NASSLLI @ USC 2022

Multiparty and Multi-floor dialogue structure
Lecture 4.1: Multifloor Dialogue Structure

David Traum
Institute for Creative Technologies
University of Southern California
traum@ict.usc.edu
https://people.ict.usc.edu/~traum/
Outline

• Annotation Review
• Multifloor Dialogue Structure Annotation Scheme
• Botlanguage Project
• Botlanguage Annotations
• Context
• “Stop!”
• Conclusions
Dialogue Structure Annotation for Multi-Floor Interaction

David Traum, Cassidy Henry, Stephanie Lukin, Ron Artstein, Felix Gervitz, Kimberly A. Pollard, Claire Bonial, Su Lei, Clare R. Voss, Matthew Marge, Cory J. Hayes, Susan G. Hill
Types of Dialogue Structure (Traum & Nakatani 1999)

**Structure Content**
- Intentional
- Linguistic
- Relational/Rhetorical
- Attentional State
- Turn-taking/floor management
- Grounding
- Participant structure

**Structure Granularity**
- Micro – within a single turn
- Meso – short subdialogue
- Macro – full conversation
Multi-floor Botlanguages Annotations: Meso-level Dialogue Structure

**Structure Types**

- **Intentional:**
  - *Transaction Units* – smallest unit of specified and performed action, including all dialogue needed to accomplish this

- **Relational/Rhetorical:**
  - *Relations* between utterances within a transaction

**Annotations**

- **TUs:** cluster of utterances
  - Not necessarily sequential

- **Relations:** Label 2\(^{nd}\) part utterance with
  - Antecedent
  - Relation type
Example:

• **Customer:** I’d like a cheeseburger
• **Waiter:** one cheeseburger.
• **Waiter:** (placing burger in bag) here you go.
• **Customer:** thanks!
• **Waiter:** would you like fries with that?
• **Customer:** Sure, a large one please!
• **Waiter:** (placing fries box in bag): one large fries.
1. **Customer**: I’d like a cheeseburger
2. **Waiter**: one cheeseburger.
3. **Waiter**: (placing burger in bag) here you go.
4. **Customer**: thanks!
5. **Waiter**: would you like fries with that?
6. **Customer**: Sure, a large one please!
7. **Waiter**: (placing fries box in bag): one large fries.

**Transaction Unit (TU)**: a group containing the initiation and (potential) fulfillment of an intent.

---

**Traum et al. 2018, LREC**
1. multi-floor dialogue: 2018 annotation schema

**Relations**: describe the structure between pairs of utterances within a TU
- **Antecedent**: the utterance that a subsequent utterance is addressing (e.g., 2 1)
- **Relation-Type**: relationship between utterance and antecedent (e.g., Acknowledgment)

1. **Customer**: I’d like a cheeseburger
2. **Waiter**: one cheeseburger.
3. **Waiter**: (placing burger in bag) here you go.
4. **Customer**: thanks!
5. **Waiter**: would you like fries with that?
6. **Customer**: Sure, large please!
7. **Waiter**: (placing fries in bag): one large fries.

*Traum et al. 2018, LREC*
Floor and Participant Structure

Participants and Floors

• Single floor Dyadic (A,B)
• Single floor Multiparty: (A,B,C,...)
• Multiple floors (with different sets of participants): \{(A B), (C D E)\}

Interactions between Floors

– Same purpose, distinct participants
– Co-located, observable
  • Participants play different roles for different floors (e.g. active participant vs overhearer)
– Some Shared participant(s)
  • multi-communicating (Rentch et al)
– Multi-floor dialogue:
  • Same purpose
  • Some Multi-communicating participant(s)
  • Content flows across floors
1. Multi-floor dialogue: introduction

Conversational floor: shares common set of speakers and observers

Multi-floor Dialogue: high-level dialogue purposes are the same, and some content is shared, but other aspects (participant structure, turn-taking expectations) are distinct.

