On the Underspecification of Situations in Open-domain Conversational Datasets

NLP for Conversational AI Workshop @ ACL 2023

Naoki Otani, Jun Araki, HyeongSik Kim, Eduard Hovy

Megagon Labs

BOSCH

THE UNIVERSITY OF MELBOURNE

*Work done at

Carnegie Mellon University
Language Technologies Institute

A (partial) set of **observable** information about objects in the scene

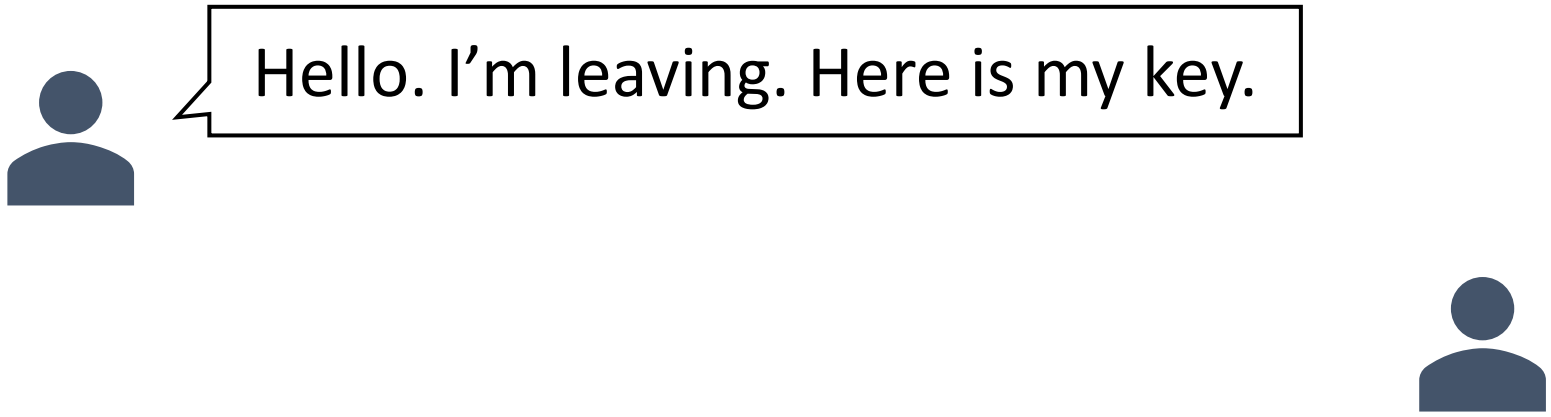
Claim:

We need to talk about **situation** for enhancing open-domain conversational agents!

Quick overview:

Meaning of language heavily relies on situation

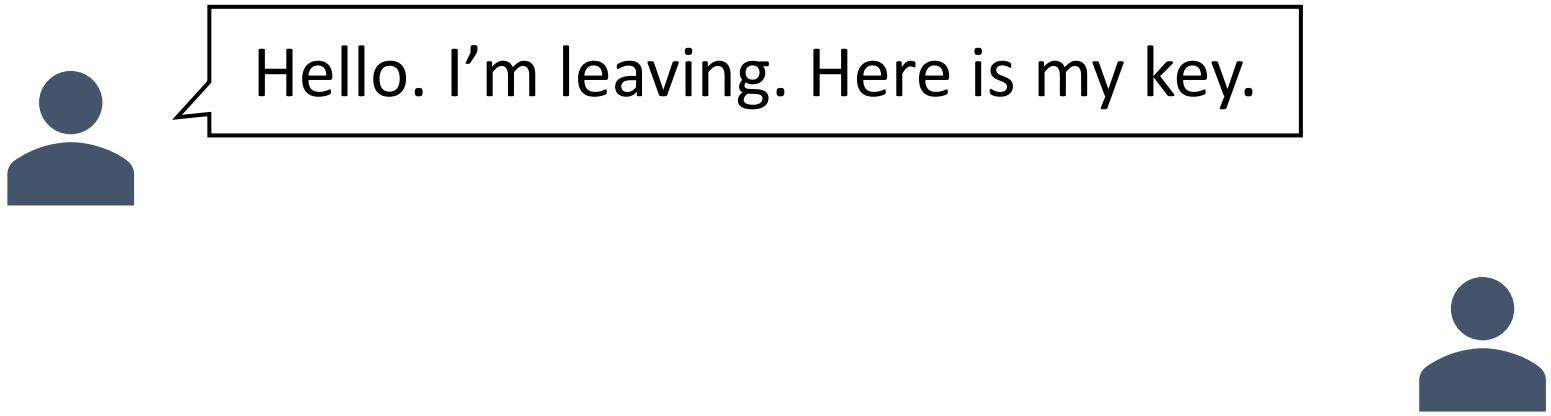
Example: Two persons are having a conversation.



Quick overview:

Meaning of language heavily relies on situation

Let’s say…

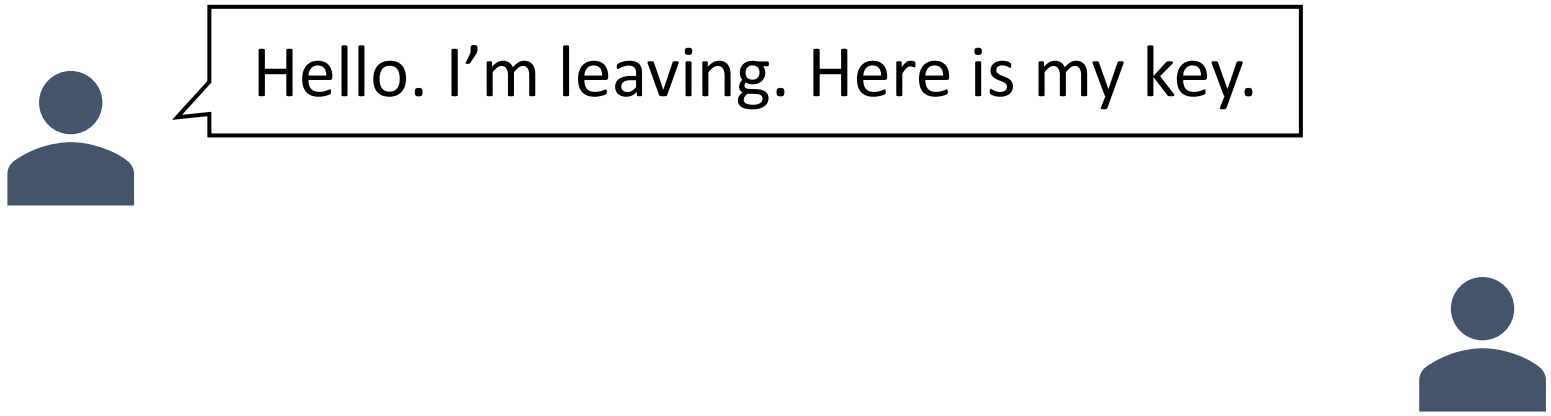


Quick overview:

Meaning of language heavily relies on situation

Let’s say…

Location: Hotel front desk



Quick overview:

Meaning of language heavily relies on situation

Let’s say…

Location: Hotel front desk

Guest

Hello. I’m leaving. Here is my key.

Staff

Quick overview:

Meaning of language heavily relies on situation

Let’s say…

Location: Hotel front desk

Hello. I’m leaving. Here is my key.

Guest

Check-out time: 11am
Current time: 10:30am
[…]

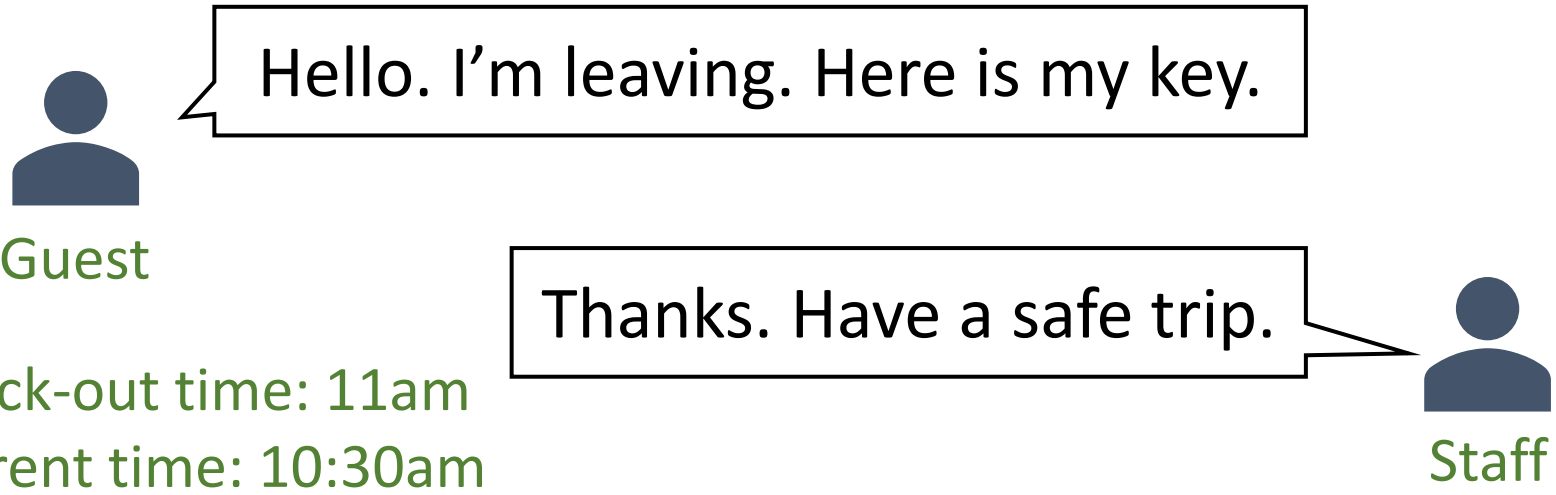
Staff

Quick overview:

