Dialogue Structure Parsing on Multi-Floor Dialogue Based on Multi-Task Learning

Seiya Kawano[1], Koichiro Yoshino[1,2], David Traum[3], Satoshi Nakamura[1,2]
[1] Nara Institute of Science and Technology
[2] RIKEN, Center for Advanced Intelligence Project AIP
[3] USC Institute for Creative Technologies

Multi-floor dialogue [traum+18]

- Our dialogues are often across multiple floors
- E.g., restaurant where some people take the customer's order and others make the food

Dining room

Kitchen room

Sever is a participant of both dialogue floors and coordinating each to achieve a shared dialogue goal

- **Multi-floor dialogue** [traum+18]
  - Our dialogues are often across multiple floors
  - E.g., military units, where orders are relay through the chain of command

Soldiers

Headquarters

Soldiers follow their commander's orders, which are decided at headquarters (one or more person coordinate communication of both floors)

Multi-floor dialogue [traum+18]

- Our dialogues are often across multiple floors
- E.g., order at restaurant, command chain of military unit, etc.

1. having the same purpose but distinct participants

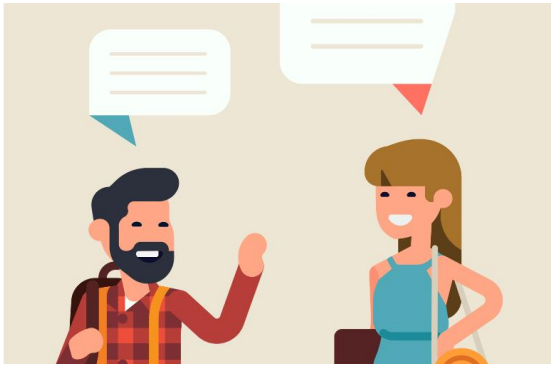
2. having one or more (but not all) participants in common, where such participants are multicommunicating [Reinsch+18]

3. Multicommunicating participants coordinating each floor for achieving a shared dialogue goal

Such situations of distributed decision-making and action are quite common

Difference with single-floor dialogue

- **Single-floor dialogue**
  - All participants can hear all communication (sharing the same dialogue flow) unlike the multi-floor dialogue
  - E.g., two people talking face to face, online conference, etc.



- **Previous work in area of discourse and dialogue**
  - Not considering multi-floor setting although whereas multi-floor situations are common
  - Identifying structures of such dialogue can be critical for building a cooperative application to address multi-floor dialogues

[Rukin+2018, Bonial+2018]

Building dialogue structure parser

Dialogue structure parsing on multi-floor dialogues

- Many annotation schemas and dialogue structure parsing model have been proposed in single-floor dialogue [Bunt+12,Prasd+15]

- However, there is no previous work on automatic dialogue structure parsing for multi-floor dialogue

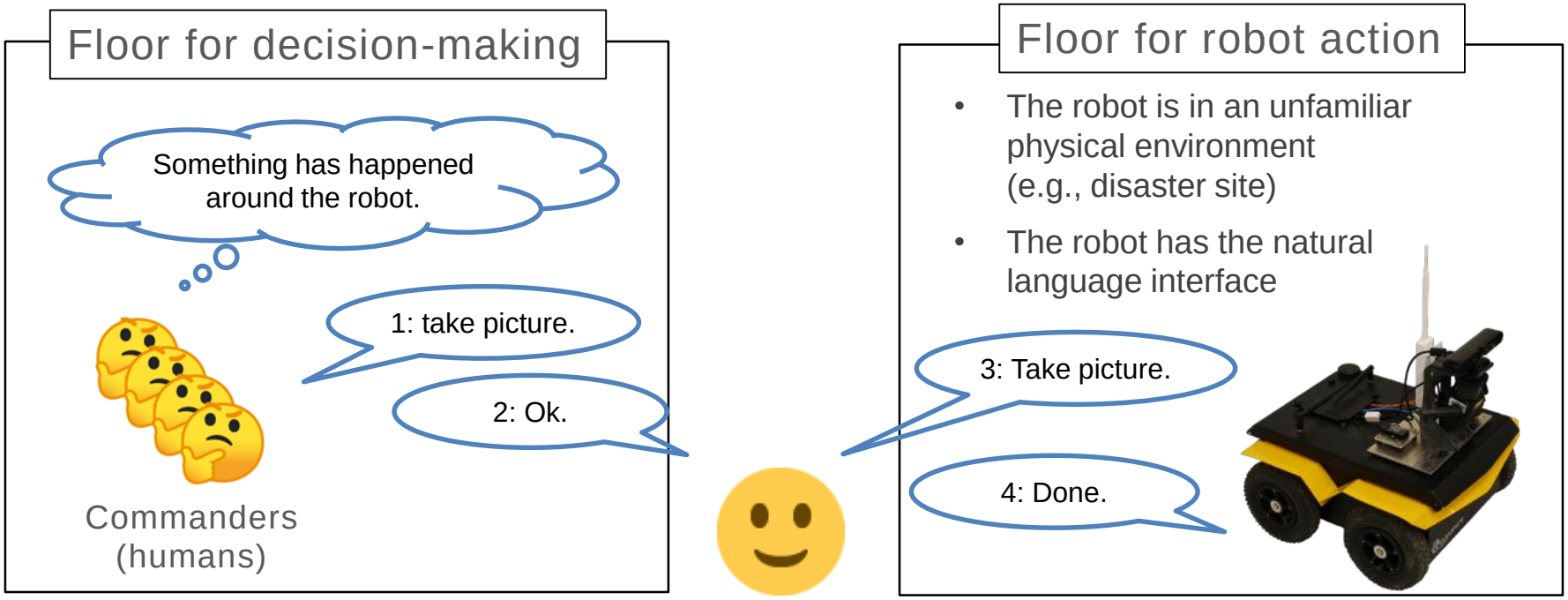
We need to substantial model for automatically identifying the structure of multi-floor dialogue

Expected contributions

1. Development of annotated resource of multi-floor dialogue
2. Development of dialogue models for addressing issues of multi-floor dialogues

Target domain and task

- **Collaborative robot navigation**
  - Natural language interaction with robot and remotely located human participants on exploration and navigation tasks



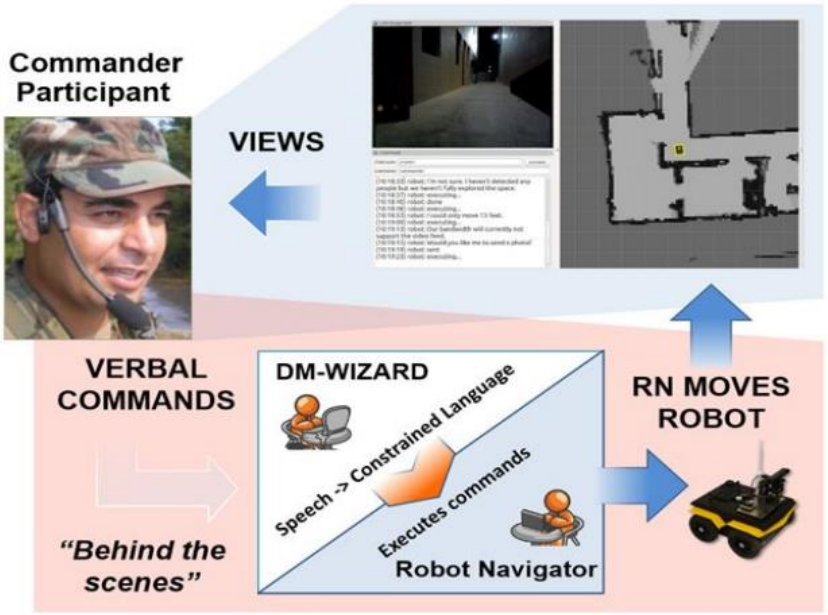
E.g., situation of urban search and rescue (USRA), etc.

Target domain and task

- **We use the dataset of multi-floor dialogue**
  - The dataset was created by Wizard-of-Woz experiments
  - Based on the minimum requirement of multi-floor dialogue

  - Three participants:
    - CM: Commander
    - DM: Dialogue manager (Multicommunicator)
    - RN: Robot navigator

  - Two different floors:
    - Floor for decision-making (CM ↔ DM)
    - Floor for robot action (rDM↔RN)



Example of multi-floor dialogue

	Left Floor		Right Floor		Annotations		
#	**Commander**	**DM → Commander**	**DM → RN**	**RN**	**TU**	**Ant**	**Rel**
1	move to where you see the first cone				1		
2		I ' m not sure which object you are referring to. Can you describe it in another way, using color or its location?			1	1	request-clarification
3	move to the cone on the right a red cone on the right				1	2	clarification-repair
4			move to face the cone on the right		1	3	translation-r
5		executing...			1	3	ack-doing
6	take another picture				2		
7				done	1	4	ack-done
8		done			1	7	translation-l
9			image		2	6	translation-r
10				image sent	2	9	ack-done
11		sent			2	10	translation-l

- Including two entangled transactions:
  - *#1*: *Move to where you see first cone*
    (*#4 is the same transaction as #1*)
  - *#2*: *Take another picture*

Annotation scheme of multi-floor dialogue

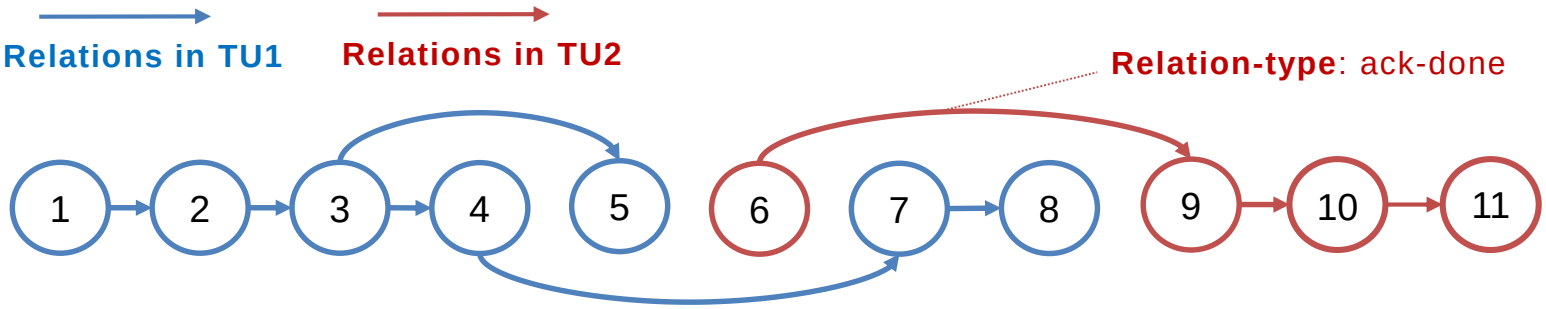
- **Transaction unit (TU)**
  - Clusters of utterances from multiple participants and floors that contribute to achieving the commander's specific intention
- **Antecedent (Ant)**
  - Link-relation of two utterances
- **Relation-type (Rel)**
  - Type of relations, that between individual utterances connect by the antecedent within the TU

[Traum+2018, LREC]

Type	Sub-types
Expansions	relate utterances that are produced by the same participant within the same floor.
	continue
	link-next
	correction
	summarization
Responses	relate utterances by different participants within the same floor.
	acknowledgment
	done
	doing
	wilco
	understand
	try
	unsure
	can't
	clarification
	req-clar
	clar-repair
	missing info
	nack
	repeat
	processing
	question-response
	answer
	non-answer
	other
	3rd turn feedback
	reciprocal response
Translations	relate utterances in different floors.
	transalation-l
	transalation-r
	comment
	quotation

Example of multi-floor dialogue structure

	Left Floor		Right Floor		Annotations		
#	Commander	DM → Commander	DM → RN	RN	TU	Ant	Rel
1	move to where you see the first cone				1		
2		I ' m not sure which object you are referring to. Can you describe it in another way, using color or its location?			1	1	request-clarification
3	move to the cone on the right a red cone on the right				1	2	clarification-repair
4			move to face the cone on the right		1	3	translation-r
5		executing...			1	3	ack-doing
6	take another picture				2		
7				done	1	4	ack-done
8		done			1	7	translation-l
9			image		2	6	translation-r
10				image sent	2	9	ack-done
11		sent			2	10	translation-l



Task definition of dialogue structure parsing

1. **Transaction-unit prediction**
   - We can formalize as a boundary classification problem

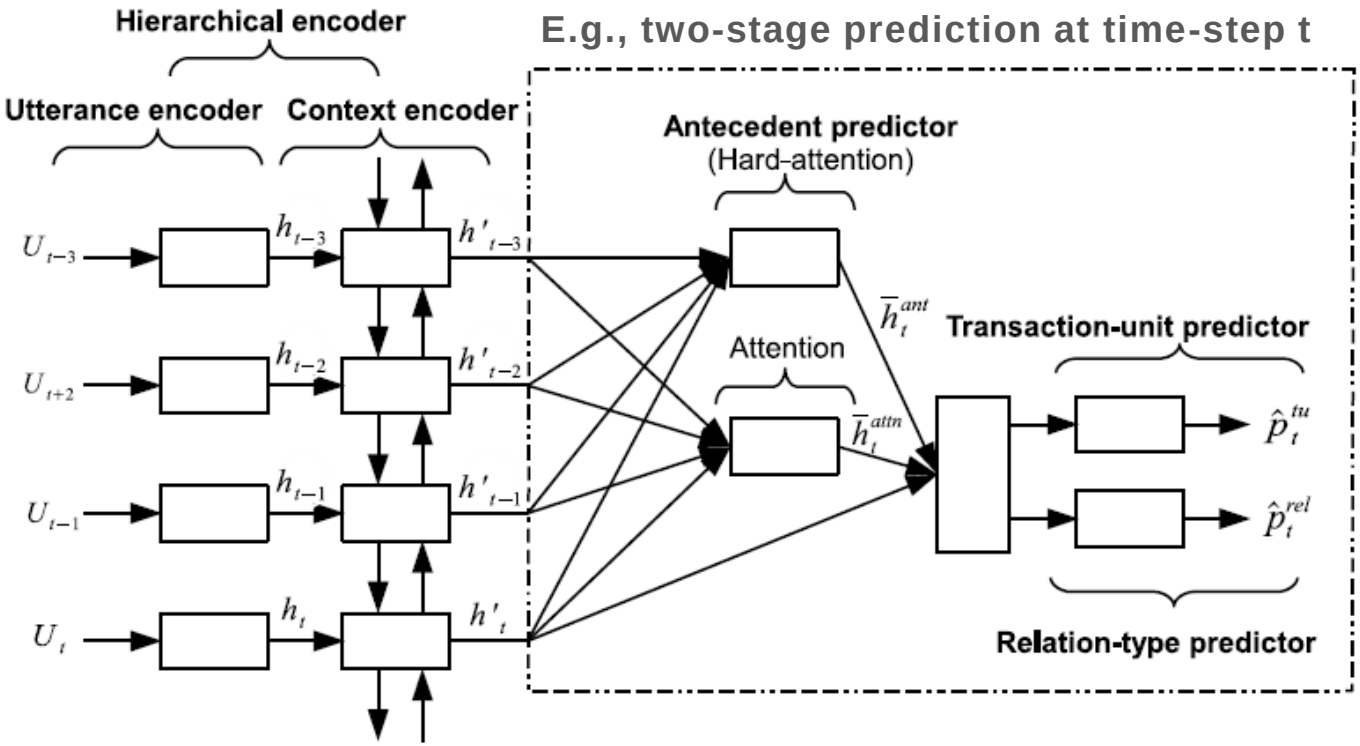
> **Start**: the utterance is the beginning of a transaction unit.
>
> **Continue**: the utterance belongs to the same transaction unit as the previous utterance.
>
> **Other**: the utterance cannot be categorized into either of the above classes.

2. **Antecedent prediction**
   - We predict the relevance of utterance pairs
   - High relevance indicates that the two are linked

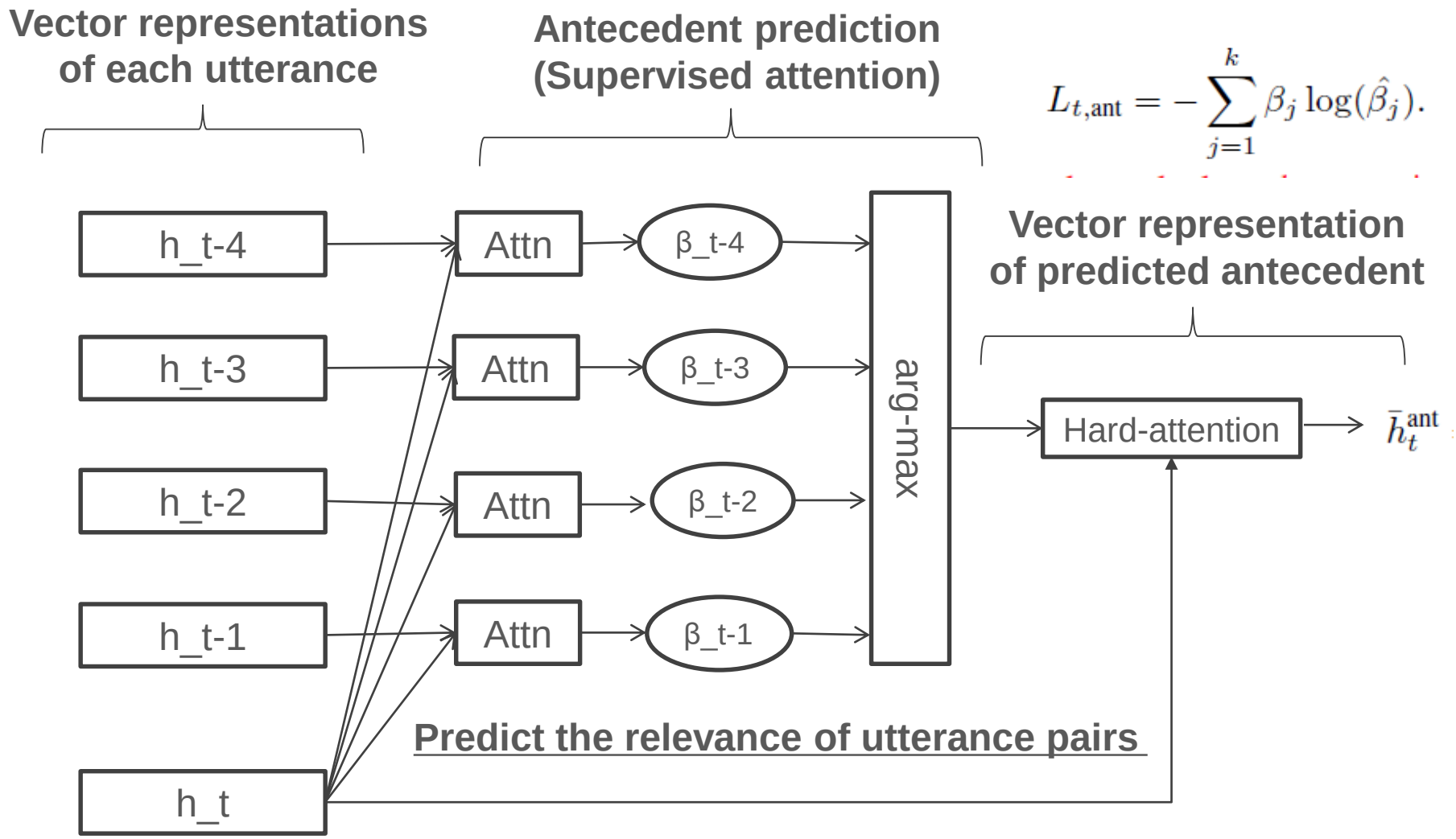
3. **Relation-type prediction**
   - We predict the relation-type of the utterance (and its antecedents)

Proposed dialogue structure parser

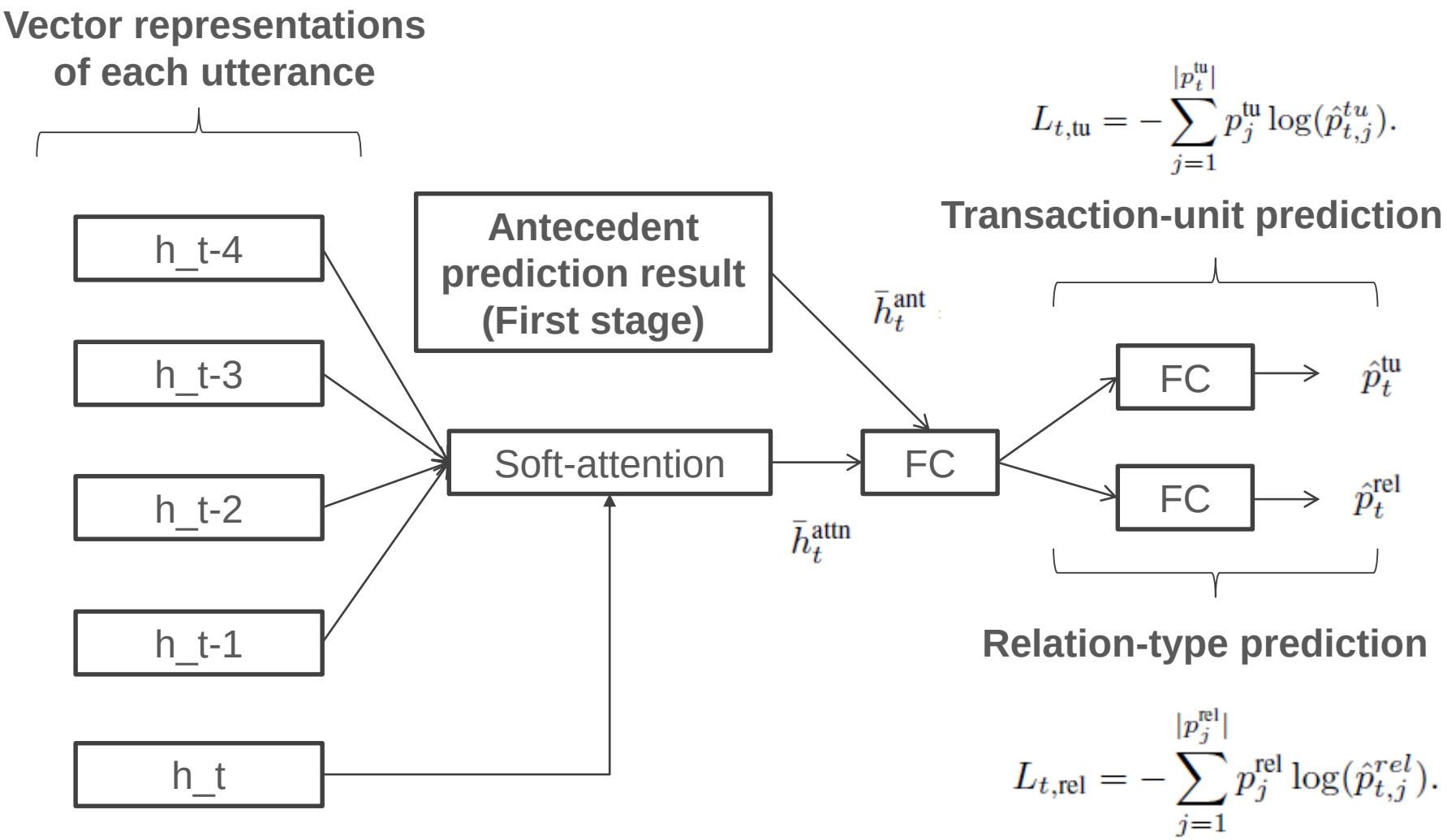
- **We introduce the multi-task prediction architecture**
  - Definitions of TU, Ant, and Rel are complementary
  - We unified each prediction task by multi-task learning and attention mechanism for improving overall parsing performance



First stage: antecedent prediction



Second stage: transaction-unit & relation-type prediction



Experimental settings

- **Training objectives**
  - Single-task: separately trained for each prediction model
  - Multi-task: proposed multi-task prediction model

$$L = \frac{1}{N}\sum_{t=1}^{N}(\gamma_{\text{ant}}L_{t,\text{ant}} + \gamma_{\text{tu}}L_{t,\text{tu}} + \gamma_{\text{rel}}L_{t,\text{rel}}).$$

antecedent transaction-unit relation-type

N is the dialogue length
set γ_{ant}, γ_{tu}, and γ_{rel} to 1

- **Online (incremental) vs. offline prediction**
  - Online prediction (unidirectional context encoder):
    only uses previous contexts without subsequent contexts
  - Offline prediction (bidirectional context encoder):
    uses both previous contexts and subsequent contexts

- **Training the floor-type embedding**
  - Add a special symbol, which indicates the types of floors,
    to prefixes and suffixes of utterances

Experimental settings

Dataset

- Annotated human-robot collaborative dialogue corpus
- Total 48 dialogues
  - Avg. dialogue length≒240
  - Avg. transactions≒34
  - Avg. transaction length≒7

	Dialogues	Utterances	Transactions
Exp. 1	24	4527	780
Exp. 2	24	6994	1049

Evaluation

- Double-cross validation (6 subsets)
- Micro-level metrics: classification accuracy (precision/recall/**F1**)
- Meso-level metrics:
  - **TuAcc**: ratio of perfectly predict the TU spans
  - **TreeAcc**: ratio of perfectly predict the link-relations within TUs
  - **TreeAcc w/ rel**:
    ratio of perfectly predicted the link-relations and types within TUs

Experimental results – overall performance

Online Prediction	TU		Ant		Rel	
	F1	TuAcc	F1	TreeAcc	F1	TreeAcc w/ rel
Majority	63.80	-	31.76	-	13.21	-
Single-task	95.44	81.19	92.34	68.12	92.53	63.80
- w/o floor	94.43	77.41	90.43	65.59	91.31	60.30
Multi-task	95.99	84.25	92.33	70.09	93.80	66.81
- w/o floor	94.62	78.18	90.82	66.86	91.58	63.33

Offline Prediction	TU		Ant		Rel	
	F1	TuAcc	F1	TreeAcc	F1	TreeAcc w/ rel
Single-task	95.33	81.46	92.40	68.83	92.91	64.62
w/o floor	94.57	77.96	91.81	67.57	91.79	62.38
Multi-task	96.06	84.52	93.21	71.35	93.90	69.09
w/o floor	94.93	78.73	92.08	69.21	92.16	68.67

Experimental results – overall performance

Online Prediction	TU		Ant		Rel	
	F1	TuAcc	F1	TreeAcc	F1	TreeAcc w/ rel
Majority	63.80	-	31.76	-	13.21	-
Single-task	95.44	81.19	92.34	68.12	92.53	63.80
- w/o floor	94.43	**77.41**	90.43	65.59	91.31	60.30
Multi-task	95.99	84.25	92.33	70.09	93.80	66.81
- w/o floor	94.62	**78.18**	90.82	66.86	91.58	63.33

Offline Prediction	TU		Ant		Rel	
	F1	TuAcc	F1	TreeAcc	F1	TreeAcc w/ rel
Single-task	95.33	81.46	92.40	68.83	92.91	64.62
w/o floor	94.57	**77.96**	91.81	67.57	91.79	62.38
Multi-task	96.06	84.52	93.21	71.35	93.90	69.09
w/o floor	94.93	**78.73**	92.08	69.21	92.16	68.67

- **w/ . vs. w/o floor**

✓ If the floor information can not be used, the parsing performance will be decreased

✓ This result suggests that modeling considering the nature of the floor is indispensable

Experimental results – overall performance

Online Prediction	TU		Ant		Rel	
	F1	TuAcc	F1	TreeAcc	F1	TreeAcc w/ rel
Majority	63.80	-	31.76	-	13.21	-
Single-task	95.44	81.19	92.34	68.12	92.53	63.80
- w/o floor	94.43	77.41	90.43	65.59	91.31	60.30
Multi-task	**95.99**	**84.25**	**92.33**	**70.09**	**93.80**	**66.81**
- w/o floor	94.62	78.18	90.82	66.86	91.58	63.33

Offline Prediction	TU		Ant		Rel	
	F1	TuAcc	F1	TreeAcc	F1	TreeAcc w/ rel
Single-task	95.33	81.46	92.40	68.83	92.91	64.62
w/o floor	94.57	77.96	91.81	67.57	91.79	62.38
Multi-task	**96.06**	**84.52**	**93.21**	**71.35**	**93.90**	**69.09**
w/o floor	94.93	78.73	92.08	69.21	92.16	68.67

- **Single vs. multi-task**

✓ Multi-task models showed improvement from single-task models

✓ Especially, showed that improvement of TuAcc and TreeAcc (Extracting a more consistent tree structure)

Experimental results – overall performance

Online Prediction	TU		Ant		Rel	
	F1	TuAcc	F1	TreeAcc	F1	TreeAcc w/ rel
Majority	63.80	-	31.76	-	13.21	-
Single-task	95.44	81.19	92.34	68.12	92.53	63.80
- w/o floor	94.43	77.41	90.43	65.59	91.31	60.30
Multi-task	95.99	84.25	92.33	70.09	93.80	66.81
- w/o floor	94.62	78.18	90.82	66.86	91.58	63.33

Offline Prediction	TU		Ant		Rel	
	F1	TuAcc	F1	TreeAcc	F1	TreeAcc w/ rel
Single-task	**95.33**	**81.46**	**92.40**	**68.83**	**92.91**	**64.62**
w/o floor	94.57	77.96	91.81	67.57	91.79	62.38
Multi-task	**96.06**	**84.52**	**93.21**	**71.35**	**93.90**	**69.09**
w/o floor	94.93	78.73	92.08	69.21	92.16	68.67

- **Online vs. offline**

✓ Offline models showed slight improvement from online-models

Experimental results – overall performance

Online Prediction	TU		Ant		Rel	
	F1	TuAcc	F1	TreeAcc	F1	TreeAcc w/ rel
Majority	63.80	-	31.76	-	13.21	-
Single-task	95.44	81.19	92.34	68.12	92.53	63.80
- w/o floor	94.43	77.41	90.43	65.59	91.31	60.30
Multi-task	95.99	84.25	92.33	70.09	93.80	66.81
- w/o floor	94.62	78.18	90.82	66.86	91.58	63.33

Offline Prediction	TU		Ant		Rel	
	F1	TuAcc	F1	TreeAcc	F1	TreeAcc w/ rel
Single-task	95.33	81.46	92.40	68.83	92.91	64.62
w/o floor	94.57	77.96	91.81	67.57	91.79	62.38
Multi-task	**96.06**	**84.52**	**93.21**	**71.35**	**93.90**	**69.09**
w/o floor	94.93	78.73	92.08	69.21	92.16	68.67

- **w/ . vs. w/o floor**

✓ If the floor information can not be used, the parsing performance will be decreased

- **Single vs. multi-task**

✓ Multi-task models showed improvement from single-task models

- **Online vs. offline**

✓ Offline models showed slight improvement from online-models

Example of parsing result – error case

	Left Floor		Right Floor		Prediction		
#	**Commander**	**DM → Commander**	**DM → RN**	**RN**	**TU**	**Ant**	**Rel**
1	take a picture				Start	#	#
2			image		Continue	1	translation-r
3				image sent	Continue	2	response-ack.
4		sent			Continue	3	translation-l
5	turn left ninety degrees				Start	#	#
6			turn left 90		Continue	5	translation-r
7		executing ...			Continue	5	response-ack.
8	take a picture after each command				Start **(Continue)**	# **(5)**	# **(expansion-cont.)**
9				done	Other **(Continue)**	6	response-ack.
10			take pic after each command		Other **(Continue)**	8	translation-r
11			image		Other **(Continue)**	8	translation-r
12				image sent	Continue	11	response-ack.
13		sent			Continue	12	translation-l

*** () is the actual label**

- Error in line#8 is propagated to the later turns(even if we assume the prediction of TU at #8 is correct, the prediction at #11 is still not correct)
- Reason for errors
  - variance of annotation quality of training data
  - entanglement of transactions due to communication delays

Summary

- **Proposal**
  - We proposed the first baseline model that automatically identifies the structure of multi-floor dialogues based on multi-task learning and an attention mechanism
  - We compared the different prediction settings

- **Results**
  - Our proposed models showed that a promising parsing performance of multi-floor dialogue structure

- **Future works**
  - Exploring the possibility of introducing powerful approaches of similar tasks related for predicting tree-structure in text
  - Developing actual dialogue system for multi-floor dialogue

End of slide.

Acknowledgement

This research and development work was supported by Grant-in-Aid for JSPS Fellows No. 20J14823. We would like to thank United States Army Research Laboratory for sharing the dataset for this research. Author Traum was supported in part by the Army Research Office under Cooperative Agreement Number W911NF-20-2-0053. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

Appendix.