Traum et al. 2018, LREC
1. **Multi-floor dialogue: introduction**

Conversational floor: shares common set of speakers and observers

Multi-floor Dialogue: high-level dialogue purposes are the same, and some content is shared, but other aspects (participant structure, turn-taking expectations) are distinct

End (final exchange)

Traum et al. 2018, LREC
1. **multi-floor dialogue: 2018 annotation schema**

**Relation Super-Types**

**Expansions** - relate utterances produced by the *same* participant within the *same* floor (4 Subtypes)

**Responses** - relate utterances by *different* participants in the *same* floor (24 Subtypes)

**Translations** - relate utterances in *different* floors (4 Subtypes)

**Example Subtypes:**

1. Customer → Waiter: I’ll have a cheeseburger
2. Customer → Waiter: and a small coke

---

1. Customer → Waiter: a small coke, please
2. Waiter → Customer: no coke, pepsi

---

1. Customer → Waiter: I’ll have a cheeseburger
2. Waiter → Cook: Cheeseburger!!
## Relations by type (1)

### Expansions
a) Continue  
b) (self-) Correction  
c) Link-next  
d) Summarization

### Translation
a) Translation<from,to>  
b) Partial  
c) Quotation  
d) Comment
Relations by type (2) Responses

a. **Processing:** positive feedback at perception level

b. **acknowledgement:** positive feedback of understanding

c. **clarification:** negative feedback of understanding

d. **question-response**

e. **reciprocal response:** e.g. ”hello” -> “hello”

f. **3rd turn feedback:** response to feedback

g. **other**
Response sub-relations

acknowledgment
• ack-done
• ack-doing
• ack-wilco
• ack-understand
• ack-try
• ack-unsure
• ack-cant

clarification
• req-clar
• clar-repair
• missing info
• nack
• req-repeat
• clar-repeat

question-response
• answer
• Non-Answer-Response (NAR)
Domain: Human-Robot Collaboration

Remote reconnaissance task
- Unfamiliar environment
- Bandwidth limitations
- User and robot not co-present

- What would the human users want to say?
  - Need to collect a corpus in order to train and evaluate the system.
- How would users naturally collaborate with this robot teammate?

(Marge et al., 2016, IEEE RO-MAN)
2. situated dialogue: human-robot dialogue corpus

What the Participant sees

- static image
- dynamic map
- messages from the robot

How far forward should I move?

What the Participant says

speaks verbal instructions

Travel forward down the hall

“Commander” Participant

Situated Corpus of Understanding Transactions: Marge et al., RO-MAN 2016
2. situated dialogue: human-robot dialogue corpus

WHAT THE PARTICIPANT SEES

last photo sent by Robot

Robot’s replies

Robot’s LIDAR map of searched area (LIDAR: Light Detection And Ranging)
2. situated dialogue: human-robot dialogue corpus

“Commander” Participant

What the participant says

speaks verbal instructions

Travel forward down the hall

What the participant sees

messages from the robot

How far forward should I move?

static image dynamic map

“Behind the scenes”

Dialogue Manager Wizard (DM)

The “brains” of the robot in the natural language interactions. May need to request clarification of participant.

Situated Corpus of Understanding Transactions:
Marge et al., RO-MAN 2016
2. situated dialogue: human-robot dialogue corpus

What the participant says

speaks verbal instructions

Go to the first door on the left

What the participant sees

messages from the robot

done

static image  dynamic map

“Behind the scenes”

Dialogue Manager Wizard (DM)

The “brains” of the robot in the natural language interactions. Transforms speech into constrained language

Robot Navigator Wizard (RN)

Navigates the robot based on the constrained instructions from the Dialogue Manager Wizard

Situated Corpus of Understanding Transactions:
Marge et al., RO-MAN 2016
Example Interaction

DM

RN

[Image of a person wearing a headset, an icon of a person sitting at a computer, an icon of a person using a mobile device, and a robotic device]
Commander

- Verbally instructs a robot
- Sees text message responses, LIDAR map, and images sent from onboard robot
Wizard #1 – Dialogue Manager

