

# 国際会議参加報告

# SIGDIAL (+YRRSDS) 2024

---

斉志揚 田中 義規

電気通信大学



Inaba Lab.  
The University of Electro-Communications



The University of  
Electro-Communications

# The 25<sup>th</sup> Annual Meeting of the Special Interest Group on Discourse and Dialogue

September 18, 2024 - September 20, 2024



**SIGdial**

# SIGDIAL

## ■ 談話と対話に関する国際会議

- 談話処理
- 対話システム
- コーパス・ツール・方法論
- 語用論・意味論的モデリング

## ■ SIGDIAL 2024 の基本情報

- 開催場所：京都大学 国際科学イノベーション棟5F
- 登録者数：161
- 発表件数：65



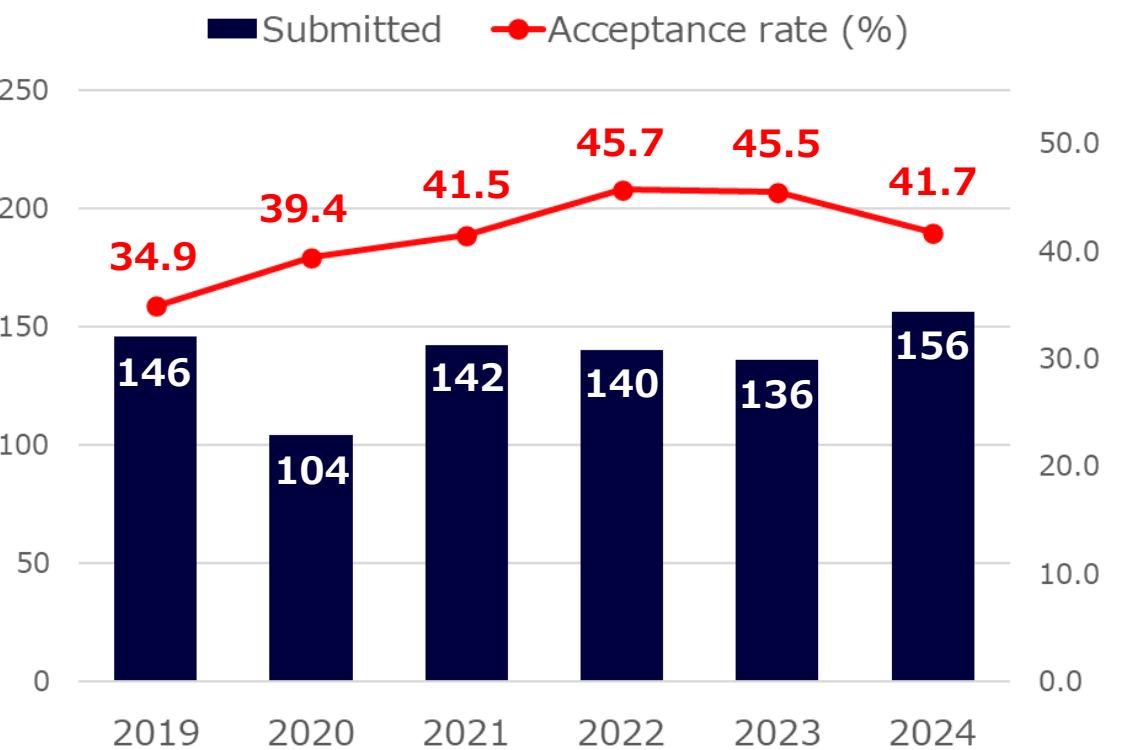
# 統計

## ■ 2024の投稿数・採択率

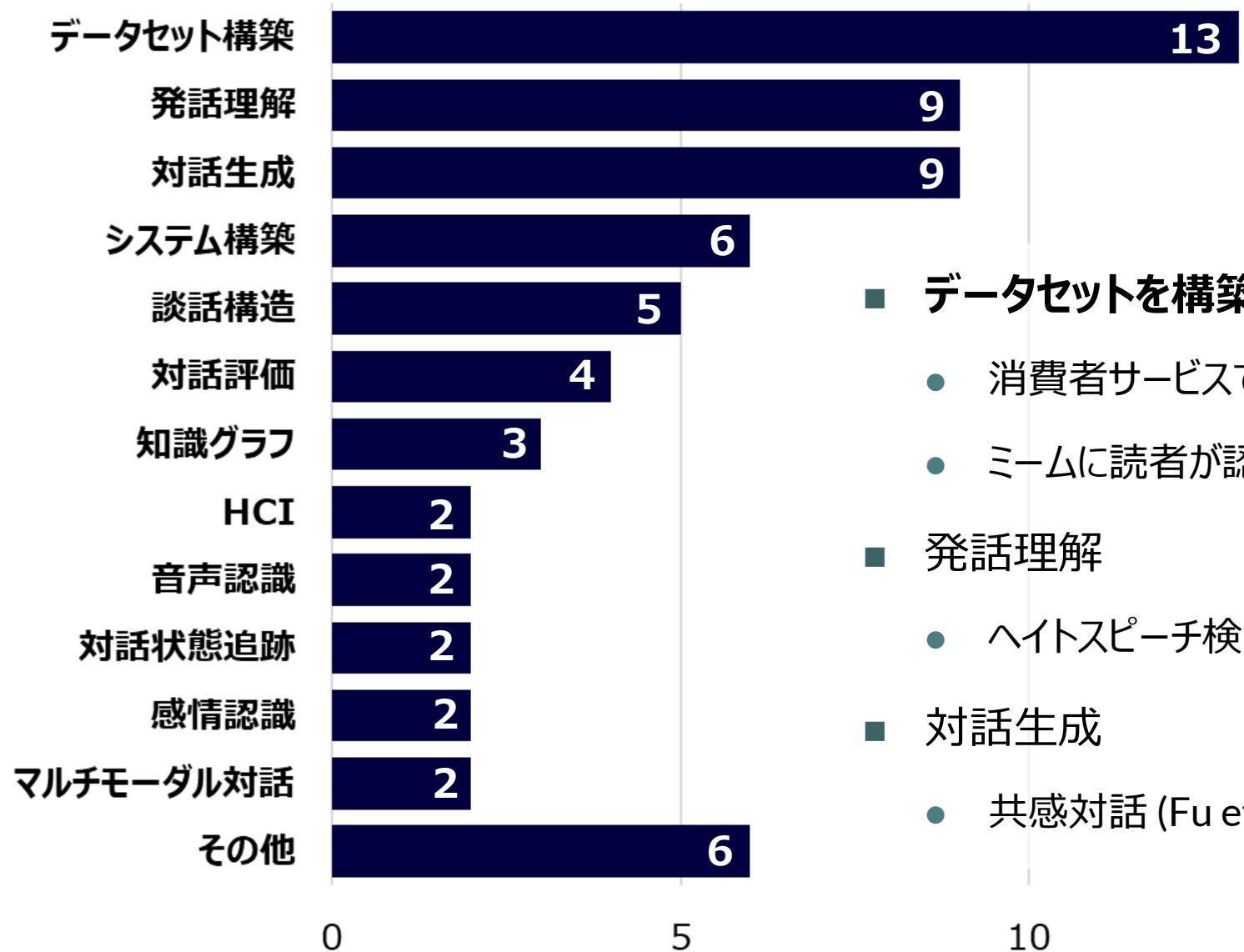
	投稿数	採択数 (率)
Long	106	44 (41.5%)
Short	44	17 (38.6%)
Demo	6	4 (66.7%)
Total	156	65 (41.7%)

- 過去2年と比較して採択数が減少
- 投稿数は過去最高

## 今年 + 過去5年の推移



# 発表内容の傾向



## ■ データセットを構築した研究が多数

- 消費者サービスでの「曖昧な参照」を含む対話を収集. (Li et al.)
- ミームに読者が認識する意図をアノテーション (Park et al.)

