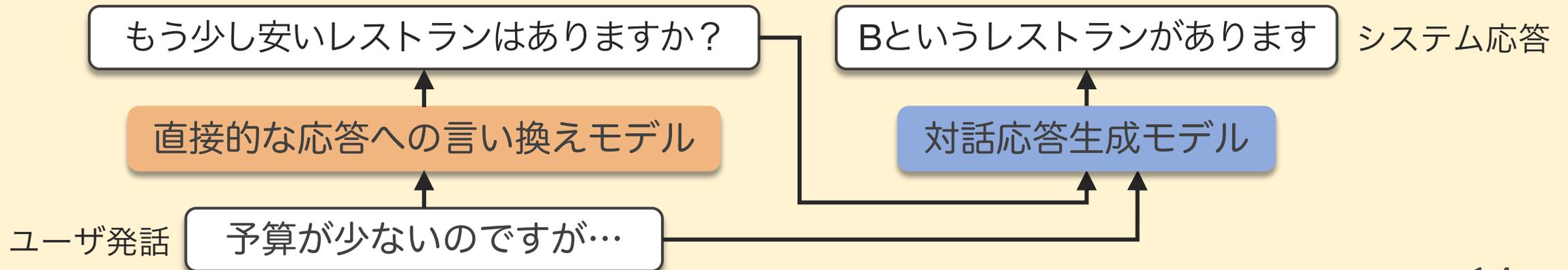
# 対話応答生成への適用

## [仮定]

対話システムにとって直接的なユーザ発話の方が間接的な発話より正しいシステム応答を生成しやすい

## [実験概要]

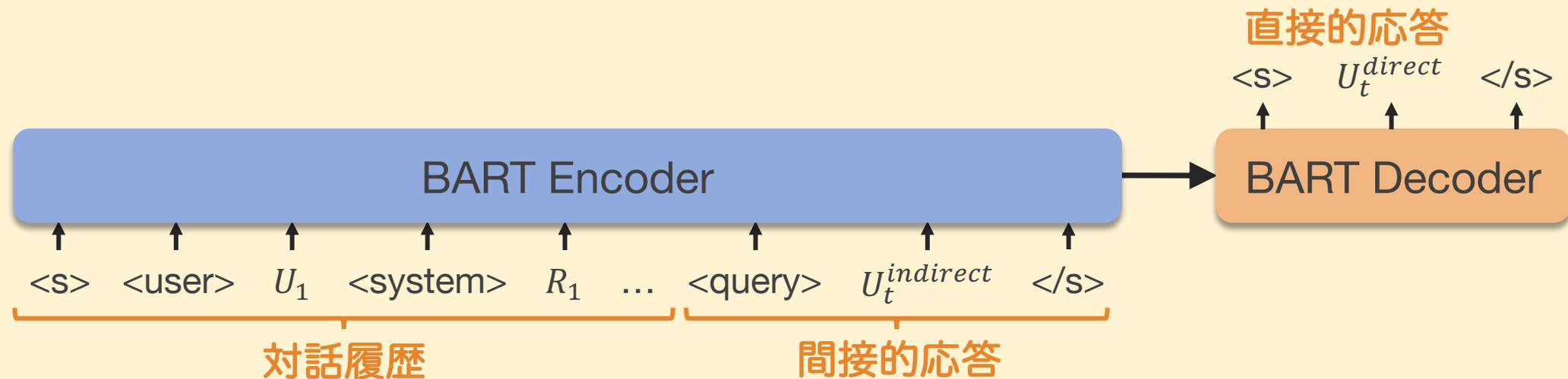
既存の対話応答生成モデルに対し，入力発話を事前に直接的な発話に言い換えて入力することで，応答生成の性能が向上するか確認する



# 間接的応答から直接的な応答への変換

事前学習済み文生成モデル (BART) を用いて、応答をより直接的に言い換え

- 対話履歴と間接的応答を入力して、直接的応答を生成
- 対話履歴の有無による影響も調査
- また、事前学習なしのベースライン (Transformer) も実験



# 変換例

USER	Hi, I'm looking for a place to eat some indian food.
SYSTEM	Do you have a price range in mind?
Indirect response	I want it to be the best place in town.
BART	I want it to be an expensive place.
- w/o history	Can you find me something expensive?
transformer	Can you find me an attraction in town?
reference	I was hoping for a much more expensive place if possible.