- Handles all language functions of "robot"
- Responds to CMD and robot navigator (RN) via text message
- Serves as mediator between RN and CMD
Wizard #2 – Robot Navigator

- Handles all navigation function of "robot"
- Constrained language received
- Separation of wizards:
  - reduces cognitive load/wizard labor
  - removes intuition of interpreting commands
Example Interaction

Proceed forward
Example Interaction

How far? You can tell me to move to an object that you see, or a distance
Example Interaction

DM

RN

Proceed forward three feet
Example Interaction

Executing…
Example Interaction

move forward three feet
Example Interaction

*moves robot forward 3 feet*
Example Interaction

DM

RN

done
Example Interaction

DM → RN

done
Data - Transcripts

- Time aligned transcripts of 4 data streams
  - 2 audio streams
    - CMD and RN
  - 2 text streams
    - DM->CMD, DM->RN
- Two conversational floors present

<table>
<thead>
<tr>
<th>Commander (Audio Stream 1)</th>
<th>DM-&gt;Commander (Chat Room 1)</th>
<th>DM-&gt;RN (Chat Room 2)</th>
<th>RN (Audio Stream 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>face the doorway on your right</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>and take a picture</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>there’s a door ahead of me on the right and one just behind me on the right. which would you like me to face?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>the door ahead of you on the right</td>
<td>move to face the door ahead of you on the right, image</td>
<td></td>
<td></td>
</tr>
<tr>
<td>executing...</td>
<td></td>
<td></td>
<td>image sent</td>
</tr>
<tr>
<td>sent</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
face the doorway on your right

and take a picture

there’s a door ahead of me on the right and one just behind me on the right. which would you like me to face?

the door ahead of you on the right

move to face the door ahead of you on the right, image executing...

sent
Left floor: CMD, DM

Commander (Audio Stream 1)

DM - Commander (Chat Room 1)

DM - RN (Chat Room 2)

RN (Audio Stream 2)

face the doorway on your right

and take a picture

there’s a door ahead of me on the right and one just behind me on the right. Which would you like me to face?

the door ahead of you on the right

move to face the door ahead of you on the right, image

executing...

image sent

sent
Right Floor: DM, RN

 Commander

 DM

 face the doorway on your right

 RN

 there’s a door ahead of me on the right and one just behind me on the right. which would you like me to face?

 DM

 image executing...

 RN

 execute the door ahead of you on the right, image

 image sent
DM translates (to) left and right

Commander (Audio Stream 1)  DM->Commander (Chat Room 1)  DM->RN (Chat Room 2)  RN (Audio Stream 2)

face the doorway on your right

and take a picture

there's a door ahead of me on the right and one just behind me on the right. which would you like me to face?

the door ahead of you on the right

move to face the door ahead of you on the right, image executing...

image sent

sent
2. situated dialogue: human-robot dialogue corpus

move forward

You can tell me to move a certain distance or to move to an object

move forward 3 feet

processing. . .

processing. . .

moving. . .

move forward 3 feet

done

done

behind the scenes…
2. situated dialogue: human-robot dialogue corpus

<table>
<thead>
<tr>
<th>ID</th>
<th>Participant (Audio Stream 1)</th>
<th>DM -&gt; Participant (Chat Room 1)</th>
<th>DM -&gt; RN (Chat Room 2)</th>
<th>RN (Audio Stream 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>move forward</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>processing. . .</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>You can tell me to move a certain distance or to move to an object</td>
<td></td>
<td>move forward 3 feet</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>go forward 3 feet</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>processing. . .</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>moving. . .</td>
<td>move forward 3 feet</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td>done</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>what do you see</td>
<td>send image</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>sent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>sent</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2. situated dialogue: training data strengths and weaknesses

- Training data establishes bi-directional associations between NL, execution behavior
2. situated dialogue: training data strengths and weaknesses