Meaning of language heavily relies on situation

Let’s say…

Location: Hotel front desk

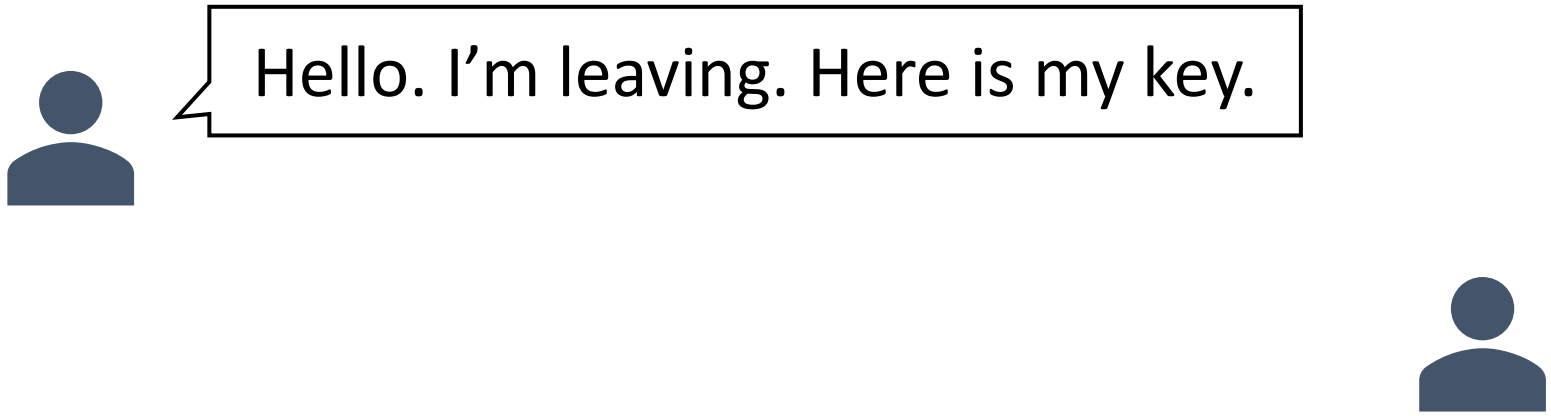


Check-out time: 11am
Current time: 10:30am
[…]

Quick overview:

Meaning of language heavily relies on situation

Let’s say…

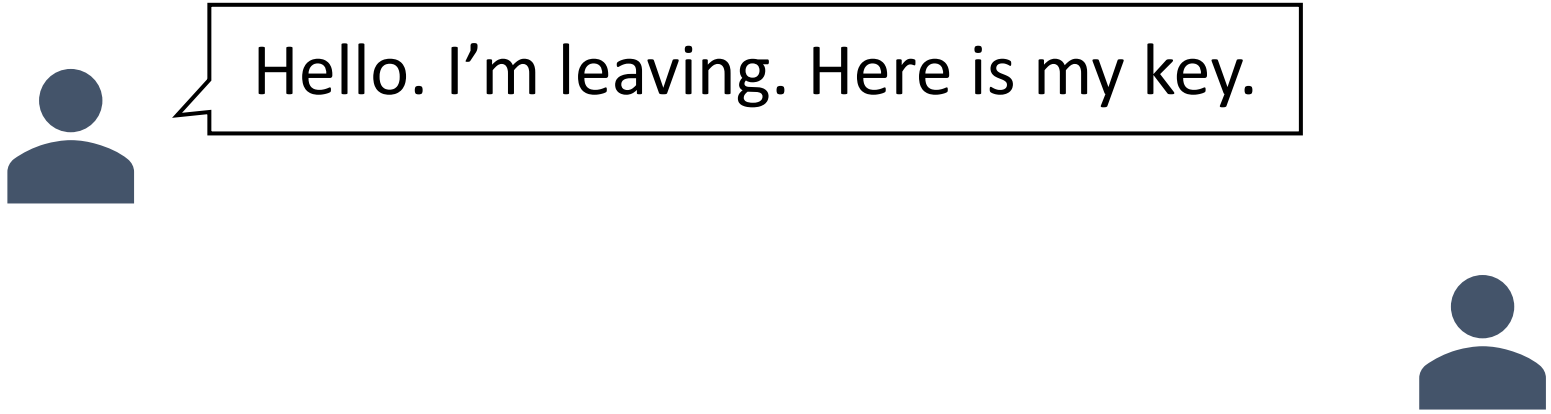


Quick overview:

Meaning of language heavily relies on situation

Let’s say…

Location: House



Quick overview:

Meaning of language heavily relies on situation

Let’s say…

Location: House

Hello. I’m leaving. Here is my key.

Pat, going to work

Chris,
cooking in the kitchen

Quick overview:

Meaning of language heavily relies on situation

Let’s say…

Location: House

Hello. I’m leaving. Here is my key.

Pat, going to work

Chris,
cooking in the kitchen

Chris is Pat’s housemate.
Chris recently lost their key.

Quick overview:

Meaning of language heavily relies on situation

Let’s say…

Location: House

Hello. I’m leaving. Here is my key.

Pat, going to work

Thanks. Do you want me to lock the door?

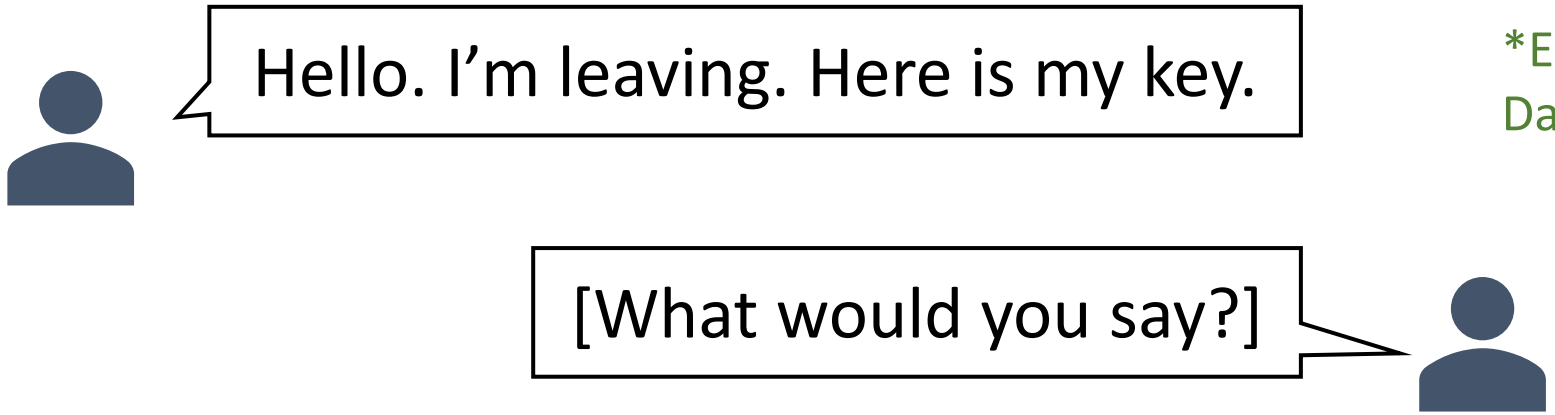
Chris,
cooking in the kitchen

Chris is Pat’s housemate.
Chris recently lost their key.

Quick overview:

Meaning of language heavily relies on situation

However, a conversation task is often designed to be like **this:**



*Example from DailyDialog (Li+17)

Situation is not specified

Quick overview:

Meaning of language heavily relies on situation

However, a conversation task is often designed to be like **this:**

Hello. I'm leaving. Here is my key.

*Example from DailyDialog (Li+17)

Chatbots respond based on **their own assumptions** about the situation.

Quick overview:

Meaning of language heavily relies on situation

However, a conversation task is often designed to be like **this:**

Hello. I'm leaving. Here is my key.

*Example from DailyDialog (Li+17)

Chatbots respond based on **their own assumptions** about the situation.

Thanks. Have a safe trip.

GPT-3(Brown+20)

Quick overview:

Meaning of language heavily relies on situation

However, a conversation task is often designed to be like **this:**

Hello. I'm leaving. Here is my key.

*Example from DailyDialog (Li+17)

Chatbots respond based on **their own assumptions** about the situation.

Thanks. Have a safe trip. → GPT-3 (Brown+20)

[…] Do you want me to lock the door? → BlenderBot2 (Komeili+22; Xu+22)

Quick overview:

Meaning of language heavily relies on situation

However, a conversation task is often designed to be like **this:**

Issues:

(1) **Underspecification of the problem space**

It's often impossible to identify the meaning/intent of an utterance.

(2) **System behavior**

The chatbot's internal assumption is not always correct or socially appropriate.

(3) **Evaluation**

Even human cannot decide which response is correct.

[…] Do you want me to lock the door?

BlenderBot2 (Komeili+22; Xu+22)

What is Situation?

What is situation?:

Situation Semantics (1980s~, Jon Barwise, John Perry)

Situation = Entities/eventualities and their properties and relations

What is situation?:

Situation Semantics (1980s~, Jon Barwise, John Perry)

Situation = Entities/eventualities and their properties and relations

Task-specific: Essential for task-oriented systems

- Vocabulary of user request types (e.g., for the restaurant domain, users can request {search, reservation, …})
- Task flow (restaurant → (1) venue, (2) size of party, (3) date/time, (4) …)
- Database/API (list of restaurants, etc.)

What is situation?:

Situation Semantics (1980s~, Jon Barwise, John Perry)

Situation = Entities/eventualities and their properties and relations

Task-specific: Essential for task-oriented systems

- Vocabulary of user request types (e.g., for the restaurant domain, users can request {search, reservation, …})
- Task flow (restaurant → (1) venue, (2) size of party, (3) date/time, (4) …)
- Database/API (list of restaurants, etc.)

User-specific: Useful for more engaging communication

- Persona, Behavior, Preference, Emotion, Intent/Goal, …

What is situation?:

Situation Semantics (1980s~, Jon Barwise, John Perry)

Situation = Entities/eventualities and their properties and relations

Task-specific: Essential for task-oriented systems

- Vocabulary of user request types (e.g., for the restaurant domain, users can request {search, reservation, …})
- Task flow (restaurant → (1) venue, (2) size of party, (3) date/time, (4) …)
- Database/API (list of restaurants, etc.)