- 文脈を利用したモデル (BART) では "best place" から "expensive place" というフレーズを生成できている
- 転移学習でないモデル (transformer) は全く関係のない "attraction" というフレーズを生成してしまっている

# 直接的な応答への変換の実験結果

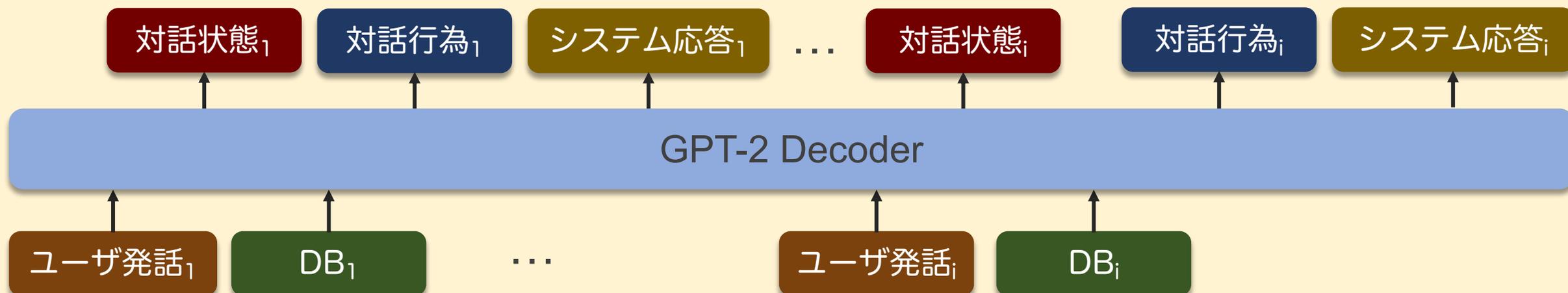
モデル	BLEU	Perplexity
Transformer w/ 対話履歴	25.23	2.66
BART w/o 対話履歴	32.51	<b>2.15</b>
BART w/ 対話履歴	<b>33.77</b>	2.16

- 履歴がある方（BART w/ 対話履歴）の方が BLEU が高い  
→履歴を利用することが意図推定に有効
- 事前訓練あり（BART）の方がなし（Transformer）より推定性能が高い  
→事前訓練によって得られる知識は間接的な応答の言い換えにも効果的

# 対話応答生成モデル UBAR [1]

事前学習済みの文生成モデルである GPT-2 を基に構築された  
タスク指向対話応答生成モデル。以下を繰り返すことで対話を行う

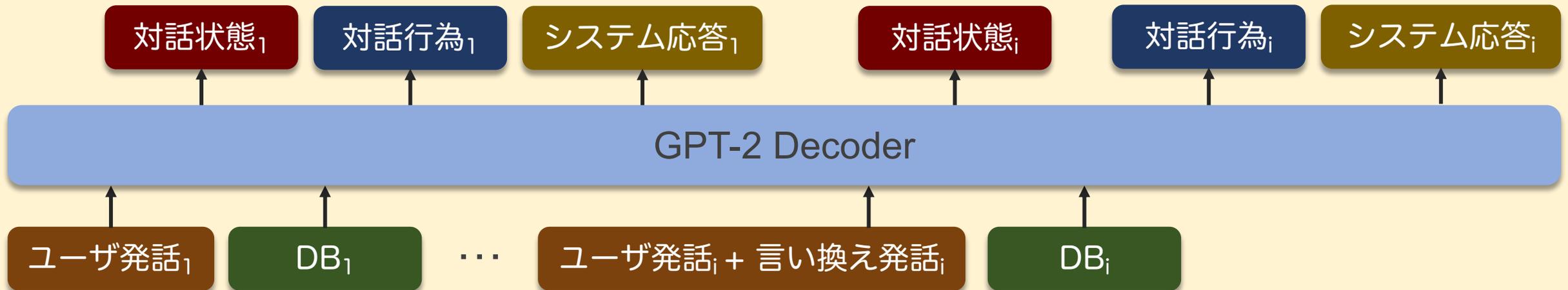
1. 入力されたユーザ発話に基づき現状の対話状態を出力
2. 対話状態を基に DB からホテルやレストランを検索し、モデルに入力
3. 検索結果に基づきシステム応答に含めるべき情報（対話行為）を出力
4. システム応答を生成



[1] Y. Yang et al., 2021. "UBAR: Towards fully end-to-end task-oriented dialog systems with gpt-2." In *Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI)*.

# 直接的なユーザ発話を考慮した応答生成

- 予測対象の応答の直前のユーザ発話を言い換えモデルを用いてより直接的な応答に言い換え
- 元のユーザ発話と言い換えた発話を特殊トークンを挟んで連結し、UBAR に入力



[1] Y. Yang et al., 2021. "UBAR: Towards fully end-to-end task-oriented dialog systems with gpt-2." In *Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI)*.

# 実験設定

[学習データ] MultiWoZ コーパス

[比較モデル]

UBAR : UBAR (既存手法) をそのまま利用

+ 直接的な発話 (生成) : 言い換えモデルが生成した直接的な応答を入力

+ 直接的な発話 (参照) : DIRECT コーパスに収録された直接的な応答 (参照文) を入力

[評価尺度]

BLEU : 生成応答と参照応答との間の n-gram ベースでの類似度

INFORM : 提示したエンティティ (ホテル名など) が正解と一致した割合

SUCCESS : 属性値 (住所や予約番号など) を正確に出力できた割合

COMBINED = BLEU + 0.5 · (INFORM + SUCCESS) : 3つの尺度を統合したもの

ハイパーパラメータ等は UBAR [1] に準拠

[1] Y. Yang et al., 2021. "UBAR: Towards fully end-to-end task-oriented dialog systems with gpt-2." In *Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI)*.

# 実験結果

モデル	BLEU	INFORM	SUCCESS	COMBINED
UBAR	15.07	90.6	77.8	99.27
+ 直接的な発話 (生成)	<b>15.39</b>	91.1	78.8	100.34
+ 直接的な発話 (参照)	15.27	<b>91.7</b>	<b>79.4</b>	<b>100.82</b>

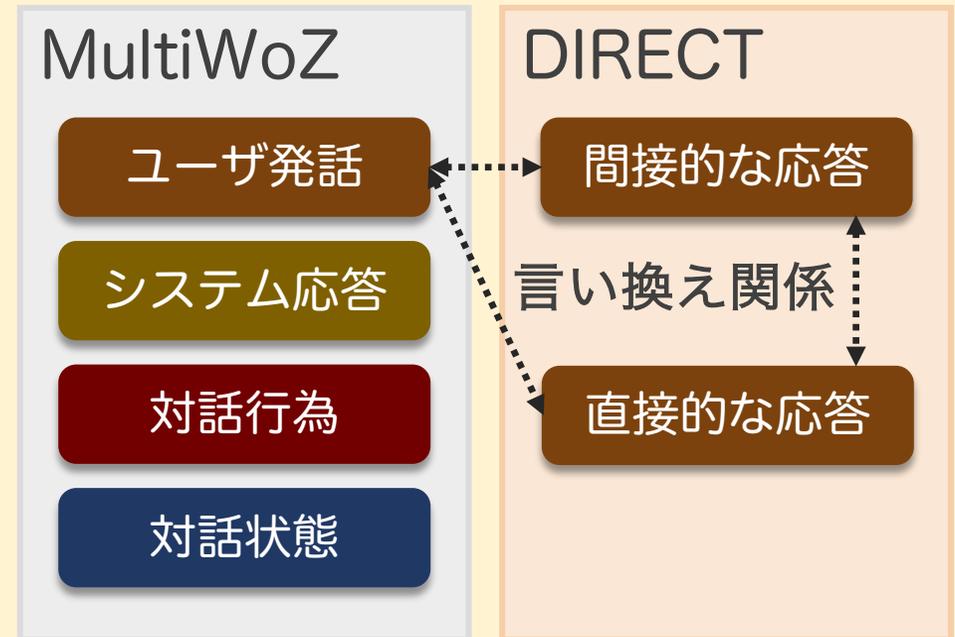
- 全ての尺度において、直接的に言い換えた発話を考慮したモデルが UBAR のスコアを上回っている  
→ ユーザ発話のより直接的な発話への言い換えは応答生成の性能向上に寄与
- 参照文を利用した方が BLEU 以外の全ての尺度でスコアが高い  
→ より正確な言い換えモデルを構築することでさらに応答生成性能が向上

# まとめ

- 間接的・直接的応答の対を含む**7万件規模**の対話コーパス **DIRECT** を構築
- 間接的応答から直接的応答への言い換えモデルを対話応答生成タスクに適用し、**応答の質を向上**

## [今後の課題]

- より高精度な言い換えモデルの構築
- 対話状態追跡タスク等の実験



DIRECT コーパス : <https://github.com/junya-takayama/DIRECT>