ScoutBot demo: Lukin et al, ACL 2018
Corpus Statistics

Basics
- **60** dialogues
  - **20** participants
  - **3** dialogues each
  - **~20** hours
- **11454** Total Utterances
  - **3,573** from commanders
  - **5,154** from DM
  - **2,727** from RN

Dialogue Structure Annotations
- **2,230** Transaction Units
- **11,058** Relations
- **644** Unique TU Tree structures
  - Classified into 5 types
## Frequent Relations

<table>
<thead>
<tr>
<th>Type</th>
<th>Subtype</th>
<th>#</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translation</td>
<td>Translate-r</td>
<td>2355</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>Translate-l</td>
<td>1911</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>comment</td>
<td>21</td>
<td>&lt;1</td>
</tr>
<tr>
<td>Expansion</td>
<td>Continue</td>
<td>1175</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Link-next</td>
<td>337</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>correction</td>
<td>50</td>
<td>&lt;1</td>
</tr>
<tr>
<td></td>
<td>summarize</td>
<td>20</td>
<td>&lt;1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type</th>
<th>Subtype</th>
<th>#</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response</td>
<td>acknowledge</td>
<td>3998</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>clarification</td>
<td>569</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>processing</td>
<td>315</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Question-response</td>
<td>212</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>other</td>
<td>48</td>
<td>&lt;1</td>
</tr>
<tr>
<td></td>
<td>3rd turn feedback</td>
<td>37</td>
<td>&lt;1</td>
</tr>
<tr>
<td></td>
<td>reciprocal</td>
<td>14</td>
<td>&lt;1</td>
</tr>
</tbody>
</table>
Structural Types of Transaction Units (TUs)

- **Minimal TU**: single instruction, acks, no repair
- **Extended-Link TU**: multiple instructions, with expansions
- **Repair TU**: contains at least one repair
  - successfully resolved or
  - abandoned
- **QA TU**: starts with question & response rather than instruction
  - simple question,
  - later instruction
- **Other TU**: none of the above (e.g. no response or translation)
## Example minimal TU

<table>
<thead>
<tr>
<th>Utt #</th>
<th>Commander</th>
<th>DM→CMD</th>
<th>DM→RN</th>
<th>RN</th>
<th>TU #</th>
<th>Antecedent</th>
<th>Relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>move forward three feet</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>ok</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>ack-wilco</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>move forward 3 feet</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>translation-r</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>done</td>
<td></td>
<td>1</td>
<td>3</td>
<td>ack-done</td>
</tr>
<tr>
<td>5</td>
<td>I moved forward 3 feet</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>4</td>
<td>translation-l</td>
</tr>
</tbody>
</table>
# Example Extended-Link TU

<table>
<thead>
<tr>
<th>Utt #</th>
<th>Left Floor</th>
<th>Right Floor</th>
<th>Annotations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Commander</td>
<td>DM→CMD</td>
<td>DM→RN</td>
</tr>
<tr>
<td>1</td>
<td>face west</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>and take a photo</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>face west, photo</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>executing...</td>
<td>1</td>
<td>2*</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td>image sent</td>
</tr>
<tr>
<td>6</td>
<td>sent</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Example Q&A TUS

<table>
<thead>
<tr>
<th>Left Floor</th>
<th>Right Floor</th>
<th>Annotations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Utt #</strong></td>
<td><strong>Commander</strong></td>
<td><strong>DM→Commander</strong></td>
</tr>
<tr>
<td>1</td>
<td>how many window openings do you see in front of you</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>three</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>do you see a yellow flashlight</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>processing...</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>I’m not sure</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>If you describe an object, you can help me to learn what it is.</td>
<td></td>
</tr>
</tbody>
</table>
## Example Other TUS

<table>
<thead>
<tr>
<th>Utt #</th>
<th>Left Floor Commander</th>
<th>DM→Commander</th>
<th>Right Floor DM→RN</th>
<th>RN</th>
<th>Annotations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>i'm ready</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>I'm also ready</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Would you like me to send a picture so you can see the room?&quot;</td>
<td></td>
<td></td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>
Frequency of TU Structures (% of corpus)