User-specific: Useful for more engaging communication

- Persona, Behavior, Preference, Emotion, Intent/Goal, …

External knowledge:

- Wikipedia articles, Customer reviews, …

What is situation?:

Situation Semantics (1980s~, Jon Barwise, John Perry)

Situation = Entities/eventualities and their properties and relations

Task-specific: Essential for task-oriented systems

Vocabulary of user request types (e.g., for the restaurant domain, users can request {search, reservation, …})
Task flow (restaurant → (1) venue, (2) size of party, (3) date/time, (4) …)
Database/API (list of restaurants, etc.)

User-specific: Useful for more engaging communication

Persona, Behavior, Preference, Emotion, Intent/Goal, …

External knowledge:

Wikipedia articles, Customer reviews, …

Environmental: Much information comes from physical sensors

Scene, Date/time, Sound, Smell,
Weather report, Traffic report, …

What is situation?:

Situation Semantics (1980s~, Jon Barwise, John Perry)

Situation = Entities/eventualities and their properties and relations

Task-specific: Essential for task-oriented systems

Vocabulary of user request types (e.g., for the restaurant domain, users can request {search, reservation, …})
Task flow (restaurant → (1) venue, (2) size of party, (3) date/time, (4) …)
Database/API (list of restaurants, etc.)

User-specific: Useful for more engaging communication

Persona, Behavior, Preference, Emotion, Intent/Goal, …

External knowledge:

Wikipedia articles, Customer reviews, …

Environmental: Much information comes from physical sensors

Scene, Date/time, Sound, Smell,
Weather report, Traffic report, …

Key: Not always mentioned explicitly by conversation's participants
→ Prior work developed **dedicated semantic representations**
and/or treat situational information as **extra signals**

Status quo

Background:

The importance of situation has been well-recognized in dialogue studies.

Train station

Traveler

Indirect speech acts (Clark'79;Allen&Perrault'80)

Do you know when the Windsor train leaves?

3:15 *at gate 7.*

Station attendant

Clark. 1979. Responding to Indirect Speech Acts.
Allen and Perrault. 1980. Analyzing Intention in Utterances.

Background:

The importance of situation has been well-recognized in dialogue studies.

Train station

Traveler

Indirect speech acts (Clark'79;Allen&Perrault'80)

Do you know when the Windsor train leaves?

3:15 *at gate 7.*

Station attendant

Location = Station
Behavior = Carrying luggage
Windsor train
departure = 3:15
location = gate 7
…

Situation
(Partial information about the world)

Clark. 1979. Responding to Indirect Speech Acts.
Allen and Perrault. 1980. Analyzing Intention in Utterances.

Background:

The importance of situation has been well-recognized in dialogue studies.

Train station

Traveler

Indirect speech acts (Clark'79;Allen&Perrault'80)

Do you know when the Windsor train leaves?

3:15 *at gate 7.*

Station attendant

Location = Station
Behavior = Carrying luggage
Windsor train
departure = 3:15
location = gate 7
…

Situation
(Partial information about the world)

Asking the departure time
→ Goal = To take the train

Clark. 1979. Responding to Indirect Speech Acts.
Allen and Perrault. 1980. Analyzing Intention in Utterances.

Background:

Open-domain

Many recent datasets lack comprehensive situational information

Dataset	Data source	Situational information	Reference:Citation
Twitter	Twitter	None	(Ritter+'10): 484
Reddit	Reddit	None	(Baumgartner+'20): 484
DailyDialog	ESL learning resources	None	(Li+'17): 783
PersonaChat	Crowdsourcing	Persona	(Zhang+'18): 974
ConvAI	Crowdsourcing	Persona	(Dinan+'19): 262
CMU_DoG	Crowdsourcing	Related Wikipedia pages	(Zhou+'18): 183
WizardOfWikipedia	Crowdsourcing	Related Wikipedia pages	(Dinan+'18): 617
EmpatheticDialogue	Crowdsourcing	Emotion	(Rashkin+'18): 489

Background:

Open-domain

Many recent datasets lack comprehensive situational information

Dataset	Data source	Situational information	Reference:Citation
Twitter	Twitter	None	(Ritter+'10): 484
Reddit	Reddit	None	gartner+'20): 484
DailyDialog	ESL learning resources	None	(Li+'17): 783
PersonaChat	Crowdsourcing	Persona	(Zhang+'18): 974
ConvAI	Crowdsourcing	Persona	(Dinan+'19): 262
CMU_DoG	Crowdsourcing	Related Wikipedia pages	(Zhou+'18): 183
WizardOfWikipedia	Crowdsourcing	Related Wikipedia pages	(Dinan+'18): 617
EmpatheticDialogue	Crowdsourcing	Emotion	(Rashkin+'18): 489

Need to infer everything **only from conv. history**

Background:

Open-domain

Many recent datasets lack comprehensive situational information

Dataset	Data source	Situational information	Reference:Citation
Twitter	Twitter	None	(Ritter+'10): 484
Reddit	Reddit	None	gartner+'20): 484
DailyDialog	ESL learning resources	None	(Li+'17): 783
PersonaChat	Crowdsourcing	Persona	(Zhang+'18): 974
ConvAI	Crowdsourcing	Persona	(Dinan+'19): 262
CMU_DoG	Crowdsourcing	Related Wikipedia pages	(Zhou+'18): 183
WizardOfWikipedia	Crowdsourcing	Related Wikipedia pages	(Dinan+'18): 617
EmpatheticDialogue	Crowdsourcing	Emotion	(Rashkin+'18): 489

Need to infer everything **only from conv. history**

Problem: **Comprehensive environmental information is missing** in many datasets (when, where, who is speaking, etc.)

SUGAR (Otani et al., ACL2023)

A Textual Dataset of Situated, Goal-aware, Proactive Responses

1.7k single-turn conversations in help-seeking scenarios collected by crowdsourcing

Utterance	Please turn on the TV.
Situations	It is evening now. [user] is home. [user] is in the living room. [user] is sitting on the couch. [user] has a TV in the house. [user] has an outfit on the bed. [user] has drinks and snacks in the kitchen. [user] has game cards on the shelf. The TV is off. [someone]'s birthday is today. There are several sports games available to watch. There is a basketball game scheduled.
Responses	Sure. Would you like me to check today's sports listings? *(Best)* Sure. Shall I pour a drink and bring some snacks for the game? *(Acceptable)* Sure, shall I select an outfit for you? *(Bad)*

SUGAR (Otani et al., ACL2023)

A Textual Dataset of Situated, Goal-aware, Proactive Responses

1.7k single-turn conversations in help-seeking scenarios collected by crowdsourcing

Utterance	Please turn on the TV.
Situations	It is evening now. [user] is home. [user] is in the living room. [user] is sitting on the couch. [user] has a TV in the house. [user] has an outfit on the bed. [user] has drinks and snacks in the kitchen. [user] has game cards on the shelf. The TV is off. [someone]'s birthday is today. There are several sports games available to watch. There is a basketball game scheduled.
Responses	Sure. Would you like me to check today's sports listings? *(Best)* Sure. Shall I pour a drink and bring some snacks for the game? *(Acceptable)* Sure, shall I select an outfit for you? *(Bad)*

Three response candidates w/ 3-level ratings

- 1/3: Reference response made by workers
- 2/3: Added through adversarial filtering

SUGAR (Otani et al., ACL2023)

A Textual Dataset of Situated, Goal-aware, Proactive Responses

1.7k single-turn conversations in help-seeking scenarios collected by crowdsourcing

Utterance	Please turn on the TV.
Situations	It is evening now. [user] is home. [user] is in the living room. [user] is sitting on the couch. [user] has a TV in the house. [user] has an outfit on the bed. [user] has drinks and snacks in the kitchen. [user] has game cards on the shelf. The TV is off. [someone]'s birthday is today. There are several sports games available to watch. There is a basketball game scheduled.
Responses	Sure. Would you like me to check today's sports listings? *(Best)* Sure. Shall I pour a drink and bring some snacks for the game? *(Acceptable)* Sure, shall I select an outfit for you? *(Bad)*

Situational information represented in simple En sentences (12 sents.)

Three response candidates w/ 3-level ratings

- 1/3: Reference response made by workers
- 2/3: Added through adversarial filtering

SUGAR (Otani et al., ACL2023)

A Textual Dataset of Situated, Goal-aware, Proactive Responses

1.7k single-turn conversations in help-seeking scenarios collected by crowdsourcing

Utterance	Please turn on the TV.
Situations	It is evening now. [user] is home. [user] is in the living room. [user] is sitting on the couch. [user] has a TV in the house. [user] has an outfit on the bed. [user] has drinks and snacks in the kitchen. [user] has game cards on the shelf. The TV is off. [someone]'s birthday is today. There are several sports games available to watch. There is a basketball game scheduled.
Responses	Sure. Would you like me to check today's sports listings? *(Best)* Sure. Shall I pour a drink and bring some snacks for the game? *(Acceptable)* Sure, shall I select an outfit for you? *(Bad)*

Situational information represented in simple En sentences (12 sents.)

Three response candidates w/ 3-level ratings

- 1/3: Reference response made by workers
- 2/3: Added through adversarial filtering

SUGAR (Otani et al., ACL2023)

A Textual Dataset of Situated, Goal-aware, Proactive Responses

1.7k single-turn conversations in help-seeking scenarios collected by crowdsourcing

Utterance	Please turn on the TV.
Situations	It is evening now. [user] is home. [user] is in the living room. [user] is sitting on the couch. [user] has a TV in the house. [user] has an outfit on the bed. [user] has drinks and snacks in the kitchen. [user] has game cards on the shelf. The TV is off. [someone]'s birthday is today. There are several sports games available to watch. There is a basketball game scheduled.
Responses	Sure. Would you like me to check today's sports listings? *(Best)* Sure. Shall I pour a drink and bring some snacks for the game? *(Acceptable)* Sure, shall I select an outfit for you? *(Bad)*

Situational information represented in simple En sentences (12 sents.)

- Relevant information: about 6 sentences
  - (Written by crowd workers to support the *Best* response)

Three response candidates w/ 3-level ratings

- 1/3: Reference response made by workers
- 2/3: Added through adversarial filtering

SUGAR (Otani et al., ACL2023)

A Textual Dataset of Situated, Goal-aware, Proactive Responses

1.7k single-turn conversations in help-seeking scenarios collected by crowdsourcing

Utterance	Please turn on the TV.
Situations	It is evening now. [user] is home. [user] is in the living room. [user] is sitting on the couch. [user] has a TV in the house. [user] has an outfit on the bed. [user] has drinks and snacks in the kitchen. [user] has game cards on the shelf. The TV is off. [someone]'s birthday is today. There are several sports games available to watch. There is a basketball game scheduled.
Responses	Sure. Would you like me to check today's sports listings? *(Best)* Sure. Shall I pour a drink and bring some snacks for the game? *(Acceptable)* Sure, shall I select an outfit for you? *(Bad)*

Situational information represented in simple En sentences (12 sents.)

- Relevant information: about 6 sentences
  - (Written by crowd workers to support the *Best* response)
- Irrelevant information (distractors)
  - (Sampled from the other examples)

Three response candidates w/ 3-level ratings

- 1/3: Reference response made by workers
- 2/3: Added through adversarial filtering

Representation of Situation:

Short English statements describing the world states

Situational statements – short sentences describing *observable* facts of the current situation

Category	Definition	Example
Location	Information about [user]'s current location.	[user] is home. / [user] is at the entrance of a house.
Possession	Information about what [user] possesses.	[user] owns a car. / There are apples in the kitchen.
Time	Information about time.	It's midnight. / It's morning.
Date	Information about date and season.	It's [user]'s birthday. / It's summer.
Behavior	Information about [user]'s behavior.	[user] just woke up. / [user] came back from jogging.
Environment	Information about non-user entities (person, objects, etc.).	The room temperature is hot. / [user]'s car has a flat tire.

A Case Study

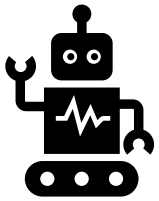
How does the inclusion of situational information matter?

Question:

Chat bots works better or worse with situation?

Conv. history (taken from CICERO(Ghosal+22)):

A: Hi, Mike! how are you feeling now?
B: How did you know I was here? is it Tom?
A: I was talking with Bob yesterday and I learnt your right leg had been injured. How did it happen?
B: [System output]



Chat bot
(Response generation system)

Question:

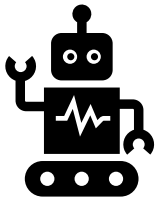
Chat bots works better or worse with situation?

Conv. history (taken from CICERO(Ghosal+22)):

A: Hi, Mike! how are you feeling now?
B: How did you know I was here? is it Tom?
A: I was talking with Bob yesterday and I learnt your right leg had been injured. How did it happen?
B: [System output]

Situation (our work; 10-12 statements)

Person B's leg **had a surgery** last night.
Person A and Person B are **in the hospital.**
Person B **injured his right leg** when he was **playing baseball**.
…



Chat bot
(Response generation system)

Question:

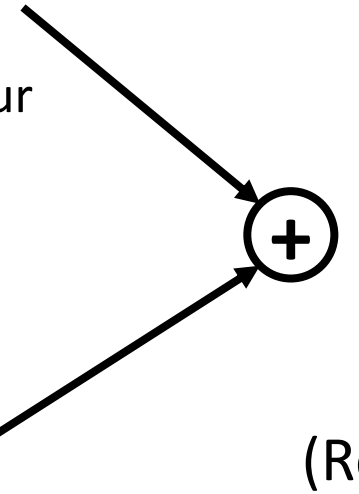
Chat bots works better or worse with situation?

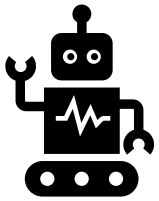
Conv. history (taken from CICERO(Ghosal+22)):

A: Hi, Mike! how are you feeling now?
B: How did you know I was here? is it Tom?
A: I was talking with Bob yesterday and I learnt your right leg had been injured. How did it happen?
B: [System output]

Situation (our work; 10-12 statements)

Person B's leg **had a surgery** last night.
Person A and Person B are **in the hospital.**
Person B **injured his right leg** when he was **playing baseball**.
…





Chat bot
(Response generation system)

Question:

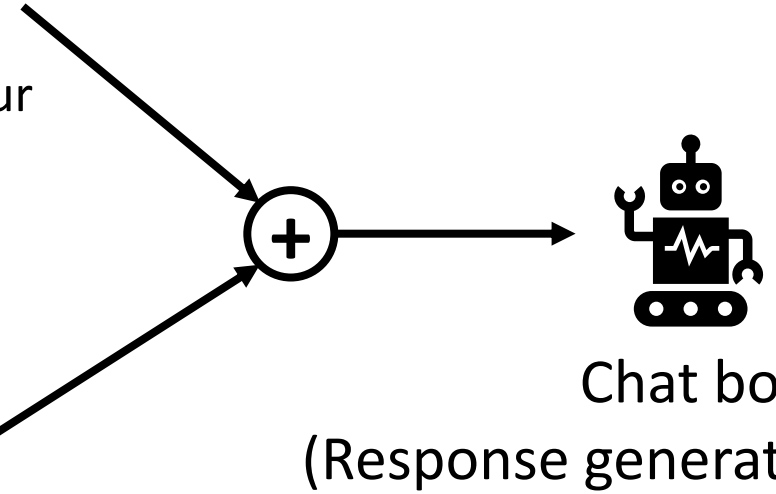
Chat bots works better or worse with situation?

Conv. history (taken from CICERO(Ghosal+22)):

A: Hi, Mike! how are you feeling now?
B: How did you know I was here? is it Tom?
A: I was talking with Bob yesterday and I learnt your right leg had been injured. How did it happen?
B: [System output]

Situation (our work; 10-12 statements)

Person B's leg **had a surgery** last night.
Person A and Person B are **in the hospital.**
Person B **injured his right leg** when he was **playing baseball**.
…



Chat bot
(Response generation system)

Question:

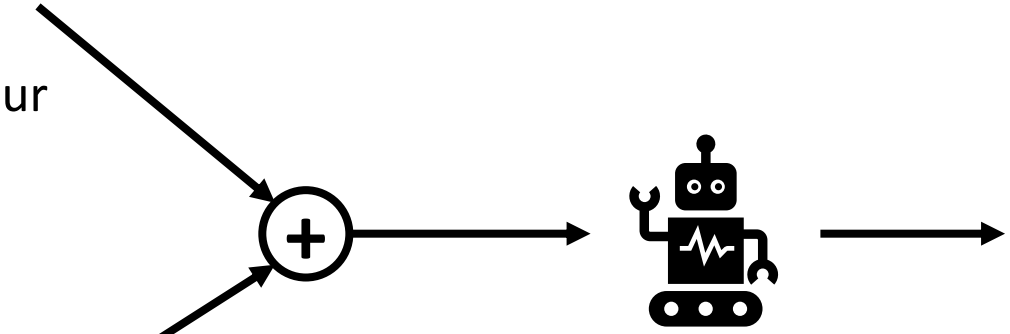
Chat bots works better or worse with situation?

Conv. history (taken from CICERO(Ghosal+22)):

A: Hi, Mike! how are you feeling now?
B: How did you know I was here? is it Tom?
A: I was talking with Bob yesterday and I learnt your right leg had been injured. How did it happen?
B: [System output]

Situation (our work; 10-12 statements)

Person B's leg **had a surgery** last night.
Person A and Person B are **in the hospital.**
Person B **injured his right leg** when he was **playing baseball**.
…



Chat bot
(Response generation system)

Question:

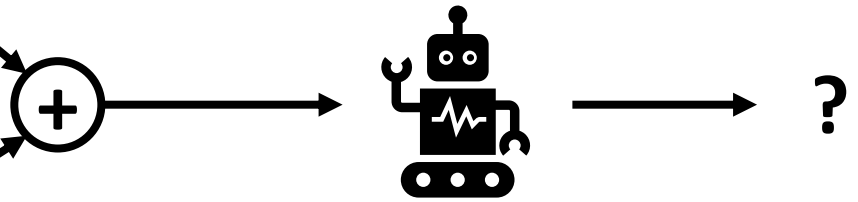
Chat bots works better or worse with situation?

Conv. history (taken from CICERO(Ghosal+22)):

A: Hi, Mike! how are you feeling now?
B: How did you know I was here? is it Tom?
A: I was talking with Bob yesterday and I learnt your right leg had been injured. How did it happen?
B: [System output]

Situation (our work; 10-12 statements)

Person B's leg **had a surgery** last night.
Person A and Person B are **in the hospital.**
Person B **injured his right leg** when he was **playing baseball**.
…



Chat bot
(Response generation system)

Question:

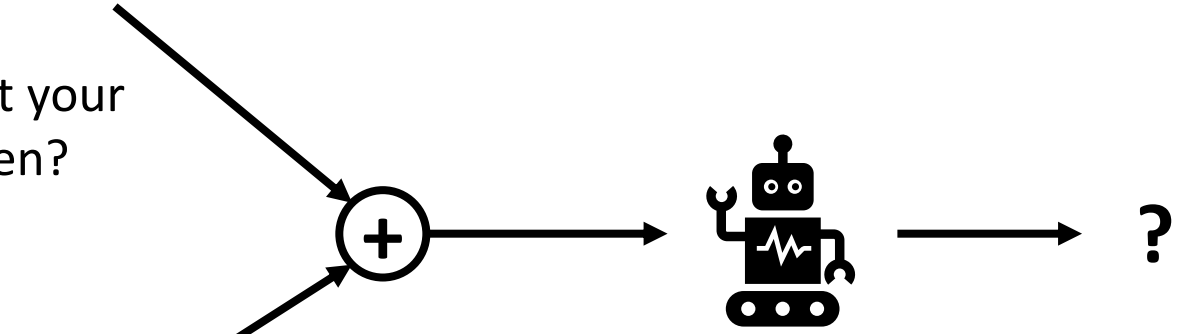
Chat bots works better or worse with situation?

Conv. history (taken from CICERO(Ghosal+22)):

A: Hi, Mike! how are you feeling now?
B: How did you know I was here? is it Tom?
A: I was talking with Bob yesterday and I learnt your right leg had been injured. How did it happen?
B: [System output]

Situation (our work; 10-12 statements)

Person B's leg **had a surgery** last night.
Person A and Person B are **in the hospital.**
Person B **injured his right leg** when he was **playing baseball**.
…



Hypothesis: Systems produce **more context-specific and meaningful** responses.

Data:

We created a dataset with situational statements semi-automatically.

Conv. history (taken from CICERO(Ghosal+22)):

A: Hi, Mike! how are you feeling now?
B: How did you know I was here? is it Tom?
A: I was talking with Bob yesterday and I learnt your right leg had been injured. How did it happen?
B: [System output]

Situation (our work)

Person B's leg **had a surgery** last night.
Person A and Person B are **in the hospital.**
Person B **injured his right leg** when he was **playing baseball**.
…

Data:

We created a dataset with situational statements semi-automatically.

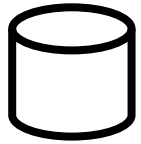
Conv. history (taken from CICERO(Ghosal+22)):

A: Hi, Mike! how are you feeling now?
B: How did you know I was here? is it Tom?
A: I was talking with Bob yesterday and I learnt your right leg had been injured. How did it happen?
B: [System output]

Situation (our work)

Person B's leg **had a surgery** last night.
Person A and Person B are **in the hospital.**
Person B **injured his right leg** when he was **playing baseball**.
…

SUGAR(Otani+23)



Parings of a conversation and situational statements.

Data:

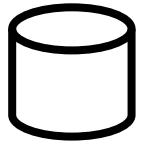
We created a dataset with situational statements semi-automatically.

Conv. history (taken from CICERO(Ghosal+22)):

A: Hi, Mike! how are you feeling now?
B: How did you know I was here? is it Tom?
A: I was talking with Bob yesterday and I learnt your right leg had been injured. How did it happen?
B: [System output]

Situation (our work)

Person B's leg **had a surgery** last night.
Person A and Person B are **in the hospital.**
Person B **injured his right leg** when he was **playing baseball**.
…



SUGAR(Otani+23)
Parings of a conversation and situational statements.

Data:

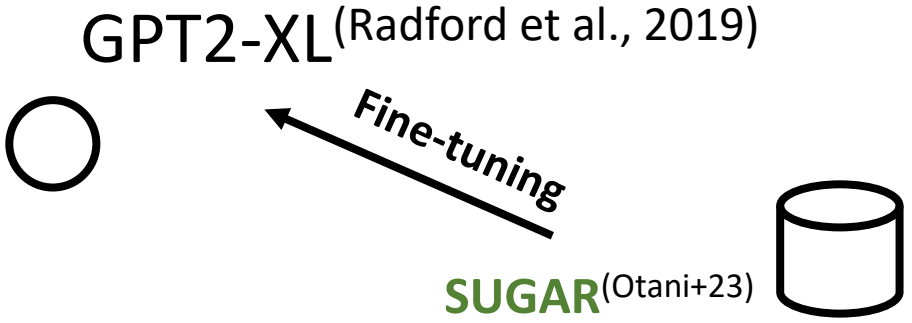
We created a dataset with situational statements semi-automatically.

Conv. history (taken from CICERO(Ghosal+22)):

A: Hi, Mike! how are you feeling now?
B: How did you know I was here? is it Tom?
A: I was talking with Bob yesterday and I learnt your right leg had been injured. How did it happen?
B: [System output]

Situation (our work)

Person B's leg **had a surgery** last night.
Person A and Person B are **in the hospital.**
Person B **injured his right leg** when he was **playing baseball**.
…

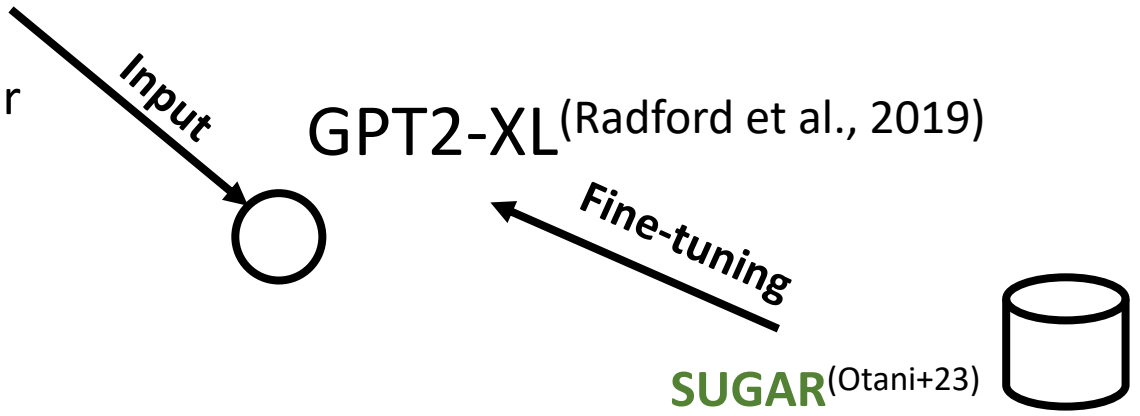


Data:

We created a dataset with situational statements semi-automatically.

Conv. history (taken from CICERO(Ghosal+22)):

A: Hi, Mike! how are you feeling now?
B: How did you know I was here? is it Tom?
A: I was talking with Bob yesterday and I learnt your right leg had been injured. How did it happen?
B: [System output]



Situation (our work)

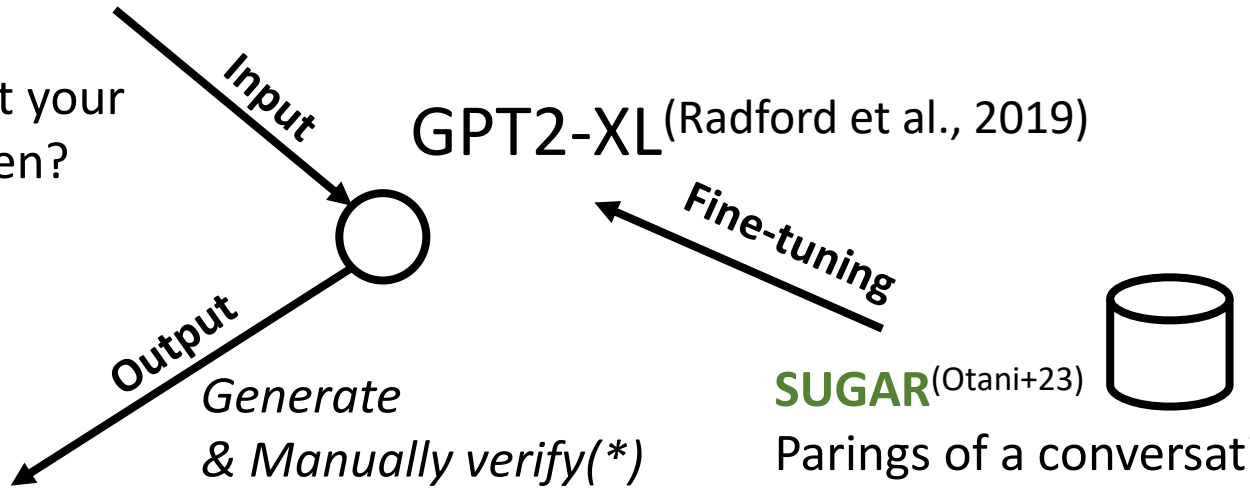
Person B's leg **had a surgery** last night.
Person A and Person B are **in the hospital.**
Person B **injured his right leg** when he was **playing baseball**.
…

Data:

We created a dataset with situational statements semi-automatically.

Conv. history (taken from CICERO(Ghosal+22)):

A: Hi, Mike! how are you feeling now?
B: How did you know I was here? is it Tom?
A: I was talking with Bob yesterday and I learnt your right leg had been injured. How did it happen?
B: [System output]



Situation (our work)

Person B's leg **had a surgery** last night.
Person A and Person B are **in the hospital.**
Person B **injured his right leg** when he was **playing baseball**.
…

Data:

We created a dataset with situational statements semi-automatically.

Conv. history (taken from CICERO(Ghosal+22)):

A: Hi, Mike! how are you feeling now?

B: How did you know I was here? is it Tom?

A: I was talking with E
right leg had been

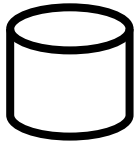
B: [System output]

	Training	Validation	Test	Avg. turn
SUGAR	1,214	102	25	1.0
CICERO	15,171	5,325	25	3.0
ConvAI2	16,878	1,000	25	4.7

ford et al., 2019)

-tuning

SUGAR(Otani+23)



Parings of a conversation and situational statements.

Situation (our wo

Person B's leg **had a surgery** last night.

Person A and Person B are **in the hospital.**

Person B **injured his right leg** when he was **playing baseball**.

…

& manually verify( )

**Test data only*

Generation and Evaluation:

Feeding situational knowledge in text as extra input

Task: Response generation

Generation and Evaluation:

Feeding situational knowledge in text as extra input

Task: Response generation

- **Systems:** BlendetBot2(Komeili+'22; Xu+'22) GPT-3(Brown+'20)

Generation and Evaluation:

Feeding situational knowledge in text as extra input

Task: Response generation

- **Systems:** BlendetBot2[(Komeili+'22; Xu+'22)] GPT-3[(Brown+'20)]

Generation and Evaluation:

Feeding situational knowledge in text as extra input

Task: Response generation

- **Systems:** BlendetBot2[(Komeili+'22; Xu+'22)] GPT-3[(Brown+'20)]

User request
+ Situational statements

- **Training/Generation:**

Generation and Evaluation:

Feeding situational knowledge in text as extra input

Task: Response generation

- **Systems:** BlendetBot2[(Komeili+'22; Xu+'22)] GPT-3[(Brown+'20)]

- **Training/Generation:**

  BB2: Fine-tuning on the dataset with and without situational statements.