## ■ 発話理解

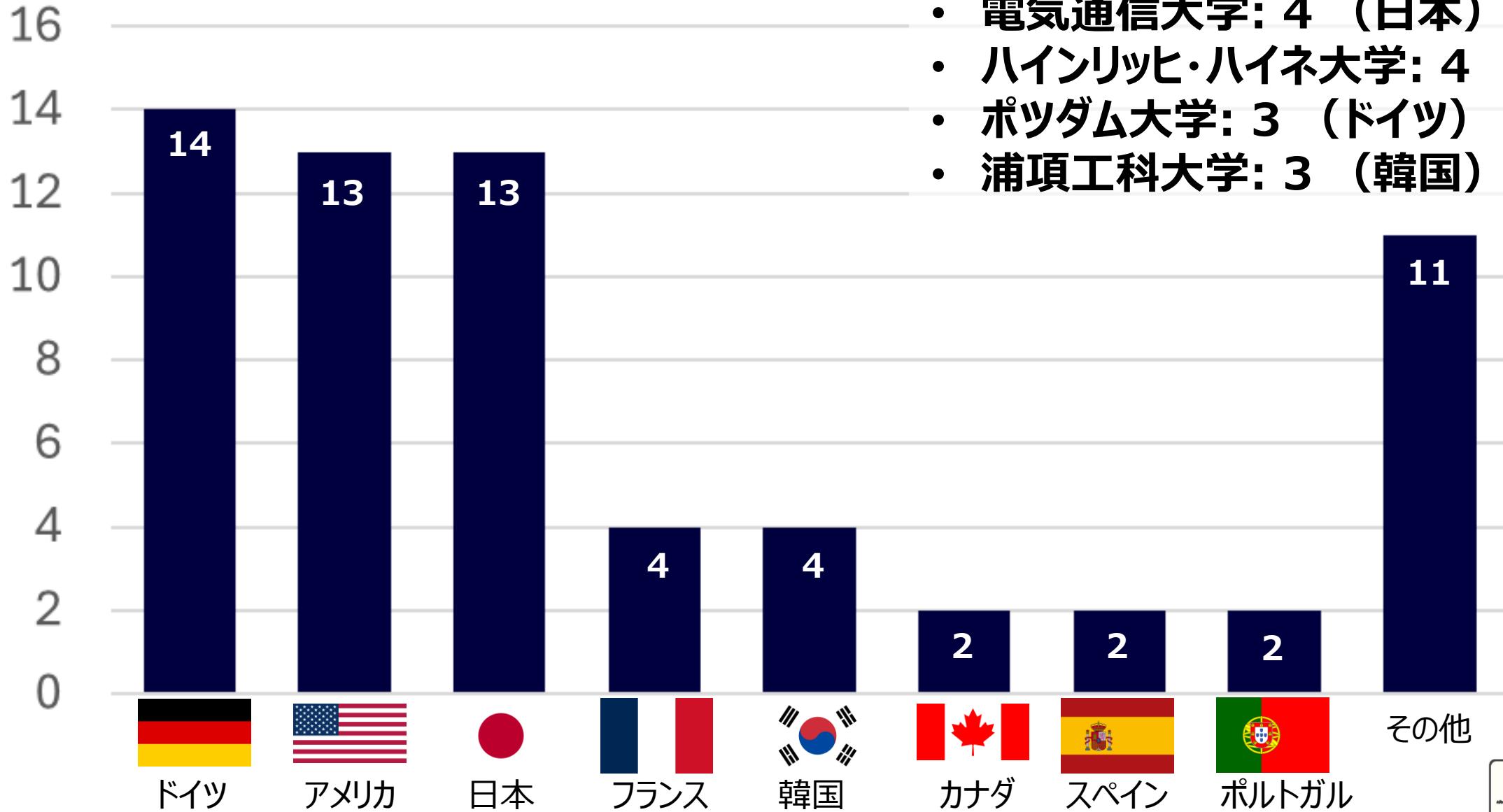
- ヘイトスピーチ検出 (An et al.), 意図分類 (Hong et al.)

## ■ 対話生成

- 共感対話 (Fu et al.), 対話データの拡張 (Qi et al., )



# 発表者 (First author) の属性

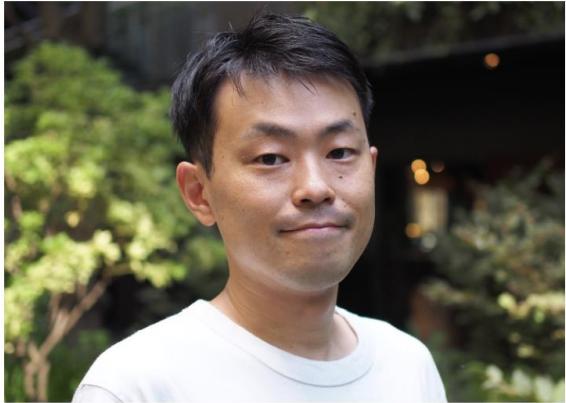


## 上位4機関

- 電気通信大学: 4 (日本)
- ハインリッヒ・ハイネ大学: 4 (ドイツ)
- ポツダム大学: 3 (ドイツ)
- 浦項工科大学: 3 (韓国)



# Keynotes



- **Yeah, Well, Haha: Generating Non-linguistic Behaviors For Human-like Conversational Robots**

- **Koji Inoue**, Kyoto University
- マルチモーダルLLMの進化とともに、人間らしい対話ロボットを実現するための非言語的行動（あいづち、笑い、ターンテイキング）研究の進展と課題



- **AI Agents Beyond ChatGPT**

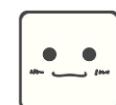
- **Zhou Yu**, Columbia University
- LLMではなく、小型で特化型のAIモデルを組み合わせてタスクを解決する方法や、それを支える技術や仕組み

# ベストペーパー候補



Best Paper

- **InteLLA: Intelligent Language Learning Assistant for Assessing Language Proficiency through Interviews and Roleplays**
  - インタビューとロールプレイを通じて、第二言語学習者の口頭能力を評価するためのマルチモーダル対話システムを提案
- **Dialogue Discourse Parsing as Generation: A Sequence-to-Sequence LLM-based Approach**
  - 談話解析タスクをSeq2SeqのLLMを用いた生成タスクとして扱いモデルを構築
- **Enhancing Dialogue Speech Recognition with Robust Contextual Awareness via Noise Representation Learning**
  - コンテキストに含まれるノイズにロバストな対話音声認識の手法を提案
- **Local Topology Measures of Contextual Language Model Latent Spaces With Applications to Dialogue Term Extraction**
  - 文脈言語モデルの潜在空間の局所的なトポロジーを測る指標を提案し、それを利用して対話用語抽出の効果を検証

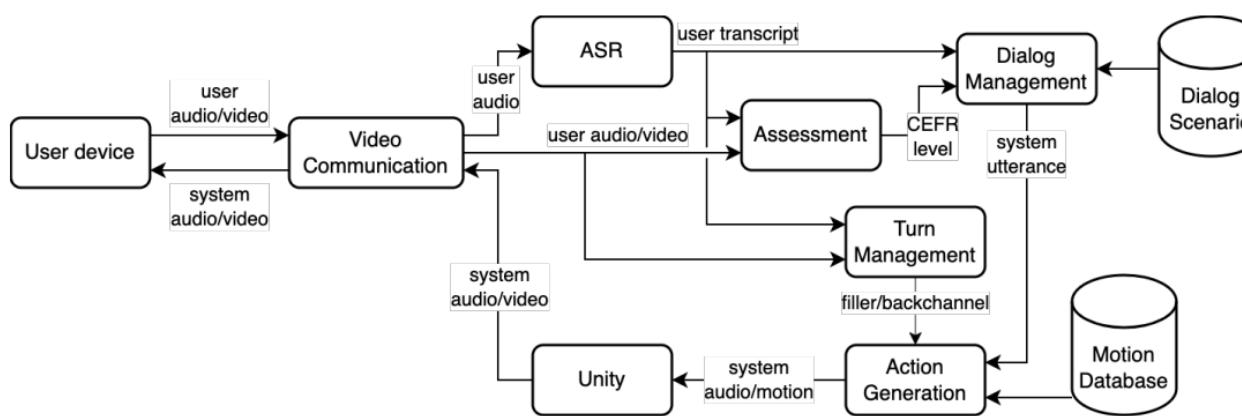


# InteLLA: Intelligent Language Learning Assistant for Assessing Language Proficiency through Interviews and Roleplays

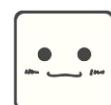


Mao Saeki, Hiroaki Takatsu, Fuma Kurata, Shungo Suzuki, Masaki Eguchi, Ryuki Matsuura, Kotaro Takizawa, Sadahiro Yoshikawa, Yoichi Matsuyama (Waseda University, Equmenopolis, Inc.)

- 従来の口頭能力評価システムは、ユーザーに一律のタスクを課すことが一般的
- インタビューやロールプレイを用いて**第二言語学習者の口頭能力を評価するマルチモーダル対話システム**
  - リアルタイム評価に基づき話題の難易度を動的に調整し、ユーザーの能力限界を引き出すことで、より正確で信頼性の高い評価を実現
- 大学生と高校生を対象とした実験では、システムによる自動評価が人間と同等の高い信頼性を示した



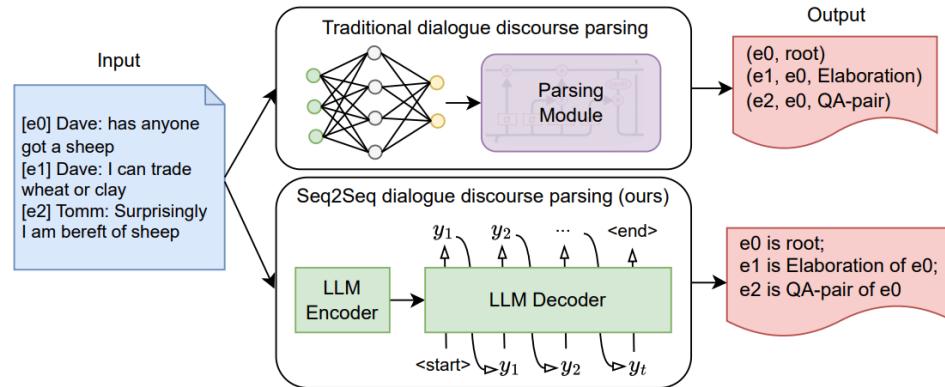
True \ Pred	A1	A2	B1	B2	C1	C2
A1	4	3	0	0	0	0
A2	0	25	6	0	0	0
B1	0	2	32	7	0	0
B2	0	0	3	49	1	0
C1	0	0	0	2	11	3
C2	0	0	0	0	0	4



# Dialogue Discourse Parsing as Generation: A Sequence-to-Sequence LLM-based Approach



Chuyuan Li, Yuwei Yin, Giuseppe Carenini (University of British Columbia)

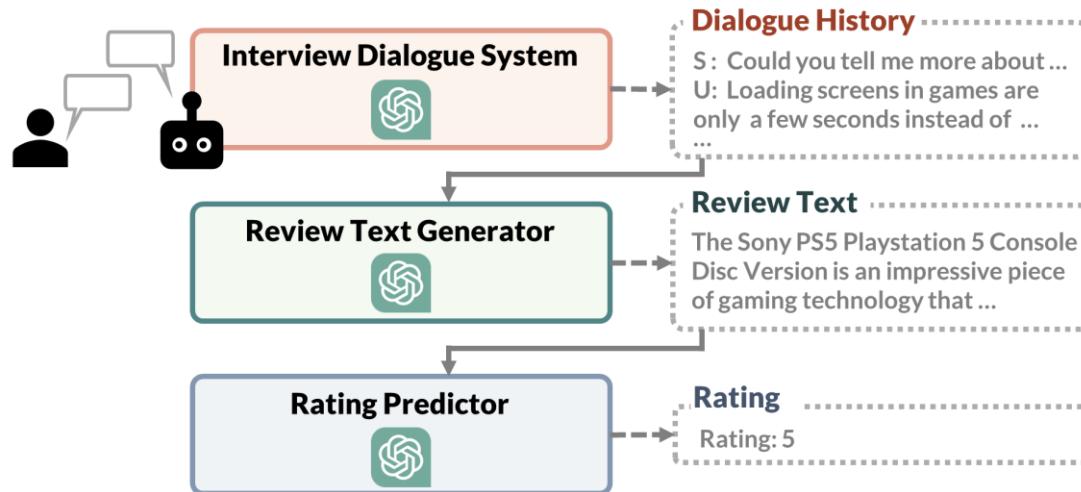


System	STAC		Molweni		
	Link ( $\Delta$ )	Full ( $\Delta$ )	Link ( $\Delta$ )	Full ( $\Delta$ )	
Natural (ours)	Seq2Seq-DDP+Transition	$70.8 \pm 0.9 (\uparrow 5.2)$	$55.1 \pm 1.0 (\uparrow 8.2)$	$83.5 \pm 0.2 (\uparrow 2.1)$	$60.3 \pm 0.1 (\uparrow 2.5)$
Focused (ours)	Seq2Seq-DDP+Transition	$72.3 \pm 0.6 (\uparrow 5.5)$	$56.6 \pm 0.6 (\uparrow 4.6)$	$83.4 \pm 0.6 (\uparrow 1.0)$	$60.0 \pm 0.5 (\uparrow 0.9)$
Shi and Huang (2019)	GRU+Pointer*	$72.9 \pm 0.4$	$54.2 \pm 0.5$	$77.9 \pm 0.4$	$54.1 \pm 0.6$
Liu and Chen (2021)	RoBERTa+Pointer	$72.9 \pm 1.5$	$57.0 \pm 1.0$	$79.0 \pm 0.4$	$55.4 \pm 1.8$
Chi and Rudnicky (2022)	RoBERTa+CLE <sup>†</sup>	$73.0 \pm 0.5$	$58.1 \pm 0.7$	$81.0 \pm 0.7$	$58.6 \pm 0.6$
Li et al. (2023c)	BERT+Biaffine+Pointer	73.0	58.5	83.2	59.8

- 談話解析タスクをSeq2SeqのLLMを用いた生成タスクに変換
  - 対話のテキスト系列を入力し、談話構造と関係ラベルを生成する
- 対話全体を一度に処理する方法と、文単位で逐次的に処理する方法を検討
- 2つのデータセットにおける性能評価の結果、SOTAと同等またはそれを上回る性能を達成

# User Review Writing via Interview with Dialogue Systems

Yoshiki Tanaka and Michimasa Inaba (The University of Electro-Communications)