- Minimal TU (48%)
- Extended-Link TU (26%)
- Repair TU (11%)
  - 9% successfully resolved
  - 2% abandoned
- QA TU (~5)%
  - 4% simple question
  - 1% lead to instruction
- Other TU (11%)
Applications of Annotated Data

- Examination of Dialogue Structure Overlap (Henry et al WiNLP 2018)
- Stylistic differences across individuals and conditions (Lukin et al Sigdial 2018)
- Automating NLU and dialogue management (Gervits et al ACL 2018 Demo)
Context Is Key:
Annotating Situated Dialogue Relations in Multi-floor Dialogue

- Claire Bonial¹,
- Mitchell Abrams², Anthony L. Baker³, Taylor Hudson³, Stephanie M. Lukin¹,
  - David Traum⁴ & Clare R. Voss¹

¹ U.S. Army DEVCOM ARL
² Institute for Human and Machine Cognition (IHMC)
³ Oak Ridge Associated Universities (ORAU)
⁴ USC Institute for Creative Technologies
Goal: Extend our *multi-floor* dialogue annotation schema to account for features of *situated* dialogue—interpretation draws upon info from physical environment, conversational history, robot’s physical form, etc.

Picture Credit: Knepper et al. 2015
2. situated dialogue: human-robot dialogue corpus

“go to the wall behind you, face north and then take a picture”
2. situated dialogue: human-robot dialogue corpus

"back up"
2. situated dialogue: human-robot dialogue corpus

"take a picture"
2. situated dialogue: human-robot dialogue corpus

“go to the other door”
2. situated dialogue: training data strengths and weaknesses

- Some associations between NL, execution behavior are only valid in particular certain situated contexts.
Modifications to annotation schema

Relation Super-Types
Expansions - relate utterances produced by the same participant within the same floor (4 Subtypes)

Responses - relate utterances by different participants in the same floor (24 Subtypes) (26 Subtypes)
Translations - relate utterances in different floors (4 Subtypes) (10 Subtypes)

Example Subtypes:
1. Customer→Waiter: I’ll have a cheeseburger
2. Customer→Waiter: and a small coke

Traum et al. 2018, LREC
3. Schema extensions: landmark and direct translation extensions

<table>
<thead>
<tr>
<th>TU</th>
<th>ID</th>
<th>Participant (Audio Stream 1)</th>
<th>DM -&gt; Participant (Chat Room 1)</th>
<th>DM -&gt; RN (Chat Room 2)</th>
<th>RN (Audio Stream 2)</th>
<th>Antecedent</th>
<th>Relation-Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>go through the doorway directly in front of you</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>and take a photo</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>processing...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>move into Conf Room</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>processing-landmark</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>then...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>link-next</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>send image</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>translation-r-direct</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>moving...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ack-doing</td>
</tr>
<tr>
<td>1</td>
<td>8</td>
<td>uh done and sent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ack-done</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
<td>done, sent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Direct Translations: Uses the same or synonymous words, where the translation is applicable in any physical or conversational context.

Landmark Translations: Refers to a unique landmark name known only to members of the right floor.
3. Schema extensions: situated and default translation extensions

<table>
<thead>
<tr>
<th>TU</th>
<th>ID</th>
<th>Participant (Audio Stream 1)</th>
<th>DM -&gt; Participant (Chat Room 1)</th>
<th>DM-&gt; RN (Chat Room 2)</th>
<th>RN (Audio Stream 2)</th>
<th>Antecedent</th>
<th>Relation-Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>turn east ninety degrees</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>and travel three feet</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>processing...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>turn left 90 degrees</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>translation-r-situated</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>then...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>link-next</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>move forward 3 feet</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>translation-r-default</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>turning...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ack-doing</td>
</tr>
<tr>
<td>1</td>
<td>8</td>
<td>moving...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>8</td>
<td>done</td>
<td></td>
<td></td>
<td>done</td>
<td></td>
<td>ack-done</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
<td>done</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Situated Translations:** Synonymous with original instruction only in the current physical context, but does not specify a unique landmark.

**Default Translations:** Supplements information by relying on some default assumption related to a robot behavior or capability.
3. Schema extensions:
History translation extension

<table>
<thead>
<tr>
<th>TU</th>
<th>ID</th>
<th>Participant (Audio Stream 1)</th>
<th>DM -&gt; Participant (Chat Room 1)</th>
<th>DM -&gt; RN (Chat Room 2)</th>
<th>RN (Audio Stream 2)</th>
<th>Antecedent</th>
<th>Relation-Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td></td>
<td>You often ask for images at the end of movement instructions. Should I send one each time?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>offer-accept</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>back up five feet</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>translation-r-direct</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td></td>
<td>back up 5 feet</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td></td>
<td>send image</td>
<td></td>
<td></td>
<td></td>
<td>translation-r-history</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td></td>
<td>executing...</td>
<td></td>
<td></td>
<td></td>
<td>ack-doing</td>
</tr>
</tbody>
</table>

**History Translations:** All or part of the translation is only relevant given the dialogue history, in which it was established that a certain instruction should be interpreted in a particular way.
3. Schema extensions: contextual translation, preparatory acknowledgments

<table>
<thead>
<tr>
<th>TU</th>
<th>ID</th>
<th>Participant (Audio Stream 1)</th>
<th>DM -&gt; Participant (Chat Room 1)</th>
<th>DM-&gt; RN (Chat Room 2)</th>
<th>RN (Audio Stream 2)</th>
<th>Antecedent</th>
<th>Relation-Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>take a picture of the wall on your left</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>processing. . .</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td></td>
<td>move to left wall</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td></td>
<td>send image</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>moving. . .</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>9</td>
<td>done, sent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Underspecified, Contextual Translations:** Draws upon situational or conversational context, but precisely what contextual information is being used is unclear, underspecified, or there are two or more factors.

**Acknowledgment – Doing/Will-Comply Preparation:** Speaker understands the command and a preparation step required for compliance with the command is underway (doing) or will be done (will-comply).
4. Annotated corpus: Analysis of type frequencies

Corpus Overview:
- 168 human-robot dialogues annotated, validated
- Total of 40,873 relations

Super-Type Relations Frequencies:
- 36.4% Acknowledgment super-type
- 36.5% Translation super-type

New Relation Types Frequencies:
- 70% Direct Translations (no situated language)
- 30% have situated language
- New preparatory acknowledgments have small, but critical impact

Table 5: Frequencies and % of updated relations.
4. Annotated corpus: Inter-annotator agreement

Reliability measured through Inter-Annotator Agreement (IAA)

- Comparable or higher IAA than original, unmodified schema
- New annotation categories are clearly identifiable

<table>
<thead>
<tr>
<th>Markable Type</th>
<th>Agreement</th>
<th>Distance Metric</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unmodified Schema</td>
<td>Modified Schema</td>
</tr>
<tr>
<td>Antecedents</td>
<td>0.72–0.82</td>
<td>0.79–0.94</td>
</tr>
<tr>
<td>Relation Types</td>
<td>0.77–0.89</td>
<td>0.83–0.93</td>
</tr>
<tr>
<td>Transaction Units</td>
<td>0.48–0.93</td>
<td>0.65–0.85</td>
</tr>
</tbody>
</table>

\(^a\)Krippendorff (1980) \(^b\)Passonneau (2006)

Table 6: IAA of the original, unmodified schema of Traum et al. (2018) and our modified schema.
5. Conclusions & future work

Extended multi-floor, multi-party dialogue structure annotations to uniquely mark situated dialogue

- Prevents inappropriate associations of NL-execution behavior in training data
- Critical step in exploration of how to relate and ground language to the context

Ongoing: Bringing together dialogue structure & propositional content with Dialogue-AMR
Making Sense of “STOP”

• Some meanings of saying “STOP” to another
  1. Action – slow down and halt motion
  2. Pause - