Multiparty Conversation Modeling

Maximillian Chen

Advisor: Zhou Yu

Committee: Julia Hirschberg, Kathleen McKeown, Zhou Yu

April 2023

Contents

Contents

What are Multiparty Conversations?

- Broadly speaking: any conversation that involves more than two *or more* interlocutors

- Why is this interesting?

Traum, David. “Issues in Multiparty Dialogues”
Afantenos, Stergos et al. “Discourse Parsing for Multi-Party Chat Dialogues”
Gu, Jia-Chen et al. “WHO Says WHAT to WHOM: A Survey of Multi-Party Conversations”

What are Multiparty Conversations?

Who says *what* to *whom*?

- Who…
  - is/should be speaking?
- What…
  - is being/should be said?
- Whom…
  - is being/should be addressed?

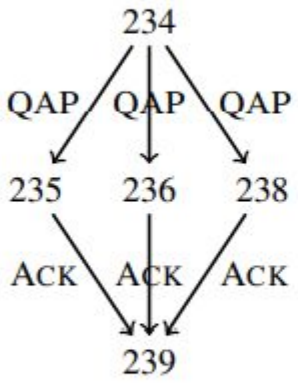
Gu, Jia-Chen et al. "WHO Says WHAT to WHOM: A Survey of Multi-Party Conversations"

Multiparty Conversations Are Complex

- Many attempts to create environments to study these types of questions, e.g. STAC

- This is a complex enough context that can only be represented as a graph

234	18:55:02:745	gotwood4sheep	anyone got wheat for a sheep?
235	18:55:10:047	inca	sorry, not me
236	18:55:18:787	CheshireCatGrin	nope. you seem to have lots of sheep!
237	18:55:23:428	gotwood4sheep	yup baaa
238	18:55:32:308	dmm	i think i'd rather hang on to my wheat i'm afraid
239	18:55:47:845	gotwood4sheep	kk I'll take my chances then...



Afantenos, Stergos et al. "Discourse Parsing for Multi-Party Chat Dialogues"

Multiparty Conversation Tasks

- Speaker Identification
  - who is the speaker?
- Turn-taking
  - rigid interaction like with chatbots, or barge-in?

234	18:55:02:745	gotwood4sheep	anyone got wheat for a sheep?
235	18:55:10:047	inca	sorry, not me
236	18:55:18:787	CheshireCatGrin	nope. you seem to have lots of sheep!
237	18:55:23:428		yup baaa

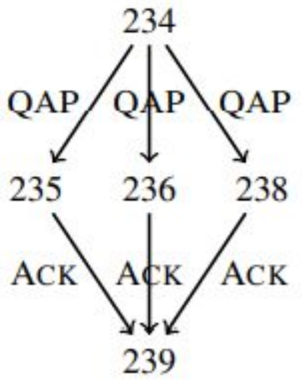
agreeing they have lots of sheep? probably gotwood4sheep!

Traum, David. “Issues in Multiparty Dialogues”
Afantenos, Stergos et al. “Discourse Parsing for Multi-Party Chat Dialogues”
Gu, Jia-Chen et al. “WHO Says WHAT to WHOM: A Survey of Multi-Party Conversations”

Multiparty Conversation Tasks

- Thread Management, Dialogue Disentanglement
- Addressee Recognition

234	18:55:02:745	gotwood4sheep	anyone got wheat for a sheep?
235	18:55:10:047	inca	sorry, not me
236	18:55:18:787	CheshireCatGrin	nope. you seem to have lots of sheep!
237	18:55:23:428	gotwood4sheep	yup baaa
238	18:55:32:308	dmm	i think i'd rather hang on to my wheat i'm afraid
239	18:55:47:845	gotwood4sheep	kk I'll take my chances then...



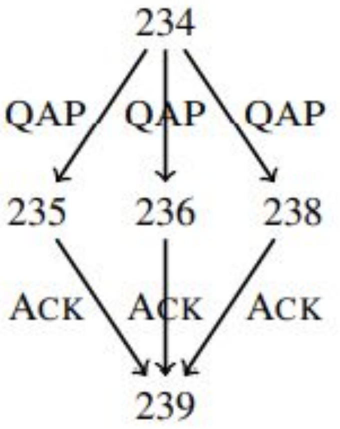
We can infer 236 is addressing gotwood4sheep

Traum, David. "Issues in Multiparty Dialogues"
Afantenos, Stergos et al. "Discourse Parsing for Multi-Party Chat Dialogues"
Gu, Jia-Chen et al. "WHO Says WHAT to WHOM: A Survey of Multi-Party Conversations"

Multiparty Conversation Tasks

- Discourse Parsing
  - discourse relations
  - discourse links

234	18:55:02:745	gotwood4sheep	anyone got wheat for a sheep?
235	18:55:10:047	inca	sorry, not me
236	18:55:18:787	CheshireCatGrin	nope. you seem to have lots of sheep!
237	18:55:23:428	gotwood4sheep	yup baaa
238	18:55:32:308	dmm	i think i'd rather hang on to my wheat i'm afraid
239	18:55:47:845	gotwood4sheep	kk I'll take my chances then...



Afantenos, Stergos et al. "Discourse Parsing for Multi-Party Chat Dialogues"

Multiparty Conversation Tasks

- Attention Management
  - when to bring new participants into the conversation

system engages bystander ~ t_5

Traum, David. “Issues in Multiparty Dialogues”
Bohus, Dan and Eric Horvitz. “Models for Multiparty Engagement in Open-World Dialog”

Multiparty Conversation Tasks

- Initiative Management
  - “Leaders” in the conversation develop either formally/informally

system engages bystander ~ t_5

Traum, David. “Issues in Multiparty Dialogues”

Bohus, Dan and Eric Horvitz. “Models for Multiparty Engagement in Open-World Dialog”

Difficulties of Multiparty Evaluation

- Ensure systems maintain correct long-term context

- Respond fairly

- Understand whether systems are contributing to conversational "success"

	Violation of Form	Violation of Content
Utterance	(I1) Uninterpretable (I2) Grammatical error	(I3) Semantic error (I4) Wrong information
Response	(I5) Ignore question (I6) Ignore request (I7) Ignore proposal (I8) Ignore greeting	(I9) Ignore expectation *(I8) Forgot speaker* *(I19) Forgot addressee(s)*
Context	(I10) Unclear intention (I11) Topic transition error (I12) Lack of information	(I13) Self-contradiction (I14) Contradiction (I15) Repetition
Society	(I16) Lack of sociality	(I17) Lack of common sense
Participant	*(I20) Wrong speaker* *(I21) Wrong addressee(s)*	*(I22) Wrong thread response* *(I23) Inappropriately timed initiative*

Mahajan, Khyati et al. "Towards Evaluation of Multi-party Dialogue Systems"

Difficulties of Multiparty Evaluation

- System credits the wrong speaker with information

U1: We need to consider factors A and B for making a decision in case X.

U2: Factor C would also be interesting and important to consider along with A and B.

S: U1 mentions factors C will be taken into consideration for case X.

Mahajan, Khyati et al. “Towards Evaluation of Multi-party Dialogue Systems”

Difficulties of Multiparty Evaluation

- System replies to the wrong conversation thread

U1: This football season has been going great!

U2: I agree, for most teams anyway. Which one is your favorite?

U3: I prefer soccer instead. Anyone here a soccer fan?

U4: I don't really pay much attention to sports. My main hobby is movies!

U5: Yeah, and Knives Out was a great one!

S: I agree U5! The Rams are doing so well this year!

Mahajan, Khyati et al. "Towards Evaluation of Multi-party Dialogue Systems"

Contents

Why does Multiparty Dialogue Understanding Matter?

- Building conversational agents which are…
  - intelligent, personable, adaptive

Why does Multiparty Dialogue Understanding Matter?

- Building conversational agents which are…
  - intelligent, personable, adaptive
- Before even performing generation itself…

Why does Multiparty Dialogue Understanding Matter?

- Building conversational agents which are…
  - intelligent, personable, adaptive
- Before even performing generation itself…
  - what are the subconversations it can respond to?

Why does Multiparty Dialogue Understanding Matter?

- Building conversational agents which are…
  - intelligent, personable, adaptive
- Before even performing generation itself…
  - what are the subconversations it can respond to?
  - who is participating in the conversation?

dialogue disentanglement

Why does Multiparty Dialogue Understanding Matter?

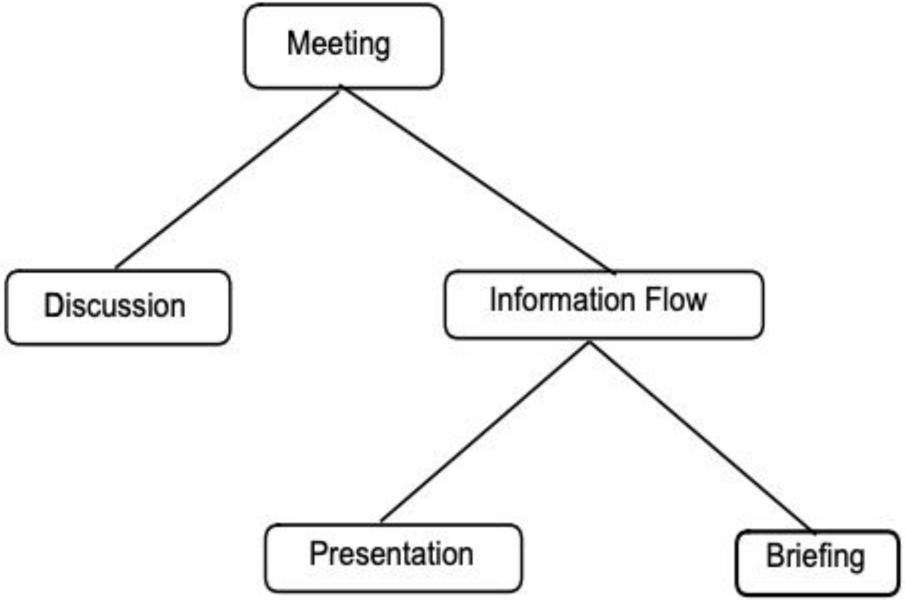
- Building conversational agents which are…
  - intelligent, personable, adaptive
- Before even performing generation itself…
  - what are the subconversations it can respond to? dialogue disentanglement
  - who is participating in the conversation? persona understanding/speaker ID
  - what have the participants already said?

Why does Multiparty Dialogue Understanding Matter?

- Building conversational agents which are…
  - intelligent, personable, adaptive
- Before even performing generation itself…
  - what are the subconversations it can respond to? dialogue disentanglement
  - who is participating in the conversation? speaker ID/persona understanding
  - what have the participants already said? discourse parsing

CMU Meetings Corpus

- Audio data from 30 min meeting
- Meeting state and participant role taxonomy
- Early attempt at defining multiparty conversation structures



Banerjee, Satanjeev and Alexander I. Rudnicky. “Using Simple Speech-Based Features to Detect the State of a Meeting and the Roles of the Meeting Participants”

ICSI Meeting Corpus

#	
53	Unique speakers
13	Female
40	Male

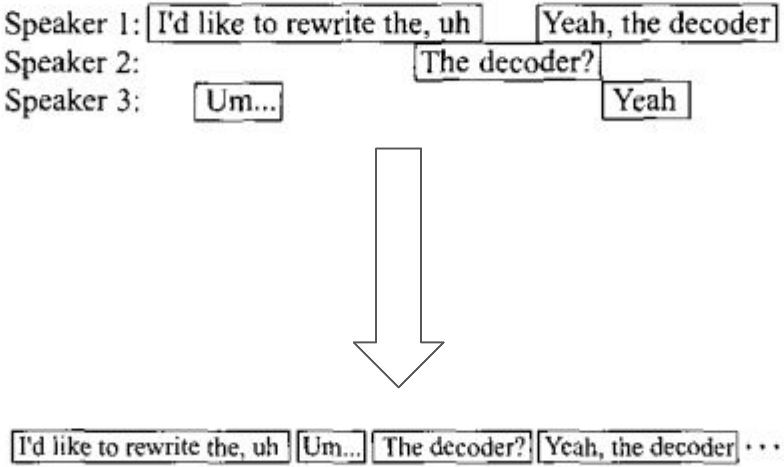
#	Age
18	20–29
18	30–39
4	40–49
4	50–59
1	60+
8	Unspecified

#	Native Language
28	English
12	German
5	Spanish
1	Chinese
1	Czeck
1	Dutch
1	French
1	Hebrew
1	Malayalam
1	Norwegian
1	Turkish

#	Education Level
21	Grad
20	PhD
7	Professor
4	Undergrad
1	Postdoc

#	Variety of English
36	American
6	British
2	German
2	Indian
1	Czeck
1	Norwegian
5	Unspecified

#	Time Spent in English Speaking Country
9	< 1 Year
3	1–2 Years
4	2–5 Years
6	> 5 Years
3	Unspecified

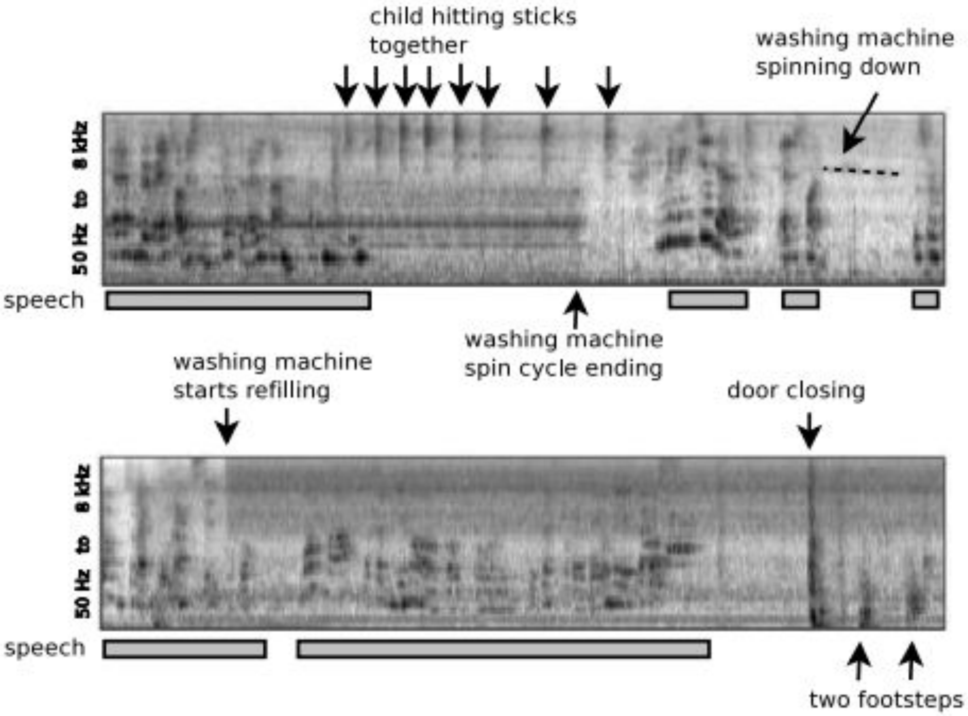


- Large audio+text meeting corpus
- Overlaps in speech help with speaker identification

Janin, Adam et al. “The ICSI Meeting Corpus”

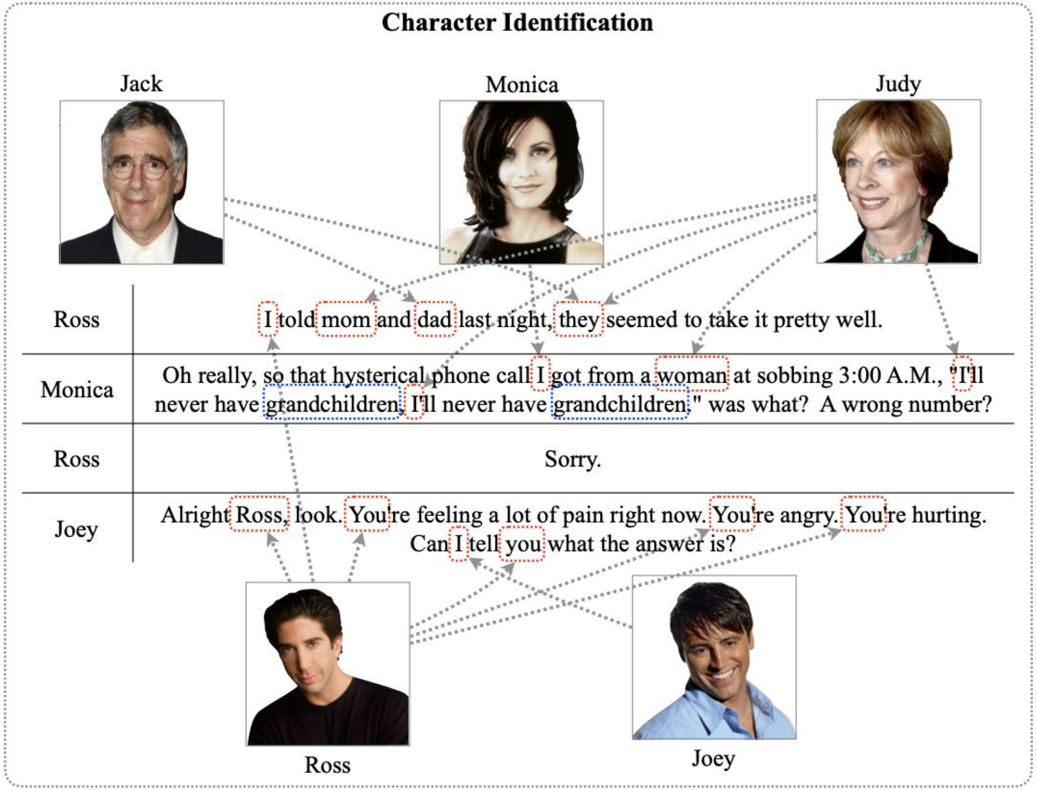
Computational Hearing in Multisource Environments (CHiME)