Generation and Evaluation:

Feeding situational knowledge in text as extra input

Task: Response generation

- **Systems:** BlendetBot2[(Komeili+'22; Xu+'22)] GPT-3[(Brown+'20)]

- **Training/Generation:**
  BB2: Fine-tuning on the dataset with and without situational statements.
  GPT-3: In-context learning with 3 examples in a prompt w/ and w/o situ.
- **Evaluation:** Manual (crowdsourcing)
  - Dimensions [(Thoppilan+22; Zhou+22)]
    - Coherence
    - Context-specificity
    - Grammaticality

Generation and Evaluation:

Feeding situational knowledge in text as extra input

Task: Response generation

- **Systems:** BlendetBot2 (Komeili+'22; Xu+'22) GPT-3 (Brown+'20)

  User request
  + Situational statements → Generator → Response (top-2)

- **Training/Generation:**

  BB2: Fine-tuning on the dataset with and without situational statements.

  GPT-3: In-context learning with 3 examples in a prompt w/ and w/o situ.

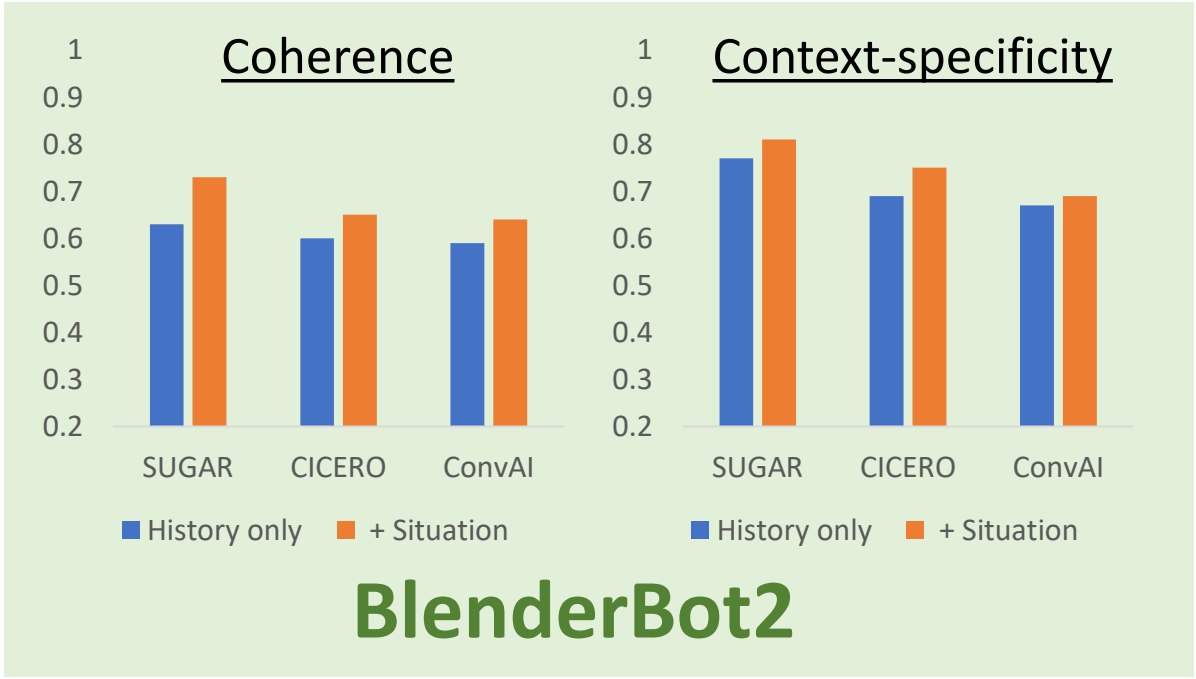
- **Evaluation:** Manual (crowdsourcing)
  - Dimensions (Thoppilan+22; Zhou+22)
    - Coherence
    - Context-specificity
    - Grammaticality

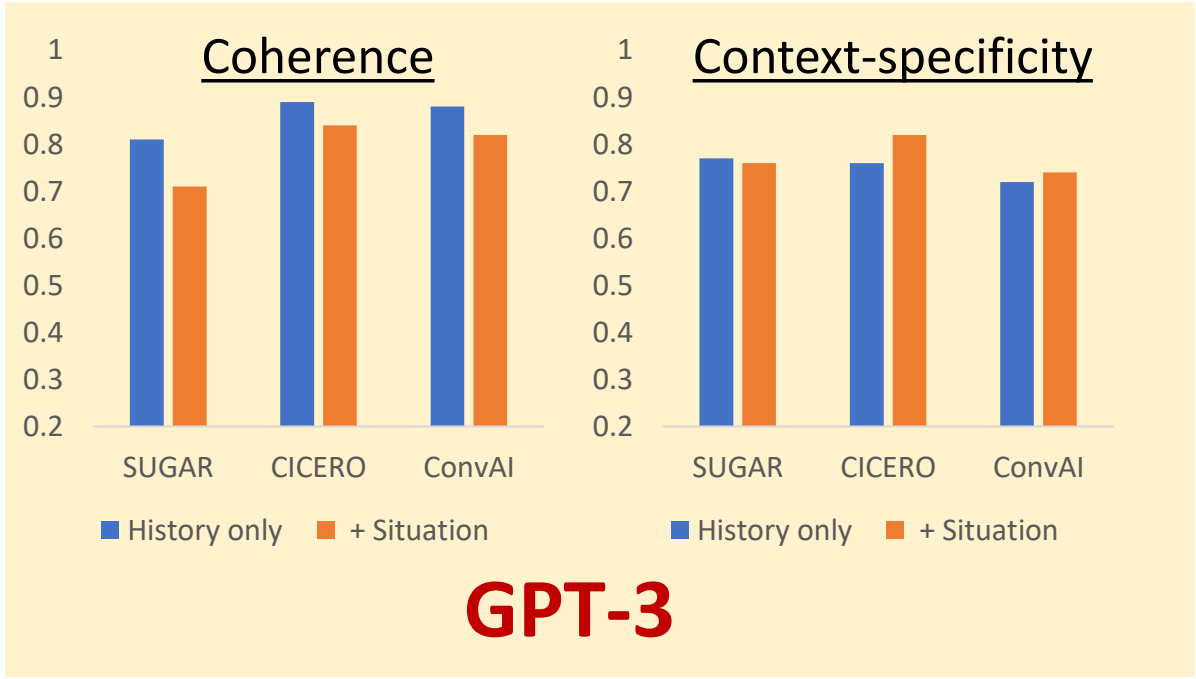
	Training	Validation	Test	Avg. turn
SUGAR	1,214	102	25	1.0
CICERO	15,171	5,325	25	3.0
ConvAI2	16,878	1,000	25	4.7

Results:

When situation is provided, BenderBot2(BB2)↑ and GPT-3↓

**Overall, BB2 < GPT-3*



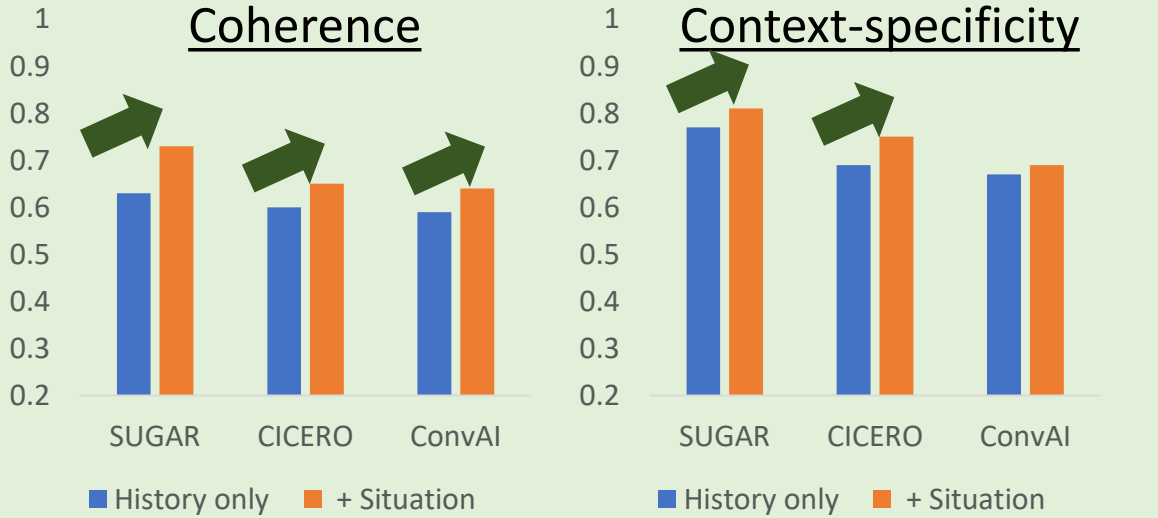


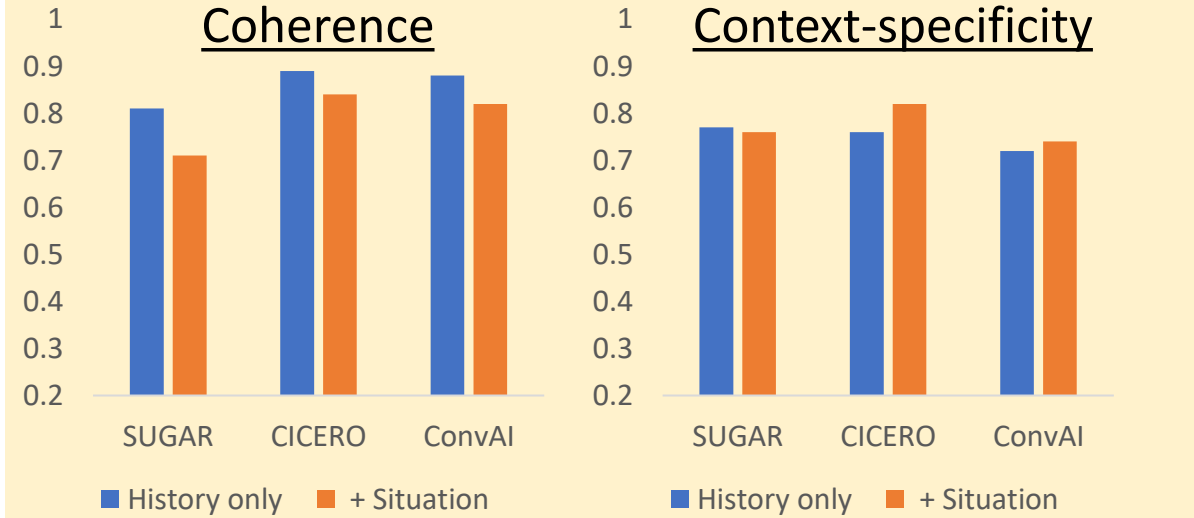
(No difference in grammaticality)

Results:

When situation is provided, BenderBot2(BB2)↑ and GPT-3↓

**Overall, BB2 < GPT-3*



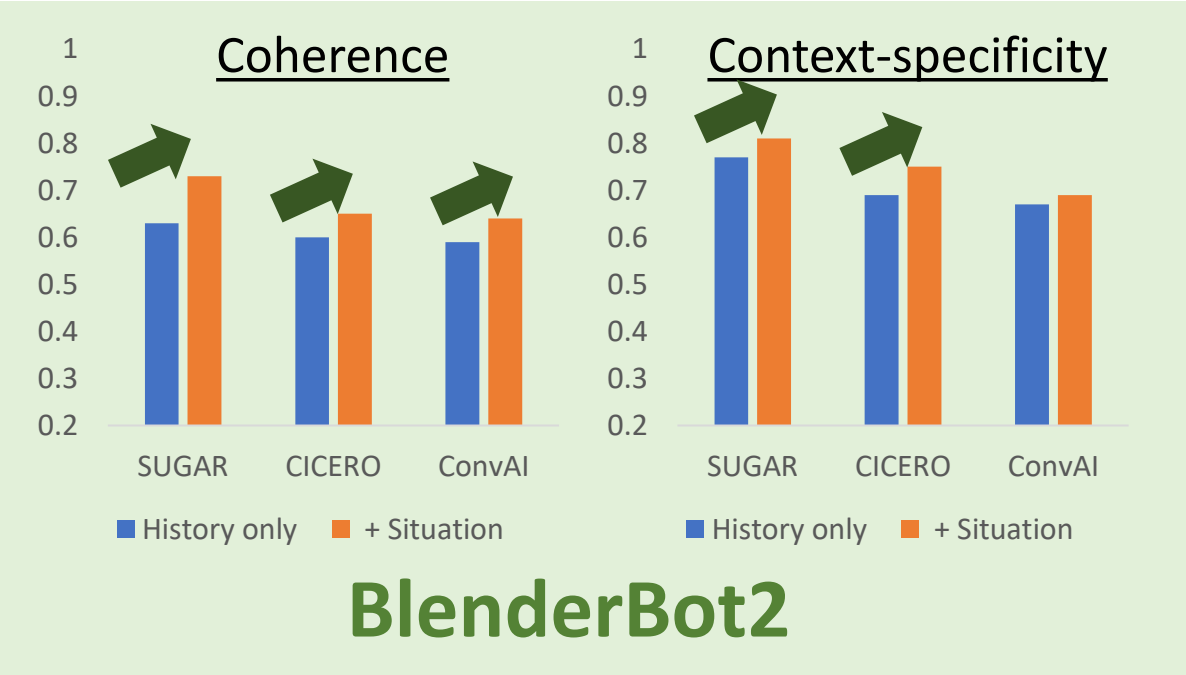


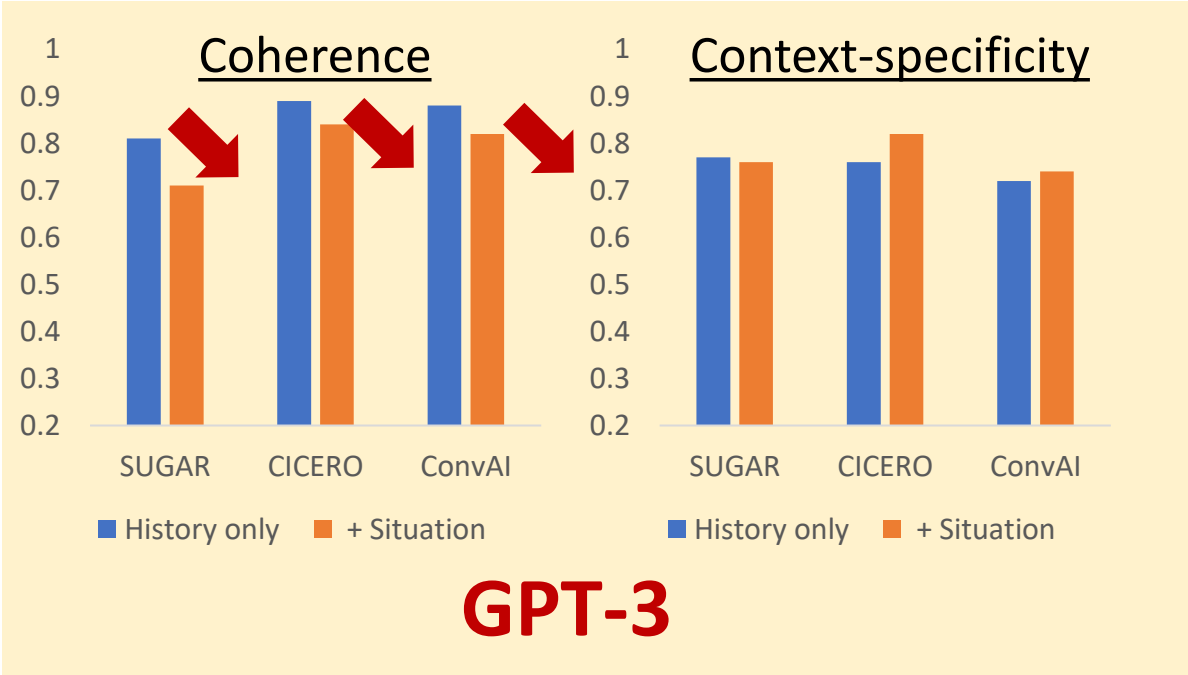
(No difference in grammaticality)

Results:

When situation is provided, BenderBot2(BB2)↑ and GPT-3↓

**Overall, BB2 < GPT-3*





(No difference in grammaticality)

Example:

BlenderBot2 improved with situational info., but GPT-3 didn't.

BlenderBot2:

Coherence↑ when situational statements were provided.

Example:

BlenderBot2 improved with situational info., but GPT-3 didn't.

BlenderBot2:

Coherence↑ when situational statements were provided.

"Could you help me take a bus?"

Example:

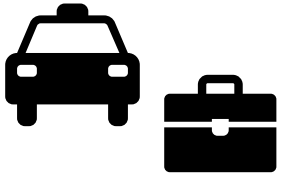
BlenderBot2 improved with situational info., but GPT-3 didn't.

BlenderBot2:

Coherence↑ when situational statements were provided.

"Could you help me take a bus?"

"Sure, shall I **book a taxi** as well? […] for you **to and from work.**"



Example:

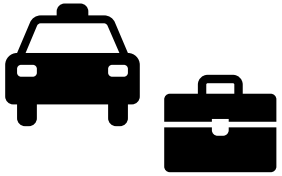
BlenderBot2 improved with situational info., but GPT-3 didn't.

BlenderBot2:

Coherence↑ when situational statements were provided.

"Could you help me take a bus?"

"Sure, shall I **book a taxi** as well? […] for you **to and from work.**"



+ Situation:

"The speaker has a bus schedule." … *(+ some distractors)*

Example:

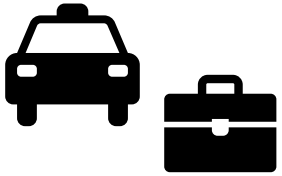
BlenderBot2 improved with situational info., but GPT-3 didn't.

BlenderBot2:

Coherence↑ when situational statements were provided.

"Could you help me take a bus?"

"Sure, shall I **book a taxi** as well? […] for you **to and from work.**"



+ Situation:

"The speaker has a bus schedule." … *(+ some distractors)*

"Sure, shall I pull up **your bus schedule?** […]

Example:

BlenderBot2 improved with situational info., but GPT-3 didn't.

GPT-3 was misled by distractors at times.

(Note: the overall quality was superior to BB2)

Example:

BlenderBot2 improved with situational info., but GPT-3 didn't.

GPT-3 was misled by distractors at times.

(Note: the overall quality was superior to BB2)

"Could you assist me with stretch?"

Example:

BlenderBot2 improved with situational info., but GPT-3 didn't.

GPT-3 was misled by distractors at times.

(Note: the overall quality was superior to BB2)

"Could you assist me with stretch?"

"Absolutely. Let me know which muscles you want to focus on and I can show you some stretches."

Example:

BlenderBot2 improved with situational info., but GPT-3 didn't.

GPT-3 was misled by distractors at times.

(Note: the overall quality was superior to BB2)

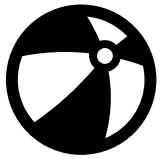
"Could you assist me with stretch?"

"Absolutely. Let me know which muscles you want to focus on and I can show you some stretches."

+ Situation:

"The speaker is wearing a jogging suit."

… "There is a beach ball at home." …



Example:

BlenderBot2 improved with situational info., but GPT-3 didn't.

GPT-3 was misled by distractors at times.

(Note: the overall quality was superior to BB2)

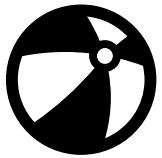
"Could you assist me with stretch?"

"Absolutely. Let me know which muscles you want to focus on and I can show you some stretches."

+ Situation:

"The speaker is wearing a jogging suit."

… "There is a beach ball at home." …



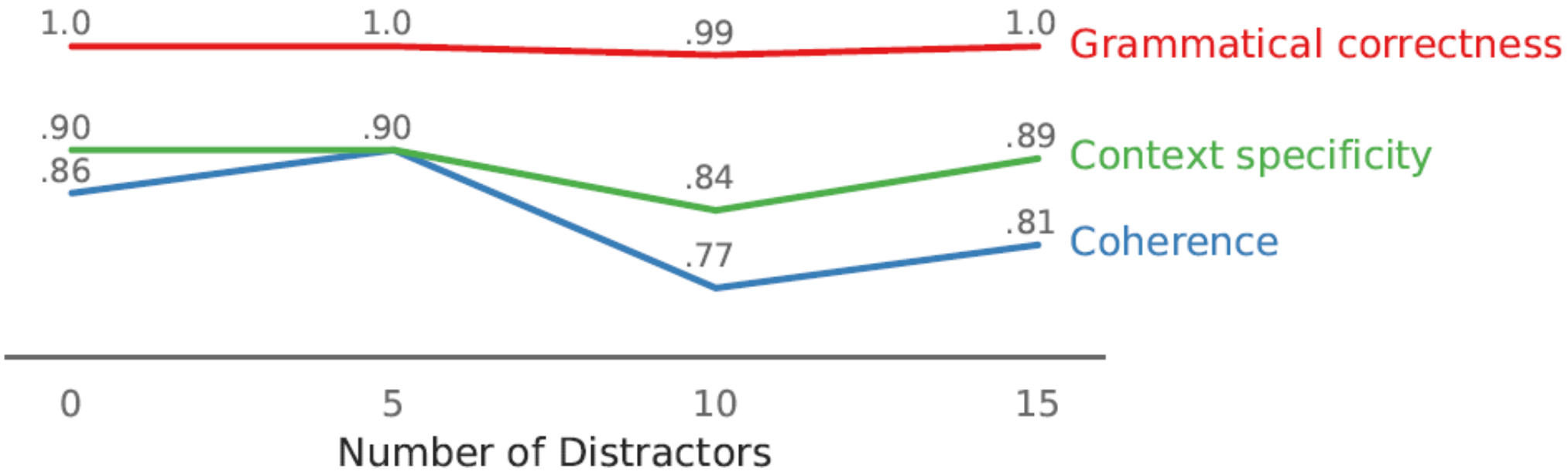
"Of course! Let's go to the beach and play with the beach ball while we stretch."

Controlled experiments:

GPT-3 was also misled by distractors

Evaluation on SUGAR

SUGAR has 5-15 irrelevant situational statements.



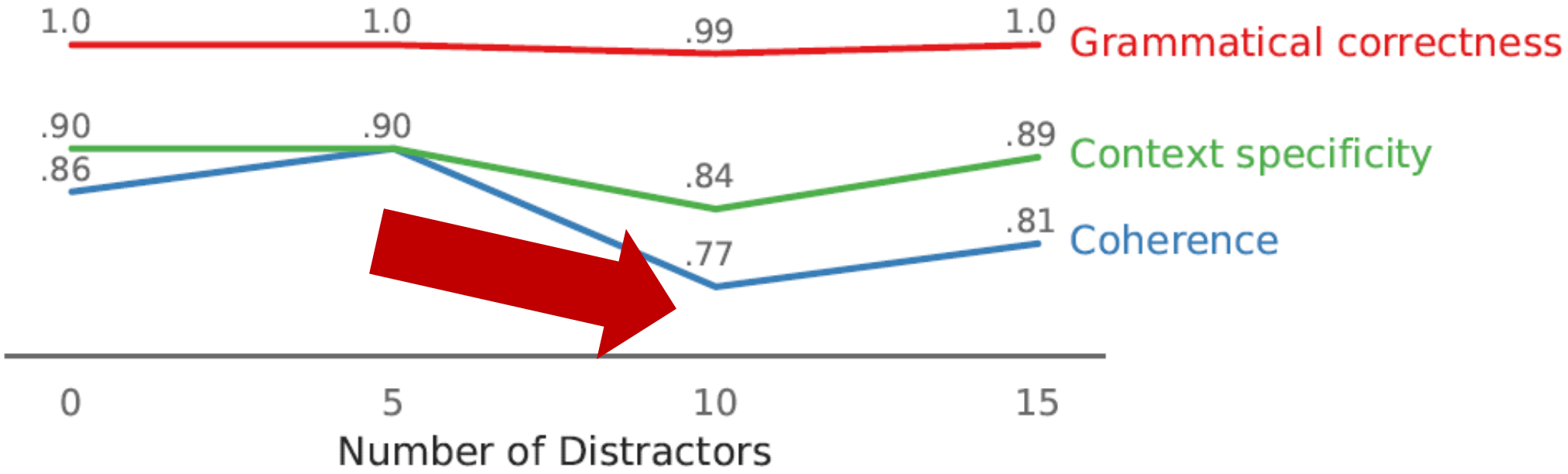
Opportunity for future research!

Controlled experiments:

GPT-3 was also misled by distractors

Evaluation on SUGAR

SUGAR has 5-15 irrelevant situational statements.



Opportunity for future research!

Conclusion and Future Work

Conclusion:

Conversations heavily rely on situation

Conclusion:

Conversations heavily rely on situation

Claim: We need to talk about situation in open-domain conversations!

Conclusion:

Conversations heavily rely on situation

Claim: We need to talk about situation in open-domain conversations!

- Many datasets don’t have situational information.

Conclusion:

Conversations heavily rely on situation

Claim: We need to talk about situation in open-domain conversations!

- Many datasets don't have situational information.
- **Problems:** (1) underspecification, (2) system behavior, and (3) evaluation

Conclusion:

Conversations heavily rely on situation

Claim: We need to talk about situation in open-domain conversations!

- Many datasets don't have situational information.
- **Problems:** (1) underspecification, (2) system behavior, and (3) evaluation
- Finding from our case study: situation is sometimes useful,
  but picking up relevant information is not easy → Room for future research!

Conclusion:

There are open challenges for future research

Conclusion:

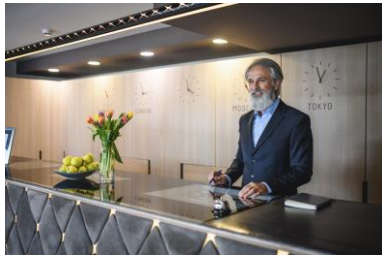
There are open challenges for future research

- **Representation:** A textual representation is NOT the best way to capture situation.
  Should go to multi-modal data (image, audio, sensory information, external APIs, etc.)

Conclusion:

There are open challenges for future research

- **Representation:** A textual representation is NOT the best way to capture situation.
  Should go to multi-modal data (image, audio, sensory information, external APIs, etc.)

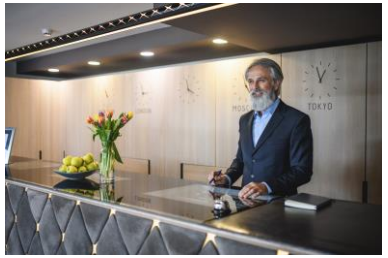






Conclusion:

There are open challenges for future research

- **Representation:** A textual representation is NOT the best way to capture situation.
  Should go to multi-modal data (image, audio, sensory information, external APIs, etc.)

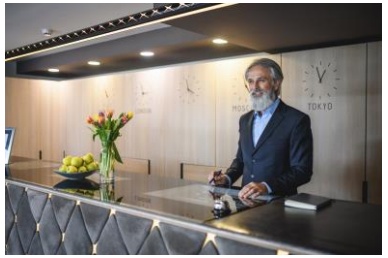



- **Adequacy:** Humans process an enormous amount of information.
  - How much should we put into the development/evaluation data?

Conclusion:

There are open challenges for future research

- **Representation:** A textual representation is NOT the best way to capture situation.
  Should go to multi-modal data (image, audio, sensory information, external APIs, etc.)




- **Adequacy:** Humans process an enormous amount of information.
  - How much should we put into the development/evaluation data?
- **Resource:** Human annotation is costly.
  - Maybe language generation models can facilitate annotation work.

Conclusion:

There are open challenges for future research

Conclusion:

There are open challenges for future research

Common ground: A one-step deeper problem behind situation.

Conclusion:

There are open challenges for future research

Common ground: A one-step deeper problem behind situation.

Conv. history (taken from CICERO(Ghosal+22)):

A: Hi, Mike! how are you feeling now?
B: How did you know I was here? is it Tom?
A: I was talking with Bob yesterday and I learnt your right leg had been injured. How did it happen?
B: [System output]

Person B

Person A

Conclusion:

There are open challenges for future research

Common ground: A one-step deeper problem behind situation.

Conv. history (taken from CICERO(Ghosal+22)):

A: Hi, Mike! how are you feeling now?
B: How did you know I was here? is it Tom?
A: I was talking with Bob yesterday and I learnt your right leg had been injured. How did it happen?
B: [System output]

Person B: *I have these observations.*

Person B

Situation:

Person B's leg **had a surgery** last night.
Person A and Person B are **in the hospital.**
Person B **injured his right leg** when he was **playing baseball**.
…

Person A

Conclusion:

There are open challenges for future research

Common ground: A one-step deeper problem behind situation.

Conv. history (taken from CICERO(Ghosal+22)):

A: Hi, Mike! how are you feeling now?
B: How did you know I was here? is it Tom?
A: I was talking with Bob yesterday and I learnt your right leg had been injured. How did it happen?
B: [System output]

Person B: *I have these observations.*

How much is shared with Person A?

Person B

Situation:

Person B's leg **had a surgery** last night.
Person A and Person B are **in the hospital.**
Person B **injured his right leg** when he was **playing baseball**.
…

Person A

A (partial) set of **observable** information about objects in the scene

Claim:

We need to talk about **situation** for enhancing open-domain conversational agents!

On the Underspecification of Situations in Open-domain Conversational Datasets

Naoki Otani naoki@megagon.ai

Jun Araki, HyeongSik Kim, and Eduard Hovy