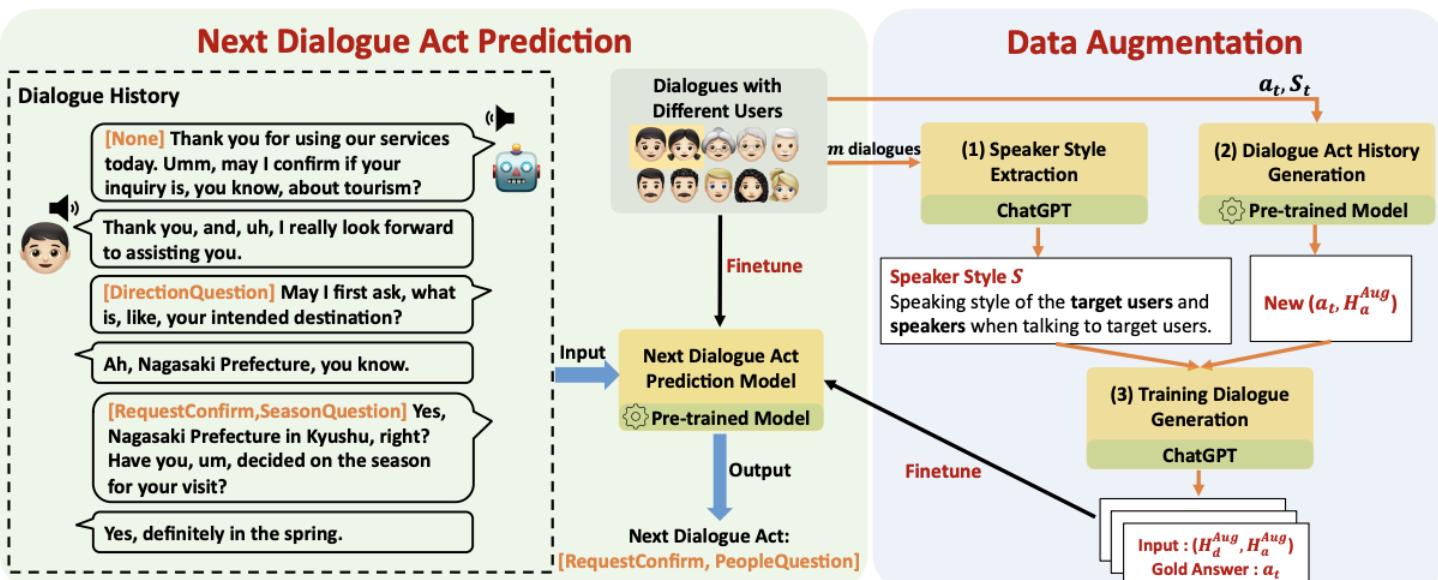
Review	Helpful.	Fluent.	Concise.	Exp.	Balance.	Detail.	Cover.	Overall.
Human	38.0	<b>47.0</b>	37.0	<b>57.0</b>	37.0	43.0	40.0	41.0
Tie	6.0	15.0	6.0	9.0	15.0	10.0	5.0	7.0
<b>Ours</b>	<b>56.0</b>	38.0	<b>57.0</b>	34.0	<b>48.0</b>	<b>47.0</b>	<b>55.0</b>	<b>52.0</b>
Baseline	38.5	28.1	45.8	21.9	34.4	35.5	35.5	37.5
Tie	12.5	33.3	6.2	16.7	15.6	12.5	10.4	17.7
<b>Ours</b>	<b>49.0</b>	<b>38.5</b>	<b>47.9</b>	<b>61.5</b>	<b>50.0</b>	<b>52.1</b>	<b>54.2</b>	<b>44.8</b>

- 対話システムを活用したユーザレビューの作成支援手法を提案
  - インタビュー対話を通じてユーザから商品を使用した際の感想・要望を引き出す
  - 対話履歴の情報をもとに、ユーザ視点のレビューを生成する
- 提案システムによるレビューが、人間によって書かれたレビューより高品質であることを確認

# Data Augmentation Integrating Dialogue Flow and Style to Adapt Spoken Dialogue Systems to Low-Resource User Groups

Zhiyang Qi and Michimasa Inaba (The University of Electro-Communications)

- 従来の対話システムは、大量の注釈付きデータを使用して学習されているが、特殊的な会話スタイルを持つユーザーのデータは限られており、収集も困難
- 未成年の対話データを例に、ターゲットユーザーグループに特化したデータ拡張フレームワークを提案
- 未成年の対話データを用いた低リソース設定において、提案手法は有効なデータが生成できることを検証



Split	Setting	Japanese GPT-NeoX		Japanese T5-base	
		Exact Match	Partial Match	Exact Match	Partial Match
1	Minors-Only	0.2451 ± 0.0117	0.3447 ± 0.0131	0.2533 ± 0.0083	0.3519 ± 0.0090
	Zero-Shot	0.2966 ± 0.0071	0.4049 ± 0.0092	0.3000 ± 0.0059	0.4066 ± 0.0053
	Low-Resource	0.3041 ± 0.0070	0.4228 ± 0.0073	0.3085 ± 0.0065	0.4232 ± 0.0064
	Low-Resource + Aug (Ours)	<b>0.3137 ± 0.0064</b>	<b>0.4320 ± 0.0094</b>	<b>0.3148 ± 0.0050</b>	<b>0.4244 ± 0.0056</b>
	Full-Resource	0.3190 ± 0.0074	0.4489 ± 0.0049	0.3125 ± 0.0029	0.4418 ± 0.0023
2	Minors-Only	0.2302 ± 0.0103	0.3677 ± 0.0105	0.2419 ± 0.0050	0.3311 ± 0.0079
	Zero-Shot	0.3162 ± 0.0069	0.4247 ± 0.0099	0.3200 ± 0.0039	0.4263 ± 0.0046
	Low-Resource	0.3220 ± 0.0071	0.4401 ± 0.0051	0.3257 ± 0.0019	0.4430 ± 0.0066
	Low-Resource + Aug (Ours)	<b>0.3290 ± 0.0083</b>	<b>0.4460 ± 0.0111</b>	<b>0.3270 ± 0.0029</b>	<b>0.4473 ± 0.0095</b>
	Full-Resource	0.3294 ± 0.0068	0.4526 ± 0.0074	0.3339 ± 0.0052	0.4486 ± 0.0075
3	Minors-Only	0.2329 ± 0.0033	0.3291 ± 0.0069	0.2528 ± 0.0038	0.3499 ± 0.0010
	Zero-Shot	0.2771 ± 0.0053	0.3878 ± 0.0075	0.2787 ± 0.0054	0.3889 ± 0.0054
	Low-Resource	0.2863 ± 0.0055	0.4070 ± 0.0019	0.2825 ± 0.0036	0.4010 ± 0.0156
	Low-Resource + Aug (Ours)	<b>0.2906 ± 0.0055</b>	<b>0.4077 ± 0.0067</b>	<b>0.2865 ± 0.0042</b>	<b>0.4097 ± 0.0090</b>
	Full-Resource	0.2889 ± 0.0069	0.4282 ± 0.0085	0.2986 ± 0.0058	0.4270 ± 0.0057
4	Minors-Only	0.2325 ± 0.0083	0.3336 ± 0.0093	0.2429 ± 0.0036	0.3480 ± 0.0091
	Zero-Shot	0.2900 ± 0.0066	0.4041 ± 0.0066	0.2947 ± 0.0047	0.4056 ± 0.0059
	Low-Resource	0.2925 ± 0.0067	0.4098 ± 0.0088	0.2983 ± 0.0031	<b>0.4156 ± 0.0120</b>
	Low-Resource + Aug (Ours)	<b>0.3005 ± 0.0069</b>	<b>0.4254 ± 0.0087</b>	<b>0.3000 ± 0.0056</b>	0.4144 ± 0.0096
	Full-Resource	0.3096 ± 0.0049	0.4425 ± 0.0098	0.3094 ± 0.0073	0.4336 ± 0.0019



# バンケット



お庭の景色



@平安神宮会館



# YRRSDS (Workshop of Young Researchers' Roundtable on Spoken Dialogue Systems)

- 対話システムの若手研究者たちが議論・交流を行うワークショップ

- 発表要件：ポジションペーパーの提出

- 「自分研究内容」、「音声対話研究の将来」、「議論の提案」に関する原稿（2ページ程度）
- 提出締め切りはSIGDIALの結果通知の後

- 発表者数は32名、参加者数はもう少し多い(50-60名程度)

- プログラム

- 口頭発表(1人5分), ラウンドテーブル(30分, 3回), キーノート(by 吉野先生, 松山先生)
  - ラウンドテーブルのトピック : LLM, マルチモーダリティ, 説明可能性, 評価, 信頼性, 倫理, SDSの未来など



# YRRSDS (Workshop of Young Researchers' Roundtable on Spoken Dialogue Systems)



@京大時計台



@銀閣寺

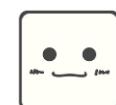
# SIGDIAL2024の所感

## ✓ 対話研究者との交流機会が多い

- Banquet, Reception, Social Activity (YRRSDS) など
- 研究に対するモチベーションUP
- 同分野における外部の研究者と知り合うことができた

## ✓ 自身の研究について質問・議論を持ちかけてくれる人も

- 研究を見つめ直す良い機会になった



# SIGDIAL2025

- 開催地：フランス・アヴィニヨン
- 日程：8.25(月) - 8.27(水)
- 場所：アヴィニヨン大学