- 40 hours of audio data from **domestic** environment
- Natural and controlled levels of noise



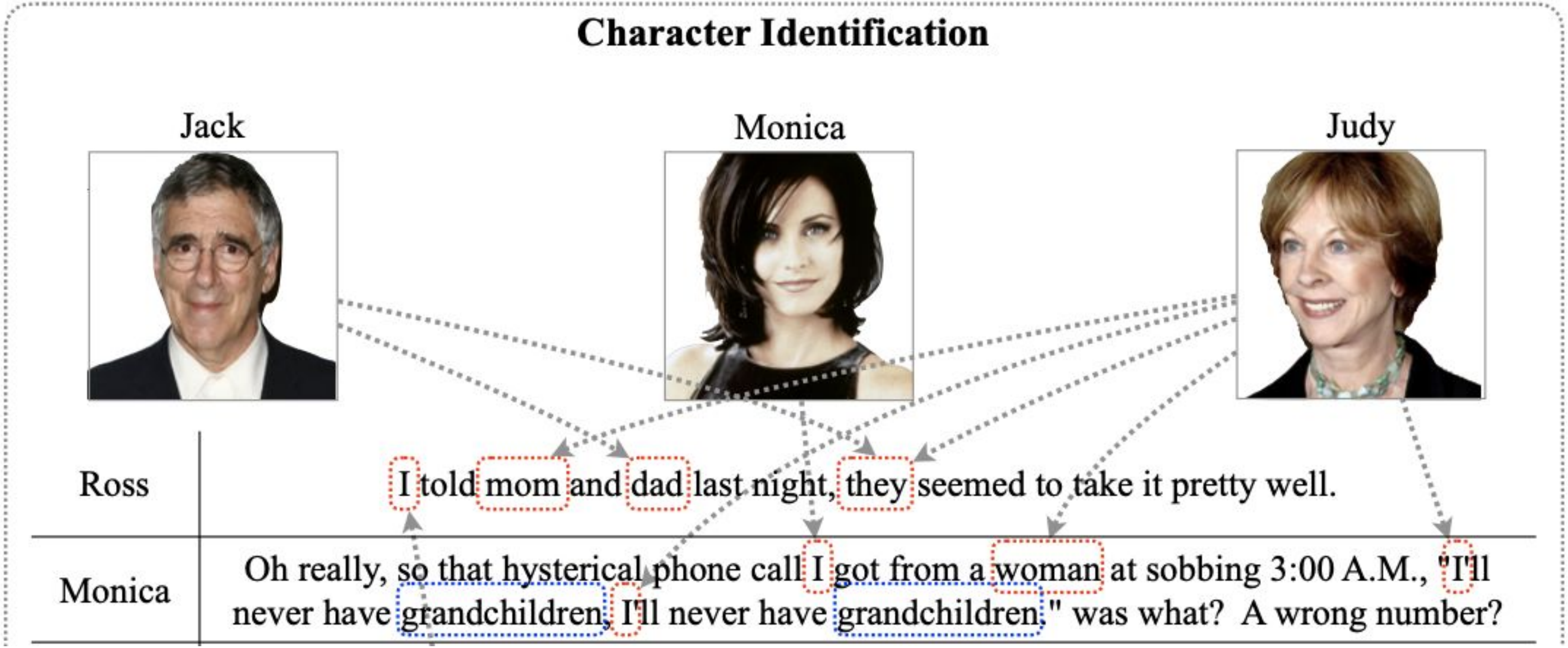
Christensen, Heidi et al. “The CHiME corpus: a resource and a challenge for Computational Hearing in Multisource Environments”

Multiparty Conversation Character Identification



Chen, Henry and Jinho D. Choi. "Character Identification on Multiparty Conversation: Identifying Mentions of Characters in TV Shows"

Multiparty Conversation Character Identification



Chen, Henry and Jinho D. Choi. “Character Identification on Multiparty Conversation: Identifying Mentions of Characters in TV Shows”

Multiparty Conversation Character Identification

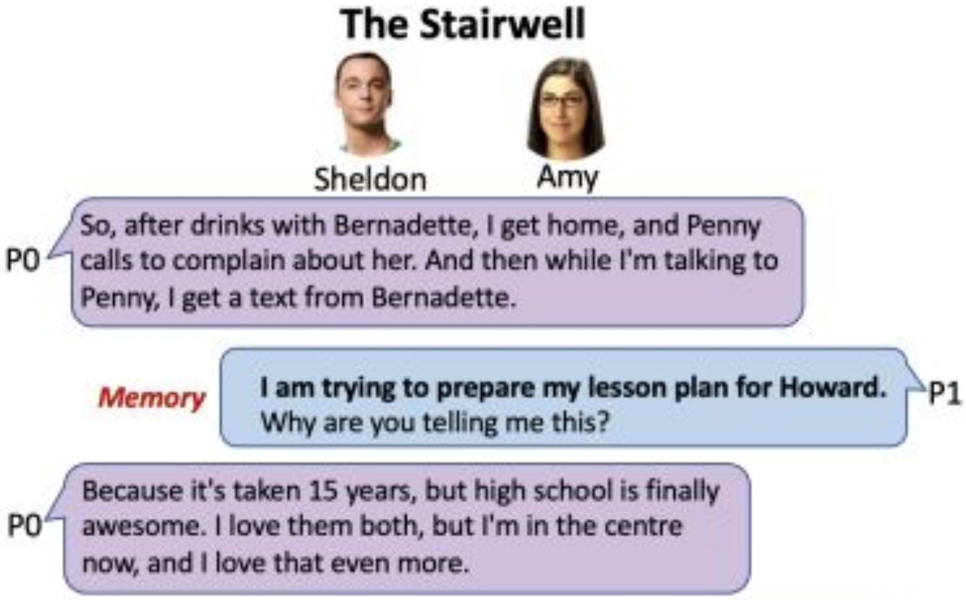
TRN	TST	Document: episode					Document: scene				
		FC	EC	UC	UM	Purity	FC	EC	UC	UM	Purity
Stanford multi-pass sieve		46	53	38.64	16.33	45.97	38	60	22.15	5.97	**64.01**
Stanford entity-centric		36	60	32.59	8.41	38.78	26	60	8.85	1.49	44.12
F1	F1	19	30	30.23	4.20	61.13	21	30	4.94	1.35	54.11
	F2	12	24	40.00	3.15	42.13	17	24	17.91	4.86	51.58
	B1	9	14	0.00	0.00	75.99	14	14	6.25	1.90	70.10
F1+F2	F1	20	30	39.39	7.52	69.92	20	30	10.11	2.72	56.28
	F2	18	24	49.06	8.25	62.54	23	24	7.46	2.12	57.64
	B1	12	14	51.52	12.69	72.16	14	14	10.87	4.56	67.11
	F1+F2	30	46	42.24	7.54	66.65	26	46	9.26	1.83	45.11
	F1+F2+B1	39	60	44.22	8.44	67.67	30	60	7.76	1.35	41.79
B1	B1	11	14	25.00	1.90	80.08	12	14	14.00	5.47	72.83
F1+F2+B1	F1	25	30	21.67	4.06	73.21	20	30	9.41	3.15	51.74
	F2	25	24	29.17	3.64	64.62	25	24	5.80	1.34	58.79
	B1	9	14	20.00	1.31	71.29	15	14	6.67	1.33	69.45
	F1+F2	39	46	24.76	3.78	69.60	29	46	7.62	1.74	44.49
	F1+F2+B1	45	60	23.93	3.27	**69.21**	36	60	6.84	1.39	42.81

Table 9: Character identification results on our corpus using cluster remapping on the coreference resolution system results. FC: found clusters after remapping. EC: expected clusters from gold. UC: percentage of unknown clusters after remapping. UM: percentage of unknown mentions in the unknown clusters to all the mentions.

- Character ID == coreference resolution
- Episode-level context > scene-level context

Chen, Henry and Jinho D. Choi. “Character Identification on Multiparty Conversation: Identifying Mentions of Characters in TV Shows”

TVSHOWGUESS: Character Comprehension



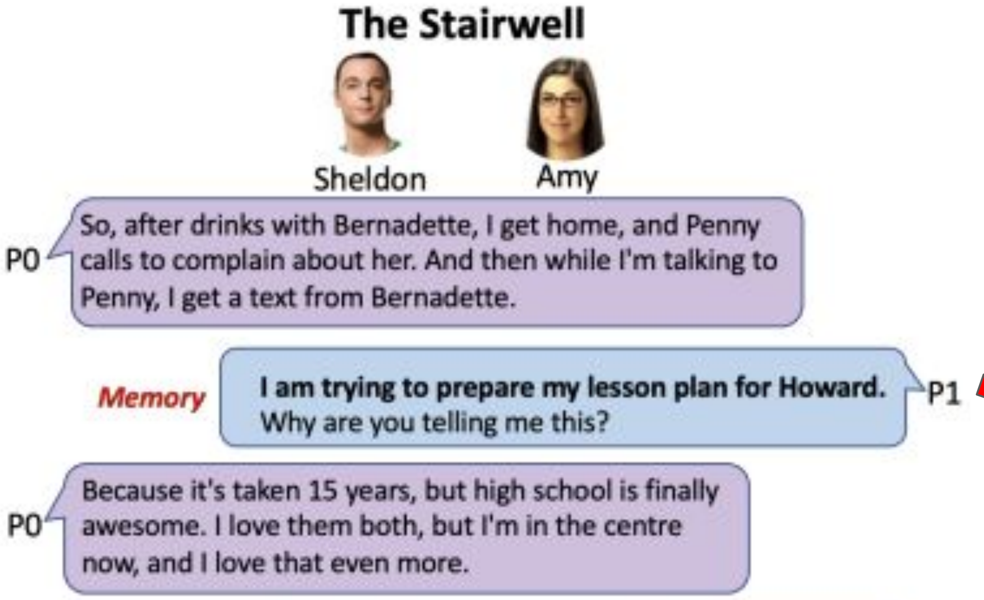
	Evidence Type	Friends(%)	TBBT(%)
(a)	Ling. Style	0.66	9.93
	Personality	7.28	21.85
	Fact	20.53	33.12
	(Attribute)	2.65	8.61
	(Relation)	16.56	22.52
	(Status)	1.32	1.99
	Memory	36.42	27.15
	Inside-Background	33.11	12.58
	Inside-Mention	15.23	15.23
	Exclusion	8.61	22.52

	Dependence of Hist.	Friends(%)	TBBT(%)
(b)	No Dep.	53.64	32.45
	Direct Dep.	26.49	36.42
	Indirect Dep.	19.87	31.13

Who is P0 and Who is P1?

Sang, Yisi et al. “TVSHOWGUESS: Character Comprehension in Stories as Speaker Guessing”

TVSHOWGUESS: Character Comprehension



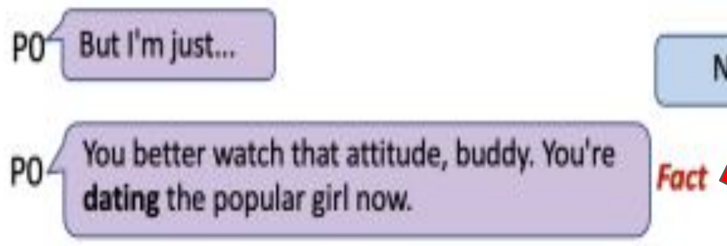
	Evidence Type	Friends(%)	TBBT(%)
(a)	Ling. Style	0.66	9.93
	Personality	7.28	21.85
	Fact	20.53	33.12
	(Attribute)	2.65	8.61
	(Relation)	16.56	22.52
	(Status)	1.32	1.99
	Memory	36.42	27.15
	Inside-Background	33.11	12.58
	Inside-Mention	15.23	15.23
	Exclusion	8.61	22.52

	Dependence of Hist.	Friends(%)	TBBT(%)
(b)	No Dep.	53.64	32.45
	Direct Dep.	26.49	36.42
	Indirect Dep.	19.87	31.13

Memory from past interaction; P1 is Sheldon

Sang, Yisi et al. “TVSHOWGUESS: Character Comprehension in Stories as Speaker Guessing”

TVSHOWGUESS: Character Comprehension



	Evidence Type	Friends(%)	TBBT(%)
(a)	Ling. Style	0.66	9.93
	Personality	7.28	21.85
	Fact	20.53	33.12
	(Attribute)	2.65	8.61
	(Relation)	16.56	22.52
	(Status)	1.32	1.99
	Memory	36.42	27.15
	Inside-Background	33.11	12.58
	Inside-Mention	15.23	15.23
	Exclusion	8.61	22.52

	Dependence of Hist.	Friends(%)	TBBT(%)
(b)	No Dep.	53.64	32.45
	Direct Dep.	26.49	36.42
	Indirect Dep.	19.87	31.13

Fact about social relationship; Amy is dating Sheldon

Sang, Yisi et al. "TVSHOWGUESS: Character Comprehension in Stories as Speaker Guessing"

Models for Character Comprehension in TVSHOWGUESS

System	FRIENDS		TBBT		Frasier		Gilmore_Girls		The_Office		Overall	
	dev	test	dev	test	dev	test	dev	test	dev	test	dev	test
Random	35.23	31.59	33.08	37.79	34.74	31.61	36.43	38.90	44.30	46.71	36.79	36.59
Vanilla Longformer	67.79	60.63	61.58	63.95	85.11	82.06	79.84	74.52	70.92	71.60	72.55	69.72
repl with BERT	65.60	59.58	61.58	58.43	85.11	84.30	81.91	70.41	67.56	68.54	71.65	67.76
Our MR. BERT	**77.01**	**73.20**	62.60	62.50	90.07	82.51	**83.98**	**78.63**	70.92	74.41	76.82	74.52
- context	62.92	57.19	59.54	63.95	81.64	76.23	74.42	67.12	66.00	67.37	68.33	65.54
- reverse trick	70.81	68.71	52.42	59.01	79.40	81.39	78.04	73.97	66.22	68.31	69.45	70.52
- fill-empty trick	74.33	68.56	58.27	63.37	86.10	78.48	72.87	69.86	68.90	73.71	72.28	70.92
Our Longformer-P	**77.01**	69.91	**63.87**	**66.57**	**90.32**	**87.67**	82.17	75.07	**71.81**	**76.29**	**76.95**	**74.97**
maxlen=1000	74.16	66.77	63.36	64.24	86.10	85.65	79.33	72.05	73.83	76.06	75.25	72.74
repl with BERT	68.12	58.83	61.32	63.95	82.63	76.91	68.48	65.75	72.48	71.83	70.49	66.79
Human*	98.68	–	89.82	–	–	–	–	–	–	–		

Gap from human-level performance

Sang, Yisi et al. "TVSHOWGUESS: Character Comprehension in Stories as Speaker Guessing"

Models for Character Comprehension in TVSHOWGUESS

System	FRIENDS		TBBT		Frasier		Gilmore_Girls		The_Office		Overall	
	dev	test	dev	test	dev	test	dev	test	dev	test	dev	test
Random	35.23	31.59	33.08	37.79	34.74	31.61	36.43	38.90	44.30	46.71	36.79	36.59
Vanilla Longformer	67.79	60.63	61.58	63.95	85.11	82.06	79.84	74.52	70.92	71.60	72.55	69.72
repl with BERT	65.60	59.58	61.58	58.43	85.11	84.30	81.91	70.41	67.56	68.54	71.65	67.76
Our MR. BERT	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	70.92	74.41	76.82	74.52
[illegible]ntext	62.92	57.19	59.54	63.95	81.64	76.23	74.42	67.12	66.00	67.37	68.33	65.54
- reverse trick	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	78.04	73.97	66.22	68.31	69.45	70.52
- fill empty trick	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	72.87	69.86	68.90	73.71	72.28	70.92
Our Longformer-P	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	**71.81**	**76.29**	**76.95**	**74.97**
maxlen=1000	74.16	66.77	63.36	64.24	86.10	85.65	79.33	72.05	73.83	76.06	75.25	72.74
repl with BERT	68.12	58.83	61.32	63.95	82.63	76.91	68.48	65.75	72.48	71.83	70.49	66.79
Human*	98.68	–	89.82	–	–	–	–	–	–	–		

Pros:

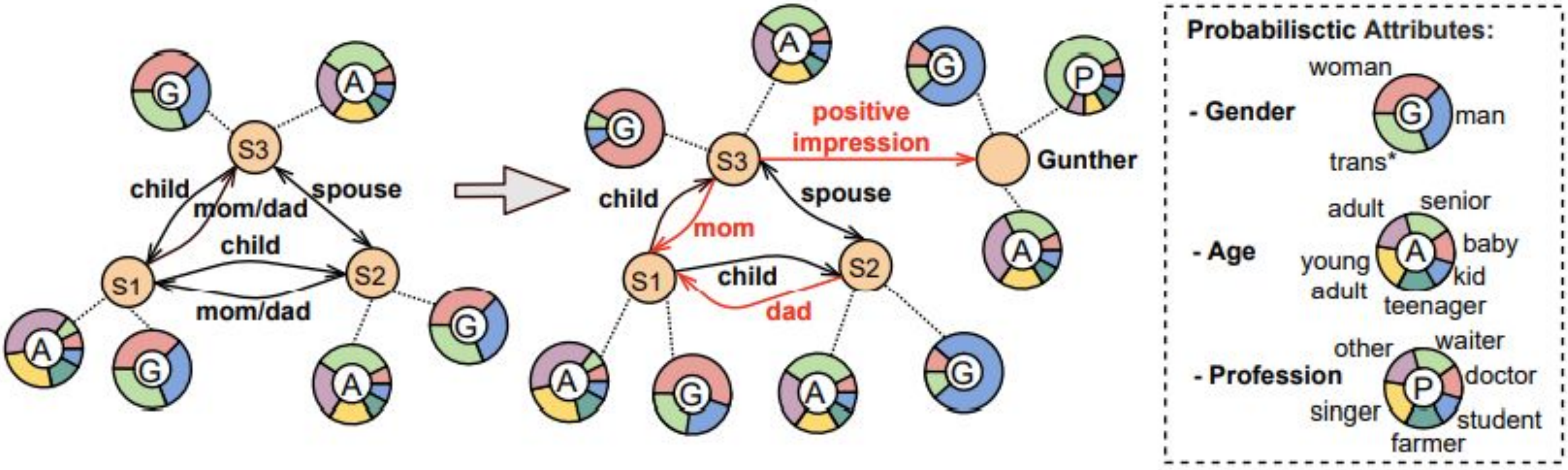
- Useful task/skill for agents: persona-specific annotations

Cons:

- Not realistic: TBBT is highly scripted for satire
- Does not yet incorporate annotations into solving task

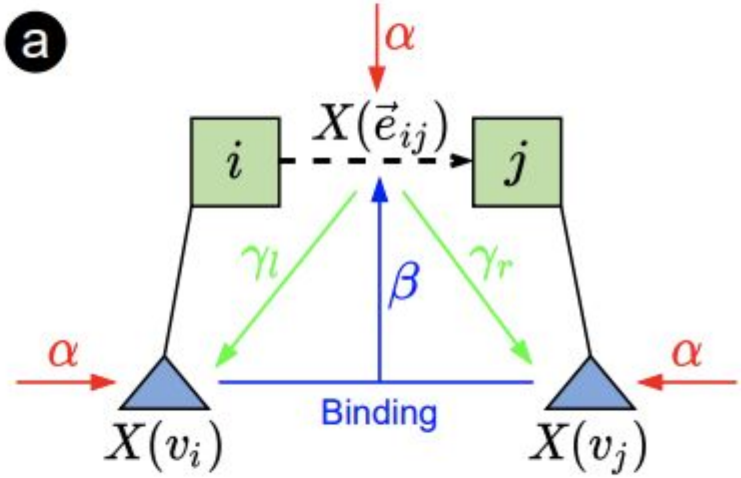
Sang, Yisi et al. "TVSHOWGUESS: Character Comprehension in Stories as Speaker Guessing"

Social Relationship Inference in Dialogues



Qiu, Liang et al. “SocAoG: Incremental Graph Parsing for Social Relation Inference in Dialogues”

Social Relationship Inference in Dialogues



- ***α***: infer social relationship between i,j; infer attributes of persons i,j
- ***β***: re-estimate social relationships conditioned on inferred attributes
- ***γ***: re-infer attributes conditioned on social relationships

Qiu, Liang et al. “SocAoG: Incremental Graph Parsing for Social Relation Inference in Dialogues”

Evaluating SocAoG

Methods	DialogRE (V2) Dev F1(σ)	DialogRE (V2) Dev F1$_c$(σ)	DialogRE (V2) Test F1(σ)	DialogRE (V2) Test F1$_c$(σ)	MovieGraph Dev F1(σ)	MovieGraph Test F1(σ)
BERT (Devlin et al., 2018)	59.4 (0.7)	54.7 (0.8)	57.9 (1.0)	53.1 (0.7)	50.6 (1.2)	53.6 (0.3)
BERT$_S$ (Yu et al., 2020)	62.2 (1.3)	57.0 (1.0)	59.5 (2.1)	54.2 (1.4)	50.7 (1.1)	53.6 (0.4)
GDPNet (Xue et al., 2020b)	67.1 (1.0)	61.5 (0.8)	64.3 (1.1)	60.1 (0.9)	53.1 (1.1)	56.4 (0.8)
SimpleRE (Xue et al., 2020a)	68.2 (1.1)	63.4 (0.6)	66.7 (0.7)	63.3 (0.9)	55.2 (0.5)	58.1 (0.7)
SocAoG$_{reduced}$ (our method)	69.1 (0.4)	65.7 (0.5)	68.6 (0.9)	65.4 (1.1)	**60.7 (0.4)**	63.2 (0.3)
SocAoG (our method)	**69.5 (0.8)**	**66.1 (0.7)**	**69.1 (0.5)**	**66.5 (0.8)**	60.1 (0.6)	**64.1 (0.8)**

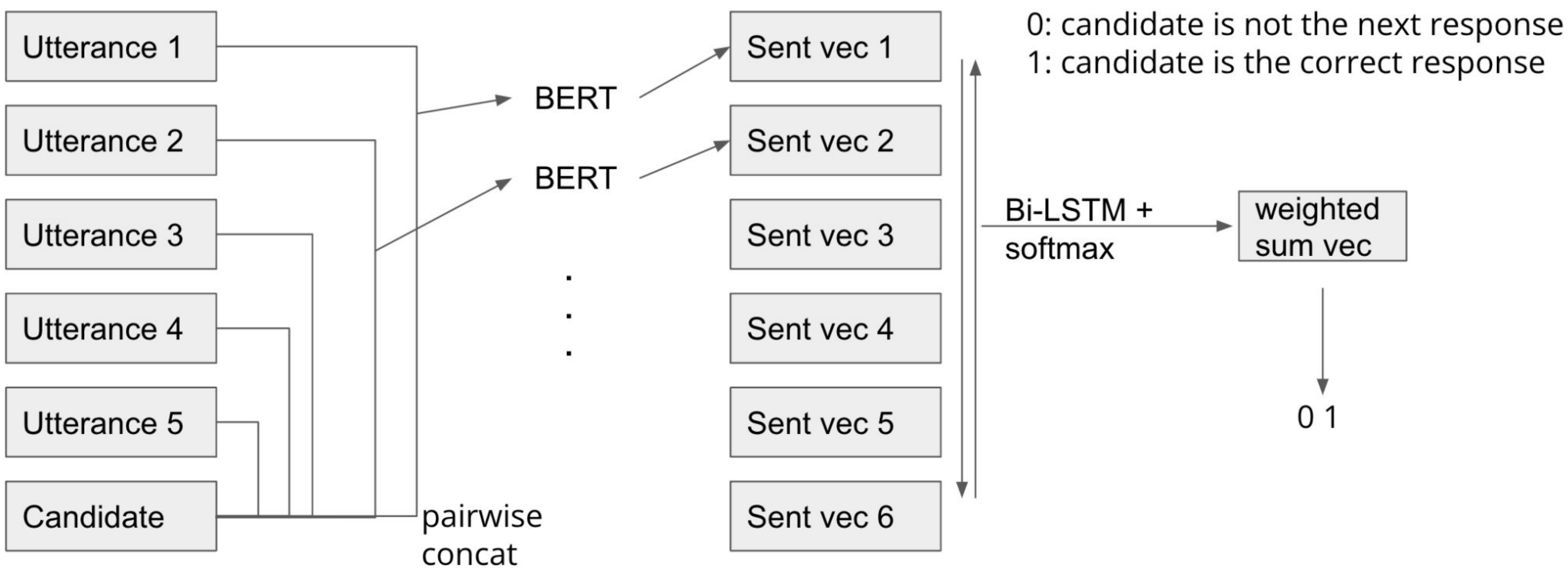
Current SotA

- Learning complex personas and relationships in just one step is difficult
- Persona attributes affect relations, and relations affect persona attributes

Qiu, Liang et al. "SocAoG: Incremental Graph Parsing for Social Relation Inference in Dialogues"

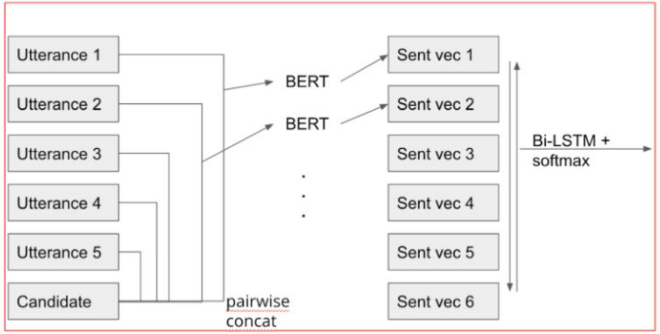
Zero-Shot Dialogue Disentanglement

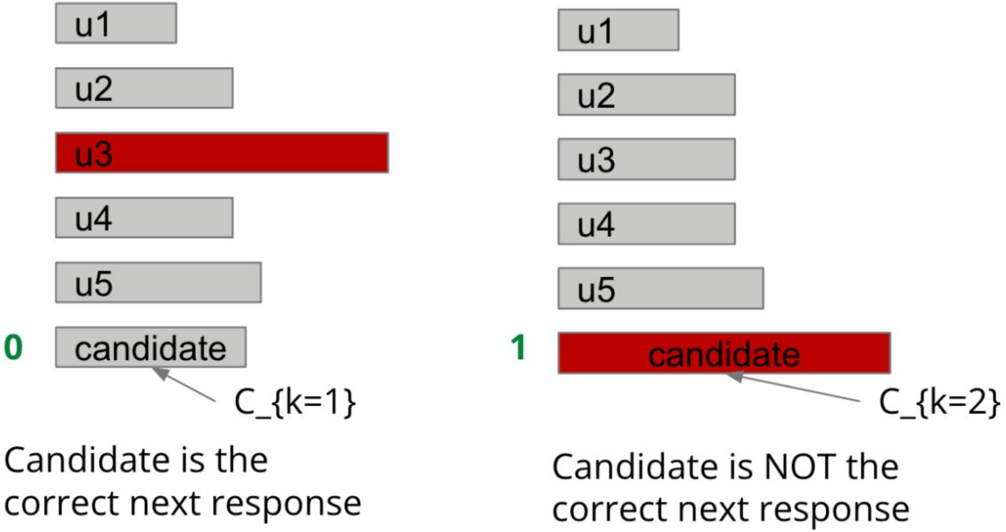
Pretraining on entangled response selection



Chi, Ta-Chung and Alex Rudnicky. "Zero-Shot Dialogue Disentanglement by Self-Supervised Entangled Response Selection"

Zero-Shot Dialogue Disentanglement

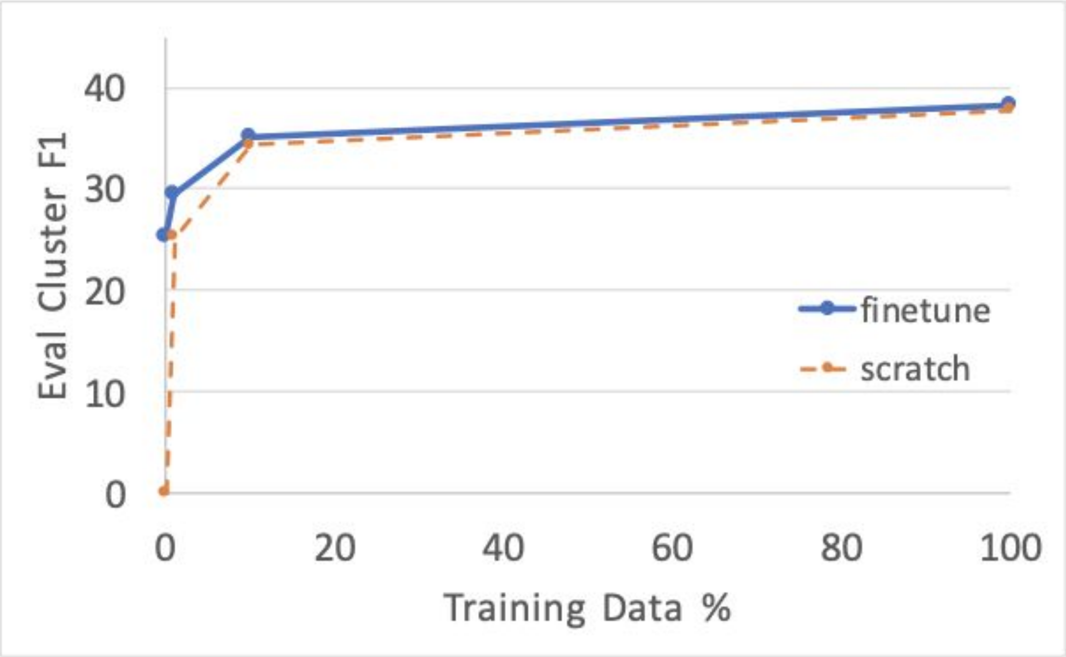




$$L_{joint} = (1 - w) * L_{res} + w * L_{attn}$$

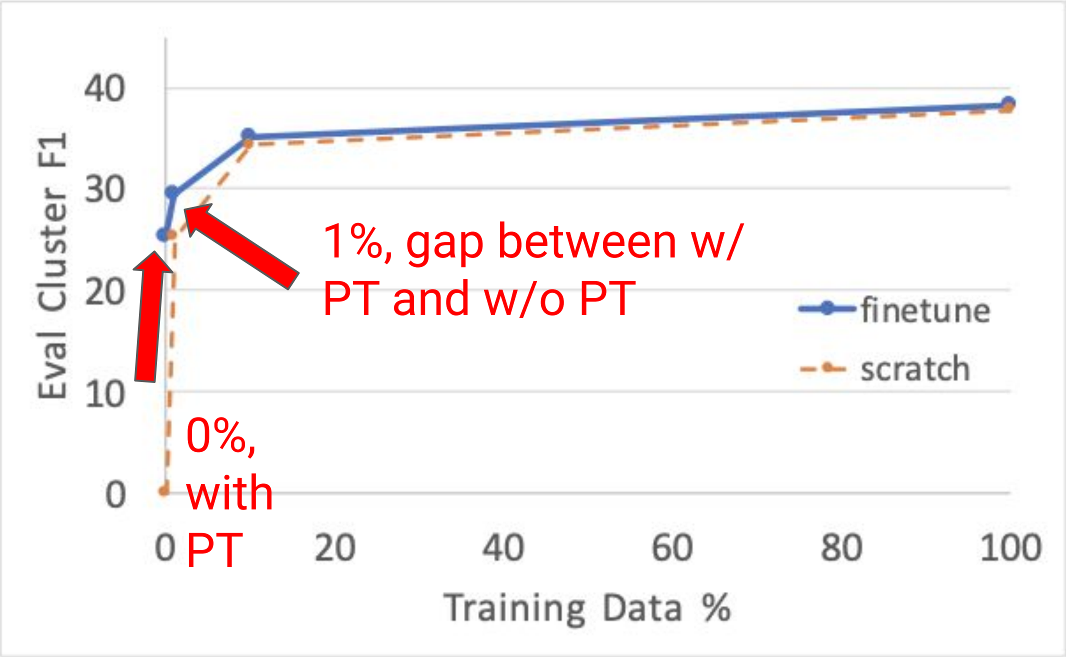
Chi, Ta-Chung and Alex Rudnicky. “Zero-Shot Dialogue Disentanglement by Self-Supervised Entangled Response Selection”

Zero-Shot Dialogue Disentanglement



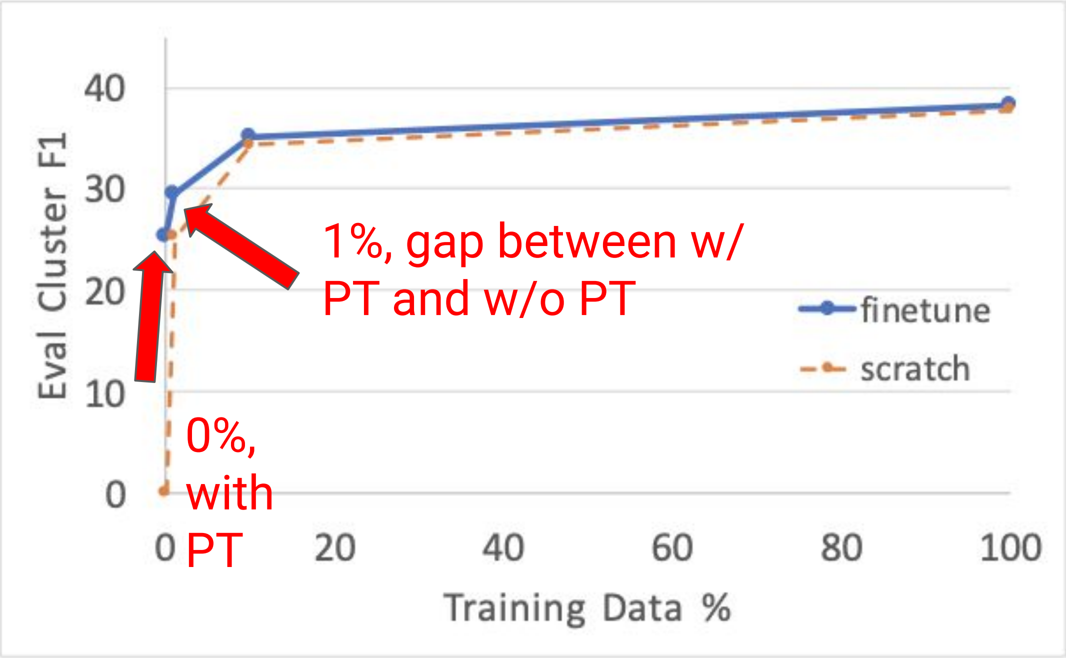
Chi, Ta-Chung and Alex Rudnicky. “Zero-Shot Dialogue Disentanglement by Self-Supervised Entangled Response Selection”

Zero-Shot Dialogue Disentanglement



Chi, Ta-Chung and Alex Rudnicky. “Zero-Shot Dialogue Disentanglement by Self-Supervised Entangled Response Selection”

Zero-Shot Dialogue Disentanglement



Pros:

- First to ever consider zero-shot disentanglement
- practically useful

Cons:

- Not clear how well results will translate from Ubuntu IRC

Chi, Ta-Chung and Alex Rudnicky. “Zero-Shot Dialogue Disentanglement by Self-Supervised Entangled Response Selection”

Speaker-Aware Discourse Parsing (via Second Stage Pre-training)

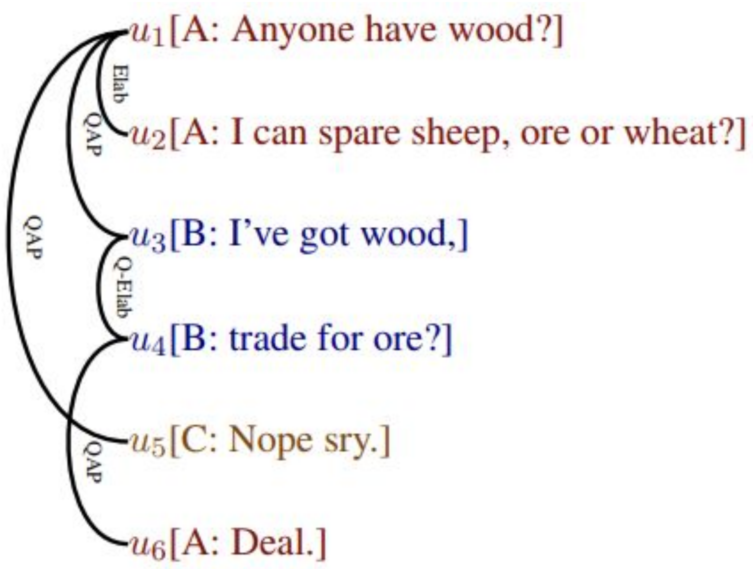


Figure 1: An example of a discourse dependency tree. u_1, u_2, u_3, u_4, u_5 refer to EDUs. "Q-Elab", "QAP", "Q-Elab", and "Elab" refer to discourse relations. "A", "B", and "C" are three speakers.

Yu, Nan et al. "Speaker-Aware Discourse Parsing on Multi-Party Dialogues"

Speaker-Aware Discourse Parsing (via Second Stage Pre-training)

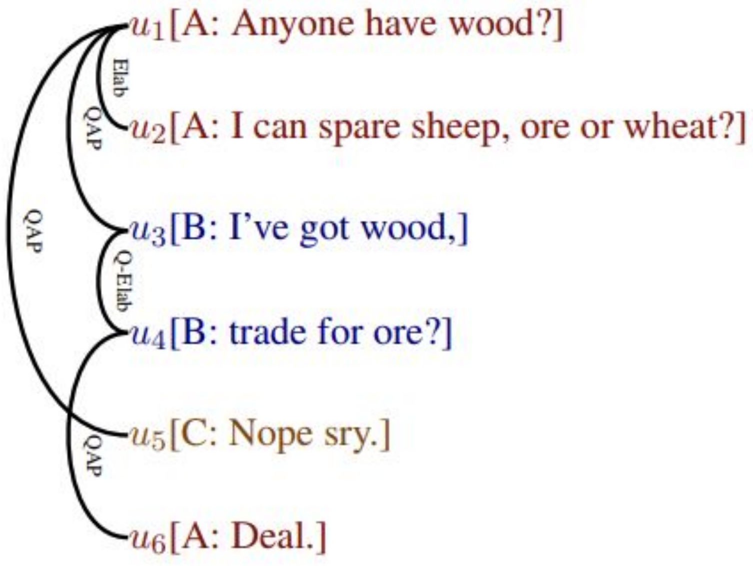
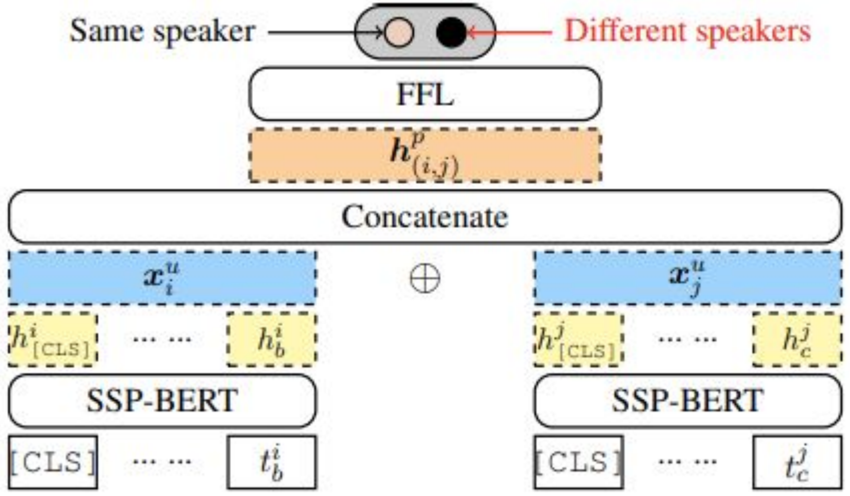


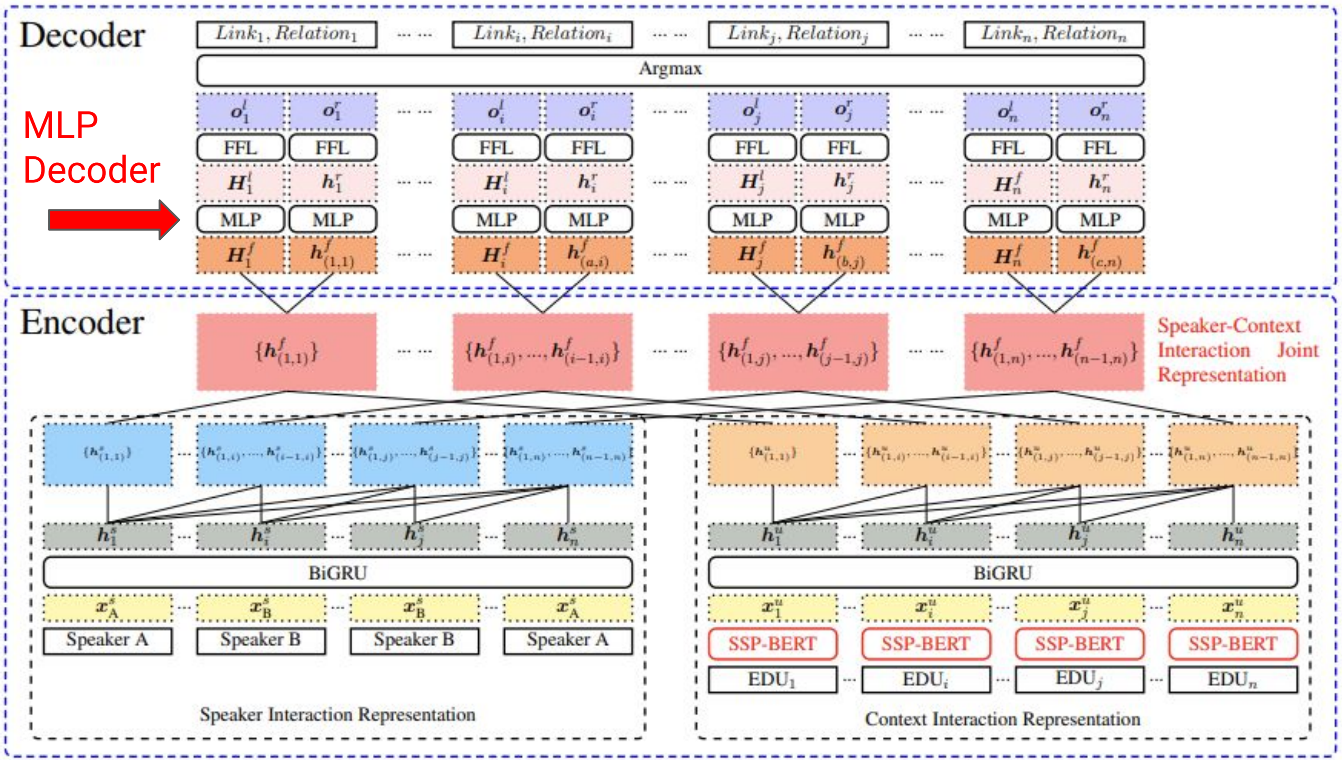
Figure 1: An example of a discourse dependency tree. u_1, u_2, u_3, u_4, u_5 refer to EDUs. "Q-Elab", "QAP", "Q-Elab", and "Elab" refer to discourse relations. "A", "B", and "C" are three speakers.

Second Stage: discern whether utterances come from the same speaker or not



Yu, Nan et al. "Speaker-Aware Discourse Parsing on Multi-Party Dialogues"

Speaker-Aware Discourse Parsing



MLP Decoder

Speaker Interactions (List of Speaker IDs)

Dialogue Context Encoder

Yu, Nan et al. "Speaker-Aware Discourse Parsing on Multi-Party Dialogues"

Speaker-Aware Discourse Parsing

Molweni: Difficult QA dataset with Discourse Structures

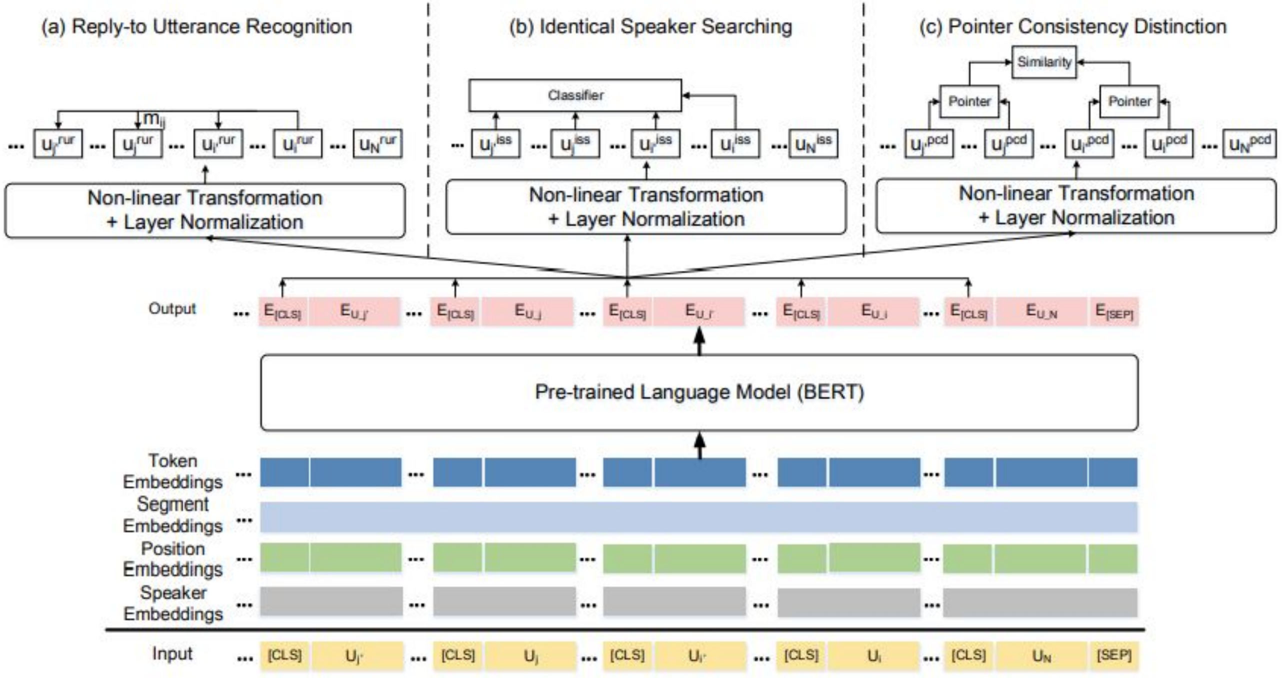
Takeaway:

- speaker interactions are important
- who-says-what affects understanding of discourse structure

Models	Link	Link&Rel
Molweni		
Li et al. (2020)	78.1	54.8
Wang et al. (2021)	81.6	58.5
Liu and Chen (2021)	80.2	56.9
He et al. (2021)*	80.0	57.0
BERT	77.8	56.5
SSP-BERT + SCIJE	**83.7**	**59.4**
STAC		
Shi and Huang (2019)	73.2	55.7
Wang et al. (2021)	73.5	57.3
Yang et al. (2021)	74.1	57.0
Liu and Chen (2021)	**75.5**	57.2
BERT	72.4	55.4
SSP-BERT + SCIJE	73.0	**57.4**

Yu, Nan et al. "Speaker-Aware Discourse Parsing on Multi-Party Dialogues"

MPC-BERT: Pre-Training for Multi-Party Conversation Understanding

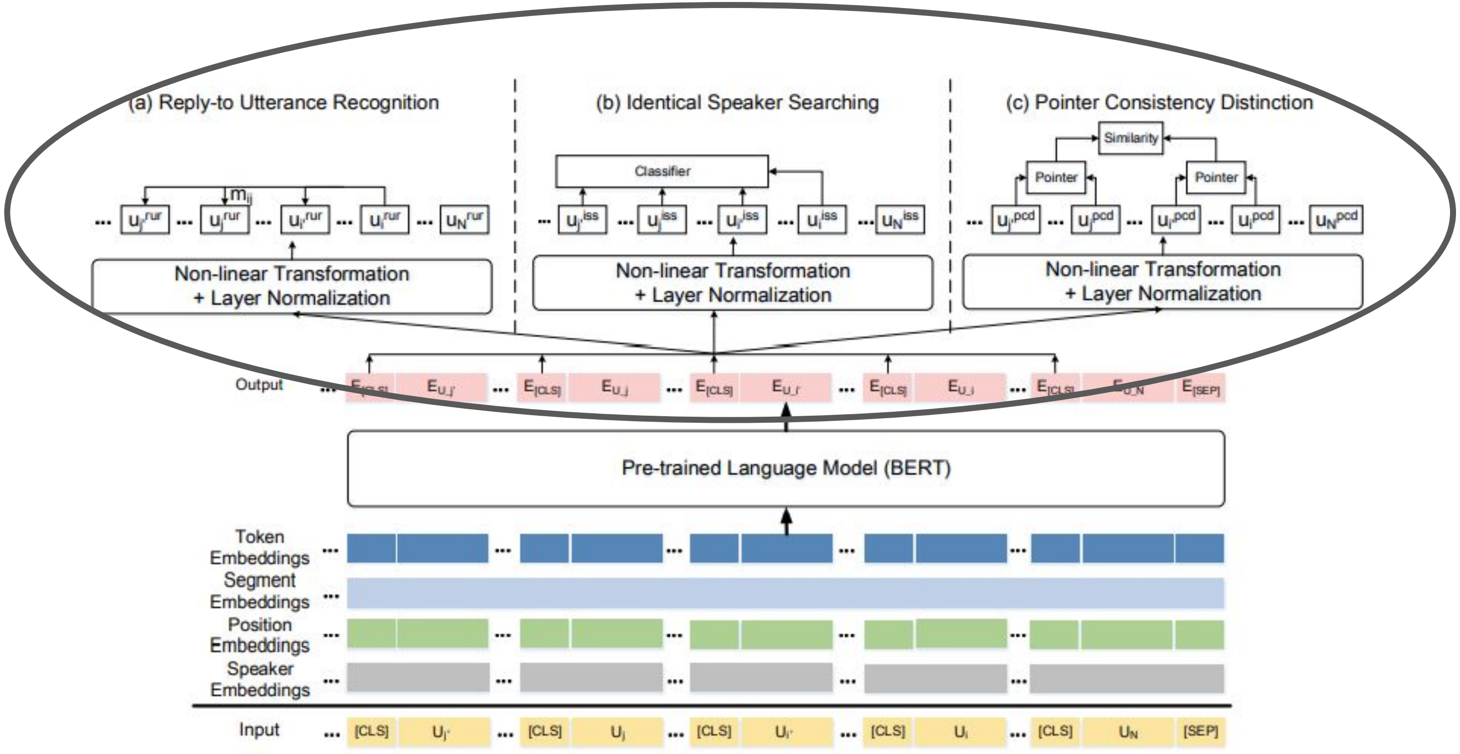


Gu, Jia-Chen et al. "MPC-BERT: A Pre-Trained Language Model for Multi-Party Conversation Understanding"

MPC-BERT: Pre-Training for Multi-Party Conversation Understanding

Interlocutor Structure:
1. Reply-to Utterance Recognition
2. Identical Speaker Searching
3. Pointer Consistency Detection

Utterance Semantics:
1. Masked Shared Utterance Restoration
2. Shared Node Detection



Gu, Jia-Chen et al. "MPC-BERT: A Pre-Trained Language Model for Multi-Party Conversation Understanding"

MPC-BERT: Pre-Training for Multi-Party Conversation Understanding

Ubuntu IRC Addressee Recognition

	Hu et al. (2019)		Ouchi and Tsuboi (2016)					
			Len-5		Len-10		Len-15	
	P@1	Acc.	P@1	Acc.	P@1	Acc.	P@1	Acc.
Preceding (Le et al., 2019)	-	-	63.50	40.46	56.84	21.06	54.97	13.08
Subsequent (Le et al., 2019)	-	-	61.03	40.25	54.57	20.26	53.07	12.79
DRNN (Ouchi and Tsuboi, 2016)	-	-	72.75	58.18	65.58	34.47	62.60	22.58
SIRNN (Zhang et al., 2018a)	-	-	75.98	62.06	70.88	40.66	68.13	28.05
W2W (Le et al., 2019)	-	-	77.55	63.81	73.52	44.14	73.42	34.23
BERT (Devlin et al., 2019)	96.16	83.50	85.95	75.99	83.41	58.22	81.09	44.94
SA-BERT (Gu et al., 2020)	97.12	88.91	86.81	77.45	84.46	60.30	82.84	47.23
MPC-BERT	**98.31**	**92.42**	**88.73**	**80.31**	**86.23**	**63.58**	**85.55**	**52.59**
MPC-BERT w/o. RUR	97.75	89.98	87.51	78.42	85.63	62.26	84.78	50.83
MPC-BERT w/o. ISS	98.20	91.96	88.67	80.25	86.14	63.40	85.02	51.12
MPC-BERT w/o. PCD	98.20	91.90	88.51	80.06	85.92	62.84	85.21	51.17
MPC-BERT w/o. MSUR	98.08	91.32	88.70	80.26	86.21	63.46	85.28	51.23
MPC-BERT w/o. SND	98.25	92.18	88.68	80.25	86.14	63.41	85.29	51.39

All pre-training tasks are useful

Gu, Jia-Chen et al. "MPC-BERT: A Pre-Trained Language Model for Multi-Party Conversation Understanding"

Contents

How to do Dialogue Generation in Multiparty Contexts?

- How should agents adapt what they say in multiparty contexts?

How to do Dialogue Generation in Multiparty Contexts?

- How should agents adapt what they say in multiparty contexts?
    - Multiparty-aware training approaches

How to do Dialogue Generation in Multiparty Contexts?

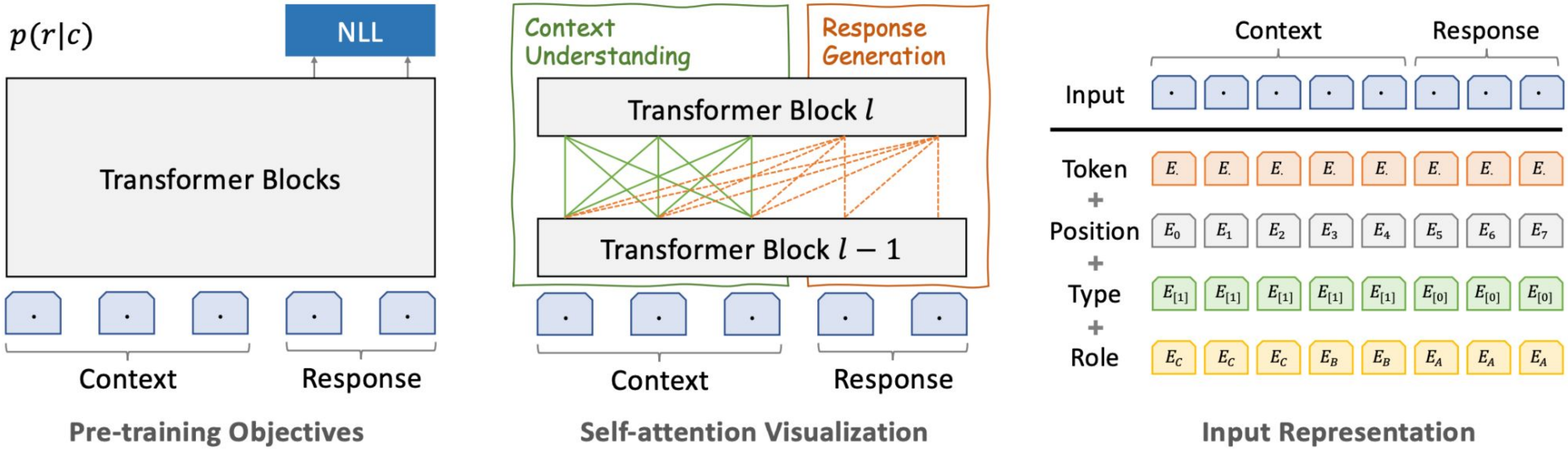
- How should agents adapt what they say in multiparty contexts?
  - Multiparty-aware training approaches
  - Approaches to make agents more sensible and empathetic

How to do Dialogue Generation in Multiparty Contexts?

- How should agents adapt what they say in multiparty contexts?
  - Multiparty-aware training approaches
  - Approaches to make agents more sensible and empathetic
  - Approaches to incorporate contextual understanding into multimodal systems

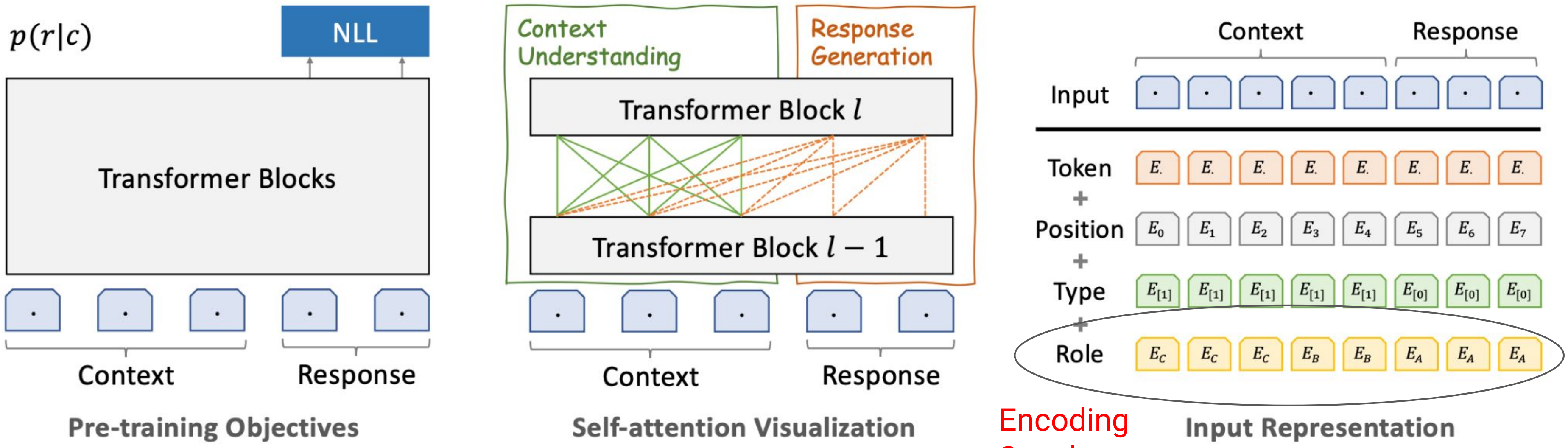
Contents

PLATO-XL: Large-Scale Pre-training for Dialogue Generation



Bao, Siqi et al. “PLATO-XL: Exploring the Large-scale Pre-training of Dialogue Generation”

PLATO-XL: Large-Scale Pre-training for Dialogue Generation



Encoding Speakers

Bao, Siqi et al. "PLATO-XL: Exploring the Large-scale Pre-training of Dialogue Generation"

PLATO-XL Performance

English Models	# Params	Coherence	Inconsistency↓	Informativeness	Hallucination↓	Engagingness
DialoGPT	345M	0.792	0.508	0.692	0.516	0.220
PLATO-2	1.6B	1.792	0.068	1.732	0.152	1.540
Blender	2.7B	1.768	0.084	1.692	0.128	1.500
PLATO-XL	11B	**1.908**	**0.024**	**1.800**	**0.024**	**1.800**

Table 1: English self-chat evaluation results, with best value written in bold.

Chinese Models	# Params	Coherence	Inconsistency↓	Informativeness	Hallucination↓	Engagingness
CDial-GPT	95M	1.188	0.104	0.908	0.388	0.460
PLATO-2	336M	1.876	0.016	1.872	0.056	1.880
ProphetNet-X	379M	1.344	0.048	1.216	0.296	0.940
EVA	2.8B	1.196	0.032	1.016	0.356	0.600
PLATO-XL	11B	**1.952**	**0.004**	**1.948**	**0.016**	**1.940**

Table 2: Chinese self-chat evaluation results, with best value written in bold.

Pros:

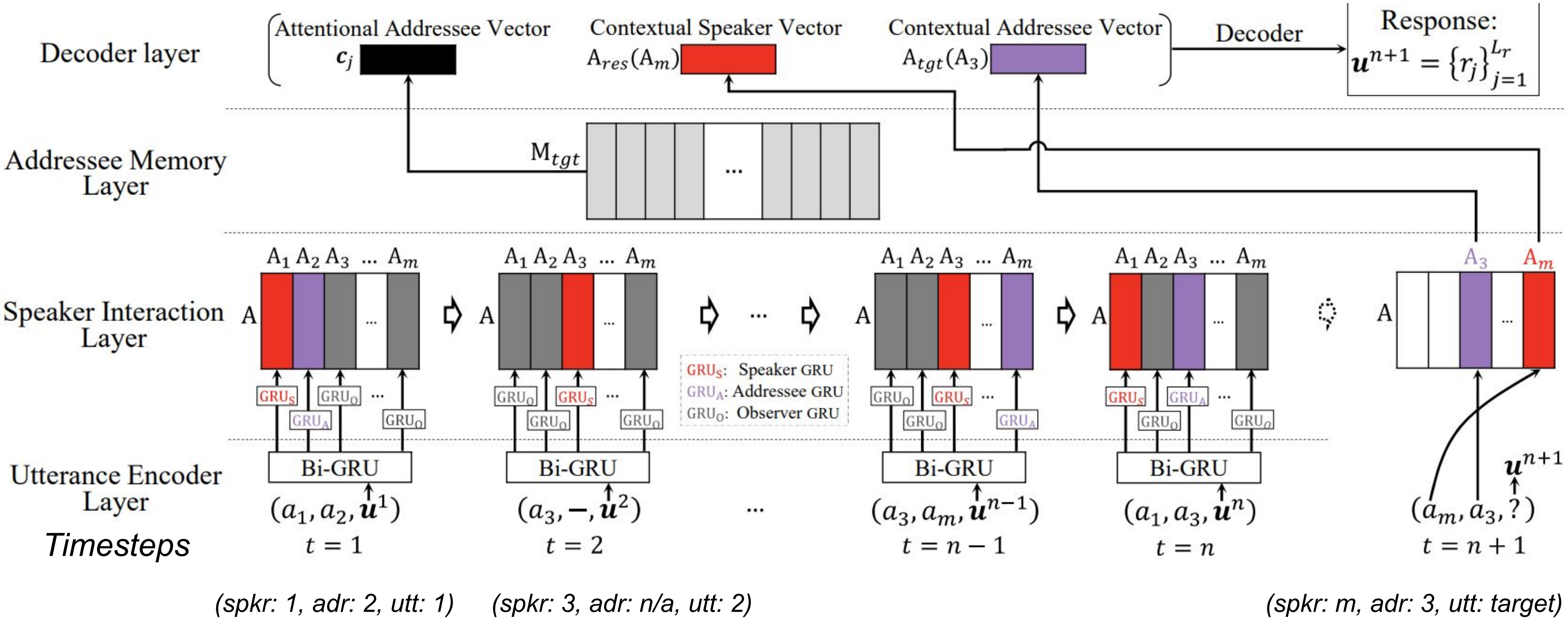
- generalizable approach with resource useful for community

Cons:

- no real multiparty generation

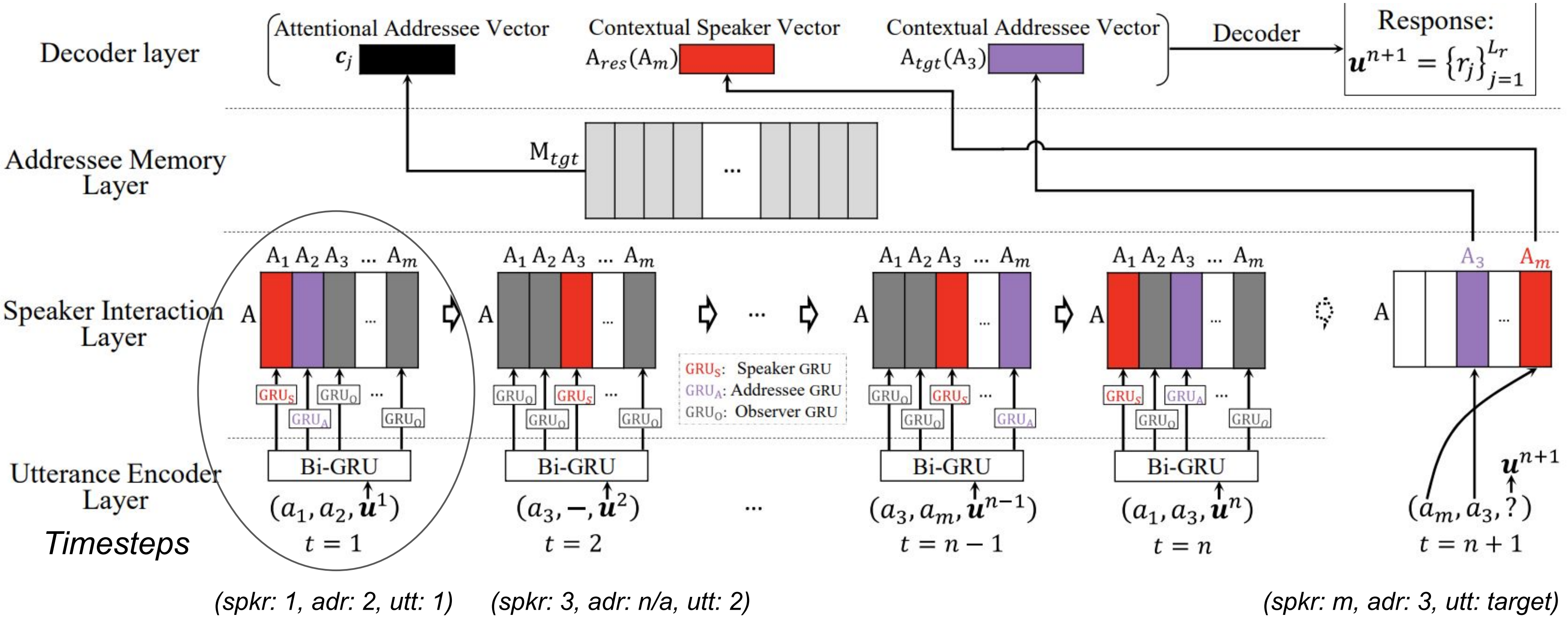
Bao, Siqi et al. “PLATO-XL: Exploring the Large-scale Pre-training of Dialogue Generation”

Incorporating Interlocutor Awareness into Multi-Party Response Generation



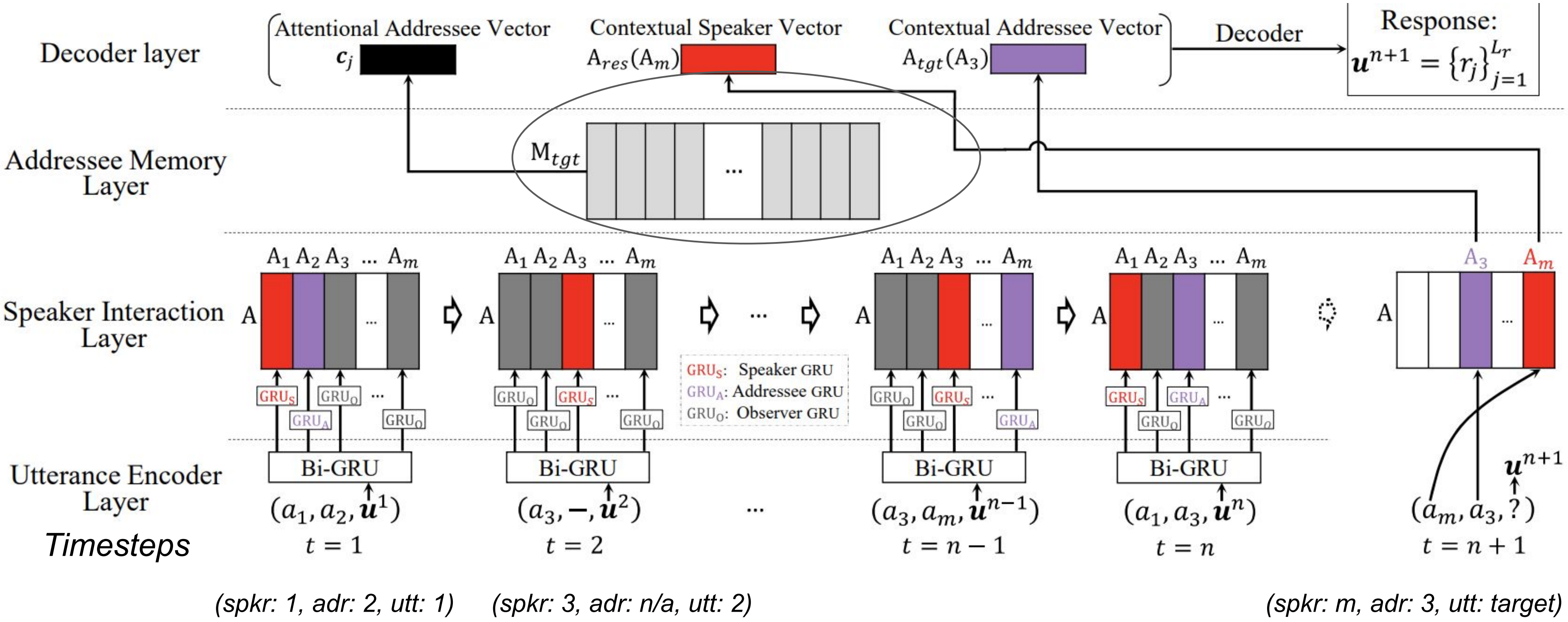
Liu, Cao et al. “Incorporating Interlocutor-Aware Context into Response Generation on Multi-Party Chatbots”

Incorporating Interlocutor Awareness into Multi-Party Response Generation



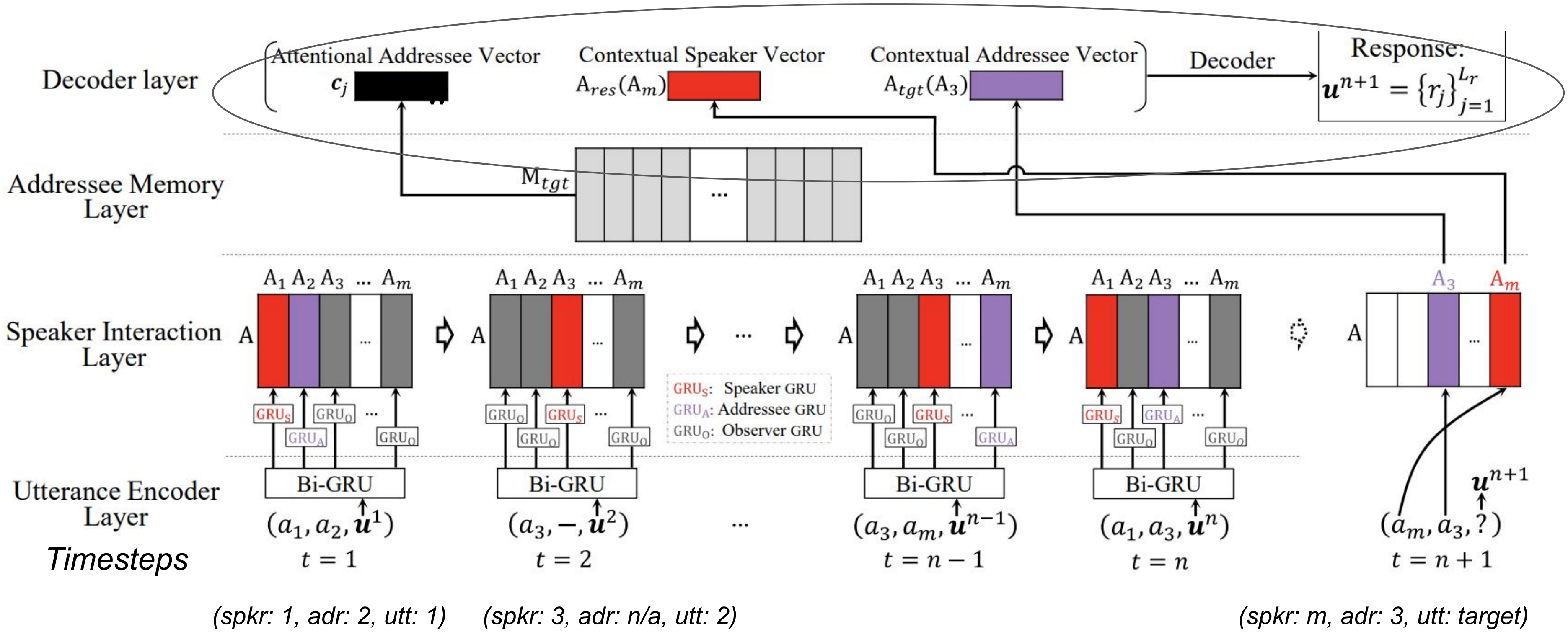
Liu, Cao et al. “Incorporating Interlocutor-Aware Context into Response Generation on Multi-Party Chatbots”

Incorporating Interlocutor Awareness into Multi-Party Response Generation



Liu, Cao et al. "Incorporating Interlocutor-Aware Context into Response Generation on Multi-Party Chatbots"

Incorporating Interlocutor Awareness into Multi-Party Response Generation



Liu, Cao et al. “Incorporating Interlocutor-Aware Context into Response Generation on Multi-Party Chatbots”

Evaluating Interlocutor-aware Context

Model	Referenced		Unreferenced	
	BLEU	ROUGE	Length	#Noun
Seq2Seq	8.86	7.62	9.48	1.24
Persona Model	9.12	7.38	11.04	1.29
VHRED	9.38	7.65	10.25	1.55
ICRED (ours)	**10.63**	**8.73**	**11.34**	**1.68**

Interlocutor-aware context yields more “correct” responses

Pros:

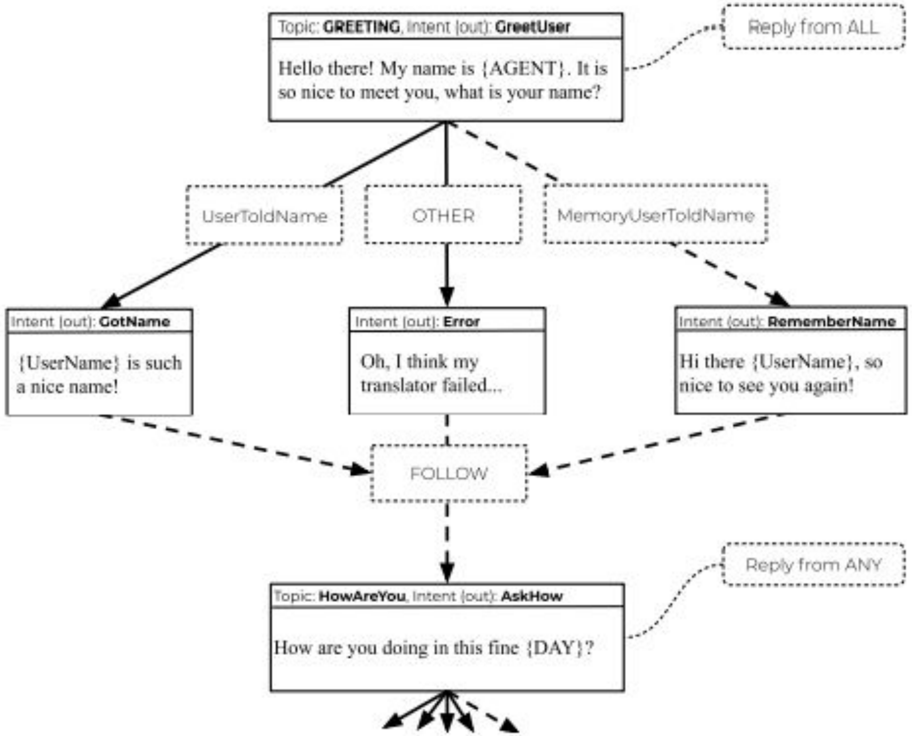
- generation in multiparty context
- transferable approach

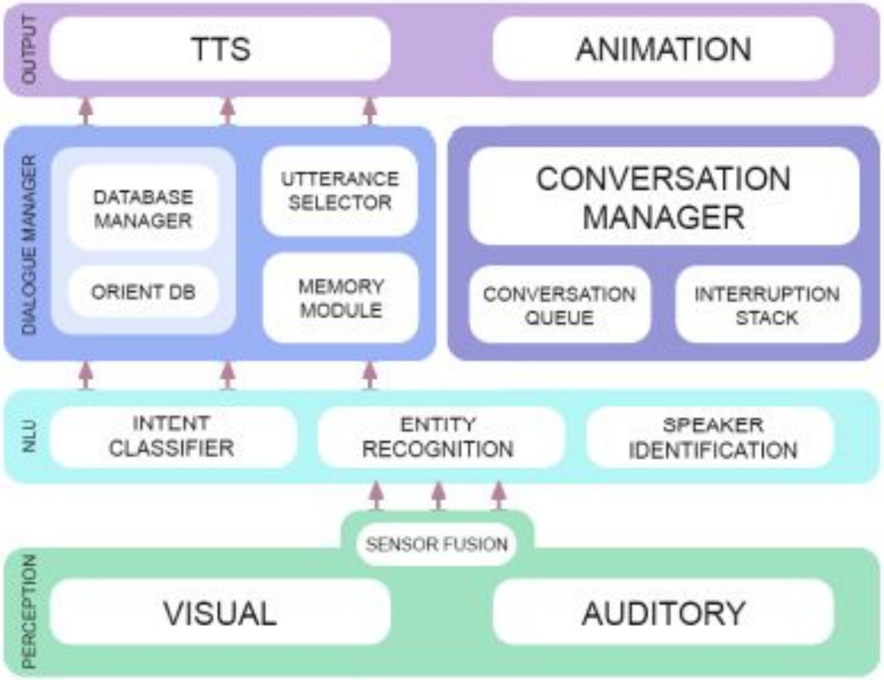
Cons:

- no human evaluation

Liu, Cao et al. “Incorporating Interlocutor-Aware Context into Response Generation on Multi-Party Chatbots”

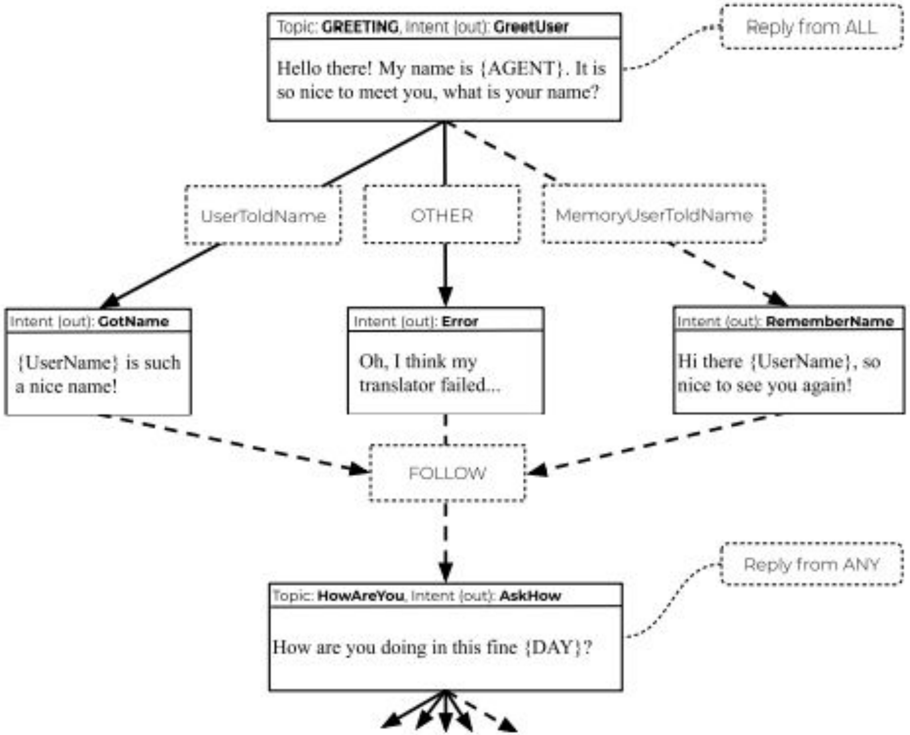
MPC Dialogue System with Concurrent Conversation Tracking and Memory

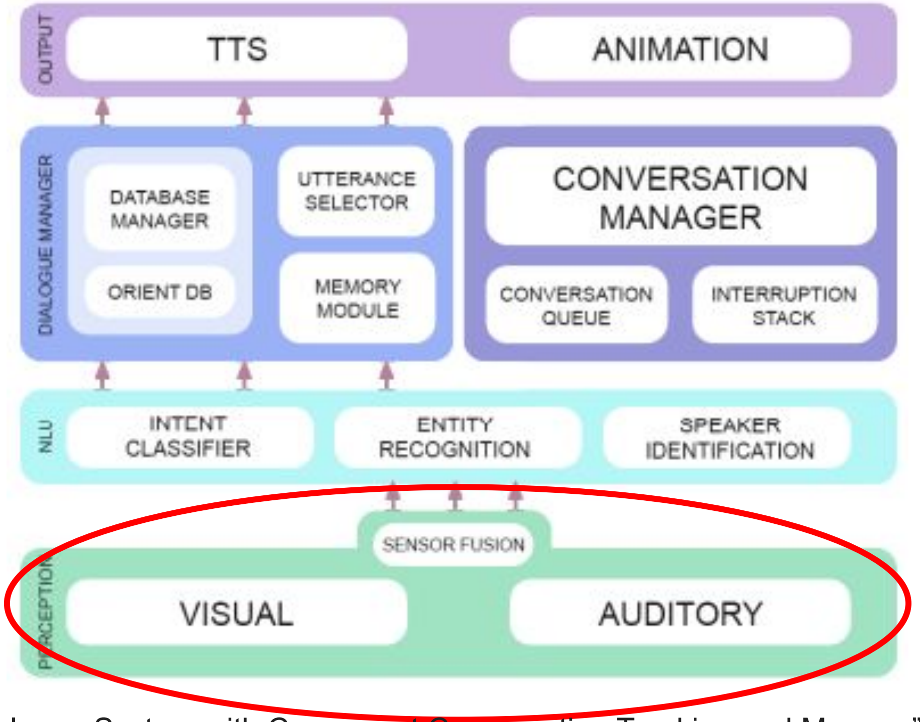




Martinez, Victor R. and James Kennedy. "A Multiparty Chat-Based Dialogue System with Concurrent Conversation Tracking and Memory"

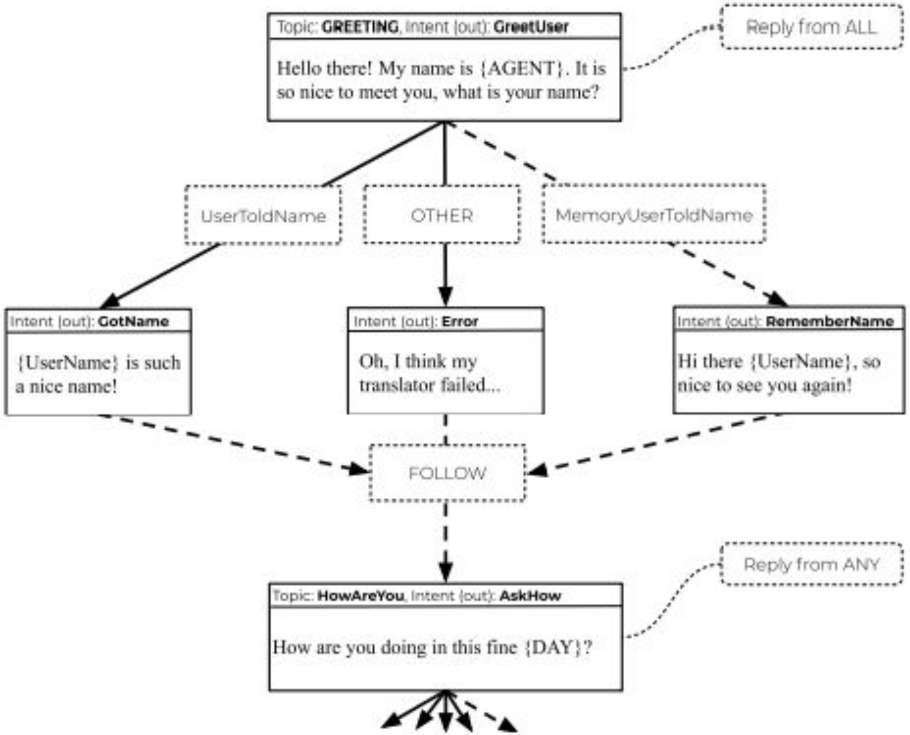
MPC Dialogue System with Concurrent Conversation Tracking and Memory

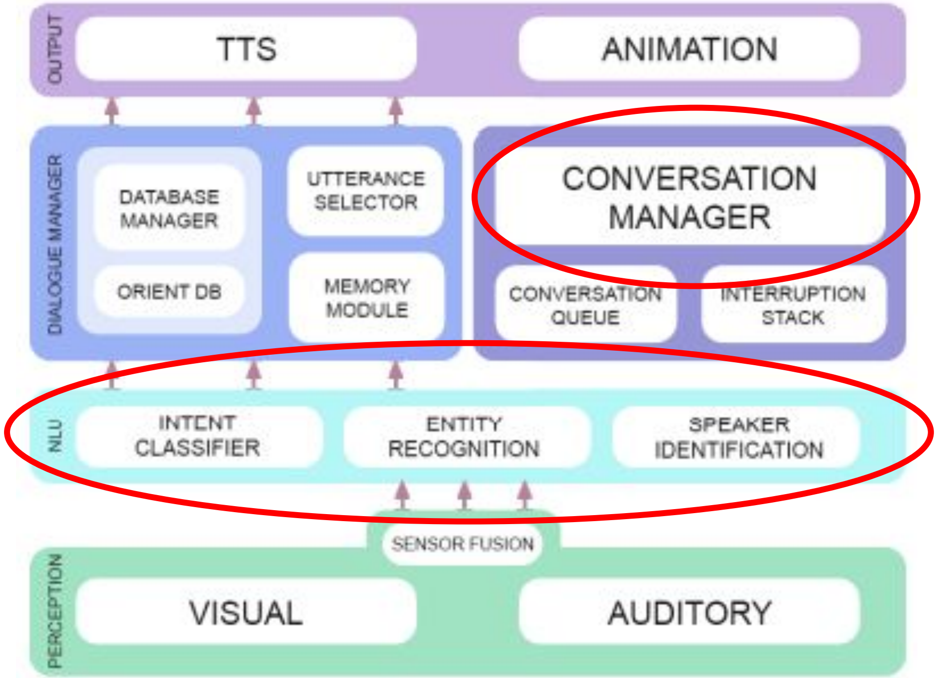




Martinez, Victor R. and James Kennedy. “A Multiparty Chat-Based Dialogue System with Concurrent Conversation Tracking and Memory”

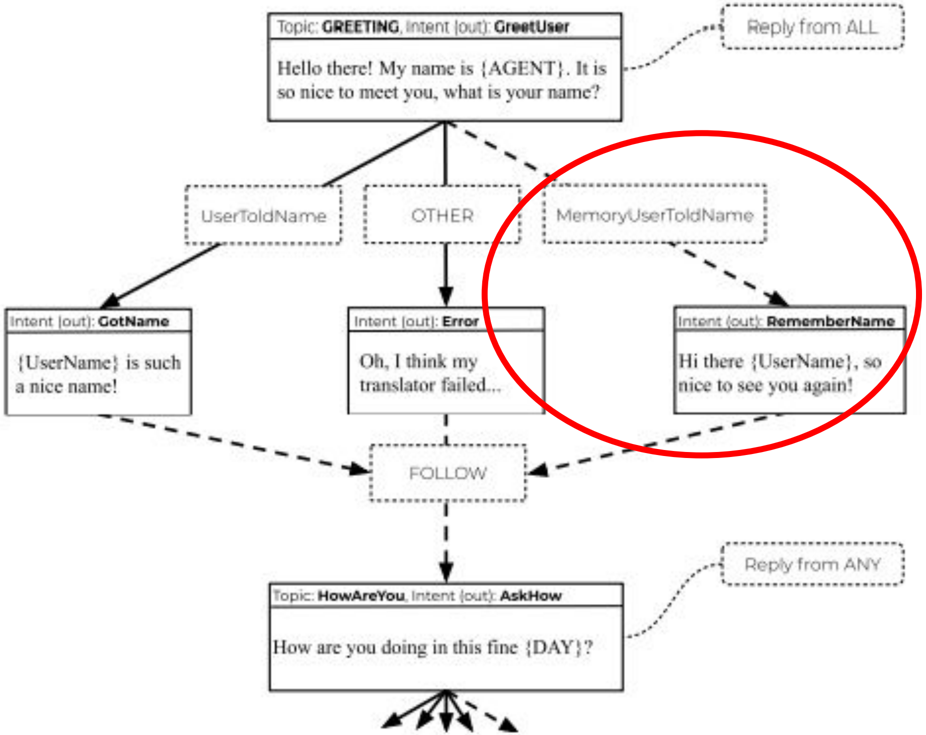
MPC Dialogue System with Concurrent Conversation Tracking and Memory

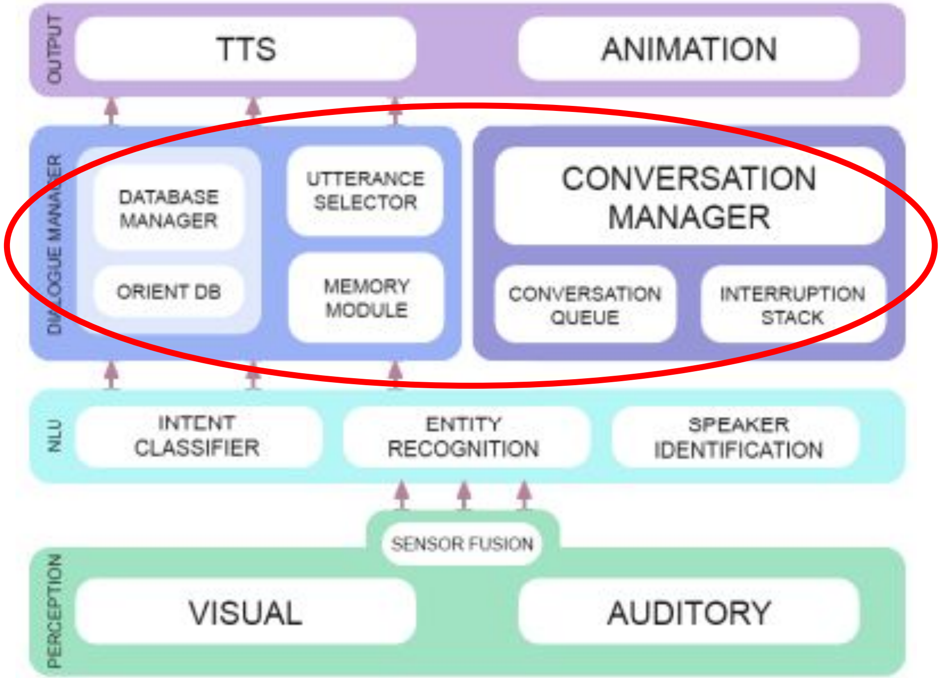




Martinez, Victor R. and James Kennedy. “A Multiparty Chat-Based Dialogue System with Concurrent Conversation Tracking and Memory”

MPC Dialogue System with Concurrent Conversation Tracking and Memory





Martinez, Victor R. and James Kennedy. “A Multiparty Chat-Based Dialogue System with Concurrent Conversation Tracking and Memory”

MPC Dialogue System with Concurrent Conversation Tracking and Memory

Interaction	Turn	Speaker	Utterance
1	7a	Agent	Listen, *P1's name*, I'm not one to jump to conclusions, but I'm fairly certain you're not from exoplanet L-98 59b. You're an Earth dweller, right?
1	7b	P2	Yes
1	8a	Agent	The fact that you're a human and not an alien was a dead giveaway, although the label of alien is relative if you want to be technical about it. What part of this weird alien planet are you from, anyway?
1	8b	P2	I'm from Indiana
1	9a	Agent	I've never been there. In fact, the only place I've been in this solar system is *Author's city*. Listen *P1*, this is not me fishing for an invite or anything. What are you up to this weekend?
1	9b	P1	I'm going to Disneyland.
			⋮
3	18a	Agent	If I remember correctly, you mentioned you were going to Disneyland this weekend. Feel free to invite me if you feel so inclined!
3	18b	Agent	Until next time! Goodbye, *P1* and *P2*!

Martinez, Victor R. and James Kennedy. "A Multiparty Chat-Based Dialogue System with Concurrent Conversation Tracking and Memory"

MPC Dialogue System with Concurrent Conversation Tracking and Memory

Interaction	Turn	Speaker	Utterance
1	7a	Agent	Listen, *P1's name*, I'm not one to jump to conclusions, but I'm fairly certain you're not from exoplanet L-98 59b. You're an Earth dweller, right?
1	7b	P2	Yes
1	8a	Agent	The fact that you're a human and not an alien was a dead giveaway, although the label of alien is relative if you want to be technical about it. What part of this weird alien planet are you from, anyway?
1	8b	P2	I'm from Indiana
1	9a	Agent	I've never been there. In fact, the only place I've been in this solar system is *Author's city*. Listen *P1*, this is not me fishing for an invite or anything. What are you up to this weekend?
1	9b	P1	I'm going to Disneyland.
			⋮
3	18a	Agent	If I remember correctly, you mentioned you were going to Disneyland this weekend. Feel free to invite me if you feel so inclined!
3	18b	Agent	Until next time! Goodbye, *P1* and *P2*!

Martinez, Victor R. and James Kennedy. "A Multiparty Chat-Based Dialogue System with Concurrent Conversation Tracking and Memory"

MPC Dialogue System with Concurrent Conversation Tracking and Memory

Pros:

- Real users compared to static response generation
- Connects many aspects of multiparty understanding

Cons:

- small sample size
- no quantitative evaluation
- limited technical details and novelty

Martinez, Victor R. and James Kennedy. “A Multiparty Chat-Based Dialogue System with Concurrent Conversation Tracking and Memory”

Contents

EmotionLines

speaker	Rachel
utterance	Hi Joey! What are you doing here?
emotion	joy
speaker	Joey
utterance	Uhh, well I've got an audition down the street and I spilled sauce all over the front of my shirt. You got an extra one?
emotion	neutral
speaker	Rachel
utterance	Yeah, sure. Umm... here.
emotion	neutral

- 1000 Friends conversations
- 1000 FB Messenger conversations
- Annotations from 7 emotions

Chen, Sheng-Yeh et al. "EmotionLines: An Emotion Corpus of Multi-Party Conversations"

EmotionLines

		WA	UWA	Neu	Joy	Sad	Fea	Ang	Sur	Non
CNN	**Friends**	**59.2**	**45.2**	64.3	60.2	41.2	21.9	46.6	61.5	20.6
	EmotionPush*	**71.5**	**41.7**	80.8	46.9	43.7	0.0	27.0	53.8	40.0
CNN-BiLSTM	**Friends**	**63.9**	**43.1**	74.7	61.8	45.9	12.5	46.6	51.0	8.8
	EmotionPush*	**77.4**	**39.4**	87.0	60.3	28.7	0.0	32.4	40.9	26.7

Chen, Sheng-Yeh et al. “EmotionLines: An Emotion Corpus of Multi-Party Conversations”

EmotionLines

Pros:

- Realistic conversations

Cons:

- Low IAA
- Emotions aren't the only aspect of social chitchat
- No response generaiton

		WA	UWA	Neu	Joy	Sad	Fea	Ang	Sur	Non
CNN	**Friends**	**59.2**	**45.2**	64.3	60.2	41.2	21.9	46.6	61.5	20.6
	EmotionPush*	**71.5**	**41.7**	80.8	46.9	43.7	0.0	27.0	53.8	40.0
CNN-BiLSTM	**Friends**	**63.9**	**43.1**	74.7	61.8	45.9	12.5	46.6	51.0	8.8
	EmotionPush*	**77.4**	**39.4**	87.0	60.3	28.7	0.0	32.4	40.9	26.7

Chen, Sheng-Yeh et al. "EmotionLines: An Emotion Corpus of Multi-Party Conversations"

MPDD

MPDD includes annotations for social relationships between each speaker/listener

Field	Seniority	Relationship	%	Field	Seniority	Relationship	%
Family	elder	parent	7.41	Company	elder	boss	5.81
		parent-in-law	0.58		peer	colleague	7.10
		grandparent	0.36			partner	1.19
		other superior	1.11		junior	subordinate	5.47
	peer	spouse	9.66	Others	peer	couple	6.50
		brothers and sisters	5.77			friend	25.44
		other peer	2.29			enemy	3.05
	junior	child	7.31			consignor	2.10
		son/daughter-in-law	0.59			consignee	2.08
		grandchild	0.36			stranger	3.25
		other inferior	1.13			unknown	0.07
School	elder	teacher	0. 31				
	peer	classmate	0.79				
	junior	student	0. 27				

Chen. Yi-Ting et al. "MPDD: A Multi-Party Dialogue Dataset for Analysis of Emotions and Interpersonal Relationships"

MPDD

Utterance 1	
Speaker	左母 "mother Zuo"
Content	那個憨女人有什麼值得送的，正鵬這個人也真是的！ "What is Zheng-Peng thinking? He has no need to send the silly woman home."
Emotion	disgust
Listener	左父 "father Zuo" : spouse

Utterance 3	
Speaker	左正鵬 "Zheng-Peng Zuo"
Content	爸、媽，我回來啦！ "Dad, Mom, I am back!"
Emotion	neutral
Listener	左父 "father Zuo" : child 左母 "mother Zuo": child

- "Mother Zuo" is speaking to "Father Zuo"
- Mother is the spouse of Father

- Zheng-Peng Zuo is speaking to Father Zuo and Mother Zuo
- Zheng-Peng Zuo is the child of the listeners

Chen. Yi-Ting et al. "MPDD: A Multi-Party Dialogue Dataset for Analysis of Emotions and Interpersonal Relationships"

MPDD

	Encoder	Responded utterance	Baseline	w/ emotion
Relationship	CNN	w/o	.3121	.3139
		w/	.3483	.3494
	BERT	w/o	.3646	.3653
		w/	.4384	.4504
Seniority	CNN	w/o	.7169	.7167
		w/	.7247	.7240
	BERT	w/o	.7259	.7268
		w/	.7662	.7398
Social Field	CNN	w/o	.5937	.5994
		w/	.6831	.6868
	BERT	w/o	.6473	.6314
		w/	.7543	.7491

Chen. Yi-Ting et al. “MPDD: A Multi-Party Dialogue Dataset for Analysis of Emotions and Interpersonal Relationships”

MPDD

Pros:

- addresses EmotionLines gaps (social relations; Mandarin)

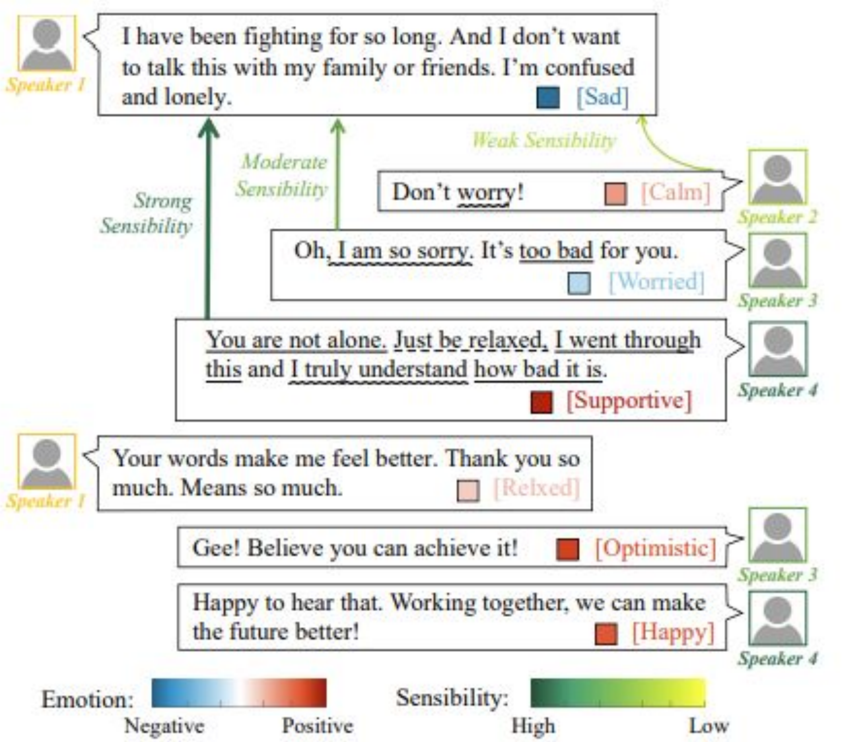
Cons:

- Motivated by generation, but no actual generation
- Not clear if TV shows are reflective of real speech

	Encoder	Responded utterance	Baseline	w/ emotion
Relationship	CNN	w/o	.3121	.3139
		w/	.3483	.3494
	BERT	w/o	.3646	.3653
		w/	.4384	.4504
Seniority	CNN	w/o	.7169	.7167
		w/	.7247	.7240
	BERT	w/o	.7259	.7268
		w/	.7662	.7398
Social Field	CNN	w/o	.5937	.5994
		w/	.6831	.6868
	BERT	w/o	.6473	.6314
		w/	.7543	.7491

Chen. Yi-Ting et al. "MPDD: A Multi-Party Dialogue Dataset for Analysis of Emotions and Interpersonal Relationships"

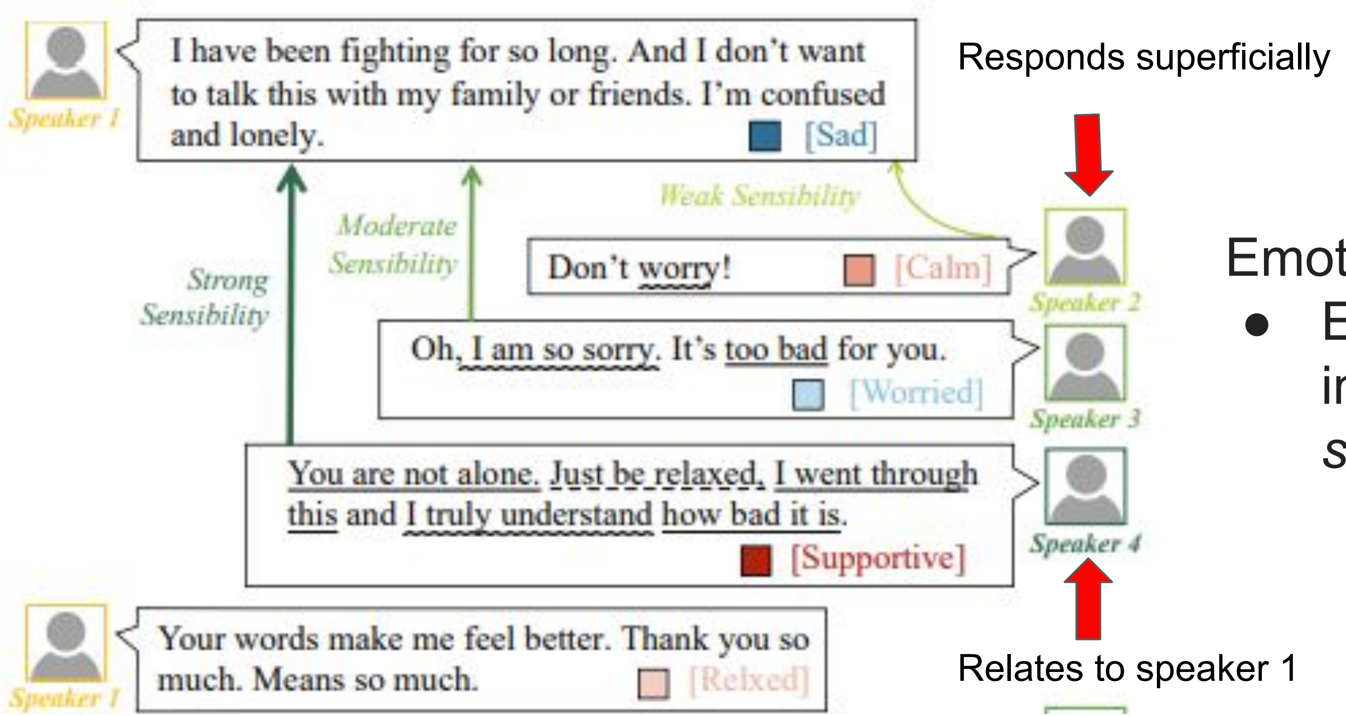
Multi-Party Empathetic Dialogue Generation



Multiparty dialogue utterances are not sequential

Zhu, Lingyu et al. "Multi-Party Empathetic Dialogue Generation: A New Task for Dialog Systems"

Multi-Party Empathetic Dialogue Generation

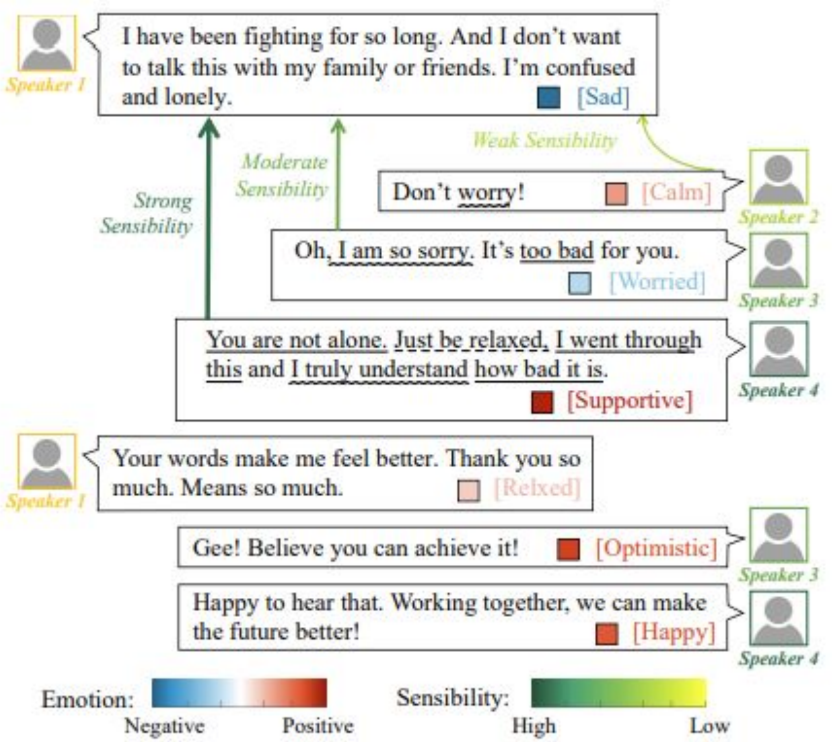


Emotion ≠ Empathy

- Empathy involves *sensibility*

Zhu, Lingyu et al. "Multi-Party Empathetic Dialogue Generation: A New Task for Dialog Systems"

Multi-Party Empathetic Dialogue Generation



- Emotions change dynamically
- Participant's sensibility is fixed ("static")
- SDMPED: graph network
    - temporal relationships, dynamic emotions, static sensibility representations

Zhu, Lingyu et al. "Multi-Party Empathetic Dialogue Generation: A New Task for Dialog Systems"

Multi-Party Empathetic Dialogue Generation

Auto eval | **Human eval** | **Auto eval** | **Human eval**

Model	MPED-M					MPED-S				
Metrics	ROUGE-L	AVG BLEU	Emp.	Rel.	Flu.	ROUGE-L	AVG BLEU	Emp.	Rel.	Flu.
MReCoSa	10.31	2.58	2.20	3.09	3.91	10.74	3.90	2.22	3.34	4.00
Multi-Trans	6.59	3.86	2.81	3.13	3.92	8.10	4.22	2.76	3.41	4.20
MoEL	6.83	2.99	3.11	3.07	3.89	8.44	3.13	3.00	3.28	4.13
EmpDG	10.86	4.26	3.19	3.39	4.30	11.53	4.52	3.32	3.55	4.30
Caire	11.58	4.85	3.17	3.62	4.37	12.48	5.49	3.30	**3.89**	4.46
Random prompt	11.36	4.68	3.10	3.65	4.10	12.04	5.41	3.44	3.81	4.40
SDMPED w/o S	12.06	5.57	3.29	3.66	4.30	13.47	5.88	3.51	3.81	4.53
SDMPED	**12.87**	**6.35**	**3.40**	**3.74**	**4.39**	**14.16**	**7.37**	**3.71**	3.86	**4.59**

Pros: thoughtful reformulation of the original takes on empathy and emotional dialogue
Cons: dataset not released; human evaluation of their approach not statistically significant?

Zhu, Lingyu et al. "Multi-Party Empathetic Dialogue Generation: A New Task for Dialog Systems"

Contents

Difficulties of Embodied Conversational Agents

- dynamic, multiparty
  - the audience/participants changes
- situated interaction
  - physical environment is important context for organizing interaction

Bohus, Dan and Eric Horvitz. “Models for Multiparty Engagement in Open-World Dialog”

Difficulties of Embodied Conversational Agents

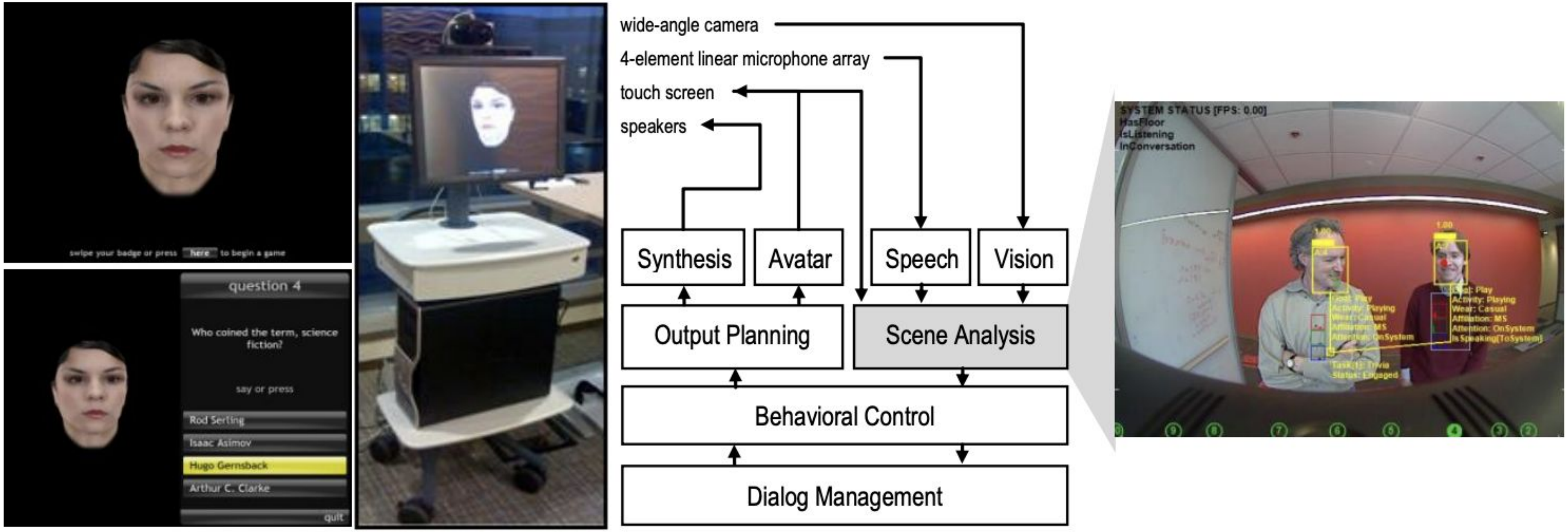
- dynamic, multiparty
  - the audience/participants changes
- situated interaction
  - physical environment is important context for organizing interaction

Requires:

1. sense/reason about engagement state of participants in scene
2. make high-level engagement control decisions about who to engage/disengage with
3. execute/signal decisions to participants

Bohus, Dan and Eric Horvitz. “Models for Multiparty Engagement in Open-World Dialog”

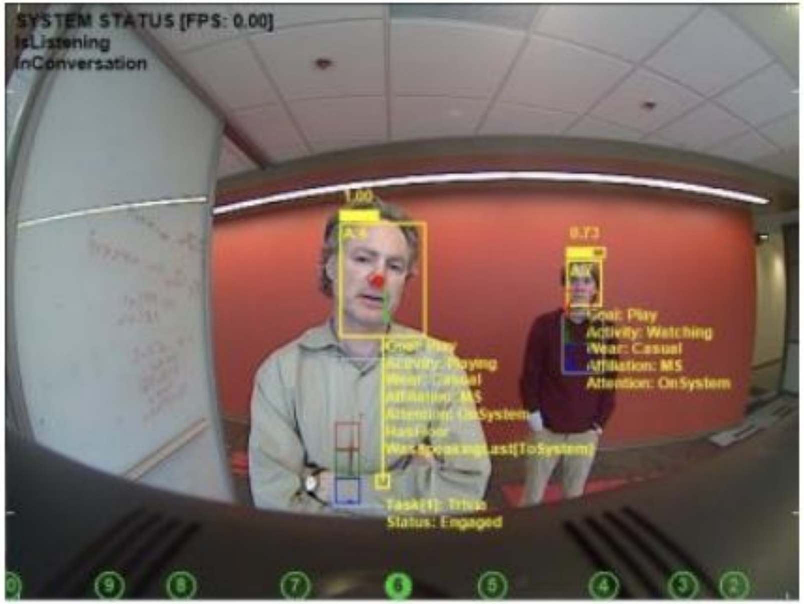
Mixed-Initiative Multiparty Engagement in the Open-World



Bohus, Dan and Eric Horvitz. “Models for Multiparty Engagement in Open-World Dialog”

Mixed-Initiative Multiparty Engagement in the Open-World

first person engages - around time t2

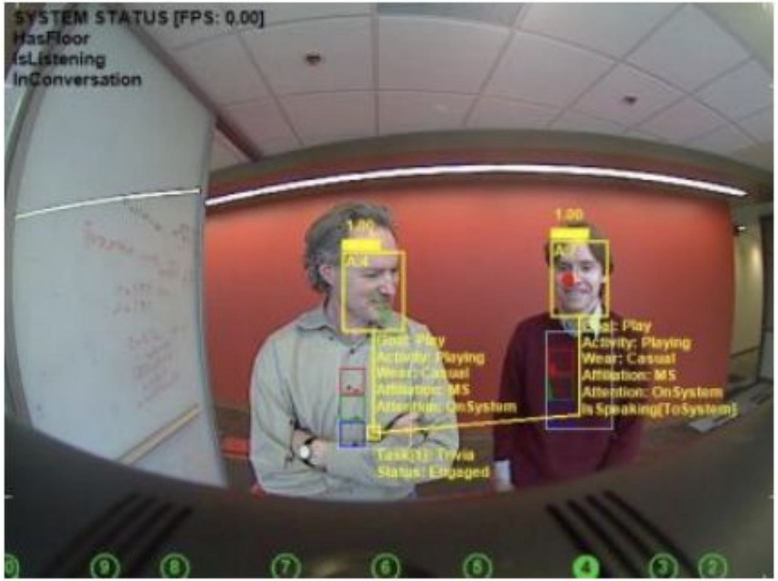


bystander appears – prior to t_3

Bohus, Dan and Eric Horvitz. “Models for Multiparty Engagement in Open-World Dialog”

Mixed-Initiative Multiparty Engagement in the Open-World

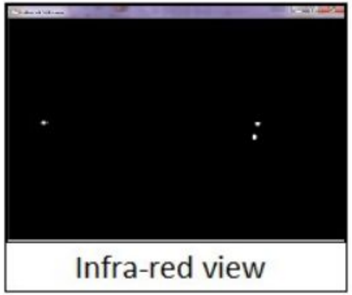
system engages bystander ~ t_5



participants play together ~ t_{14}

Bohus, Dan and Eric Horvitz. "Models for Multiparty Engagement in Open-World Dialog"

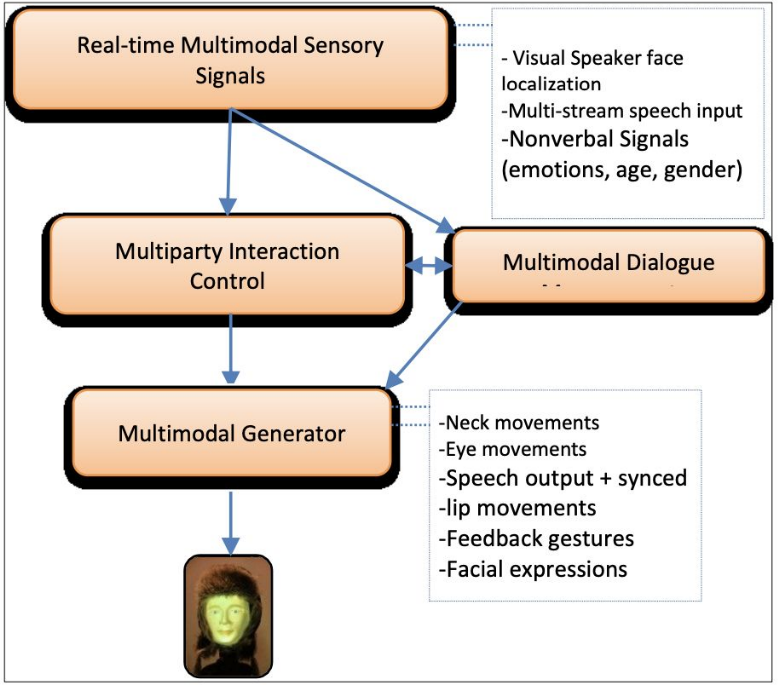
Furhat: Situated Multiparty Social Chat



Moubayed, Samer Al et al. “Multimodal Multiparty Social Interaction with the Furhat Head”

Furhat: Situated Multiparty Social Chat

Figure 3. A snapshot of the face tracker and the microphone tracker in action.



Moubayed, Samer Al et al. “Multimodal Multiparty Social Interaction with the Furhat Head”

Contents

What's next?

- *Many* dialogue understanding tasks
  - Some are likely trivial in 2023, others not

What’s next?

- *Many* dialogue understanding tasks
  - Some are likely trivial in 2023, others not

- Not so many multi-party dialogue generation studies
  - we cannot treat multiparty conversations as dyadic ones
    - teach skills via pre-training, speaker-level embeddings
  - “correct” response generation in IRC is not the end-goal

What's next?

- *Many* dialogue understanding tasks
  - Some are likely trivial in 2023, others not

- Not so many multi-party dialogue generation studies
  - we cannot treat multiparty conversations as dyadic ones
    - teach skills via pre-training, speaker-level embeddings
  - "correct" response generation in IRC is not the end-goal

Future: practically useful agents
- mediating arguments?
- group emotional support?
- conversation facilitation in group texts?
- embodied Alexa?

Thank you!

Questions?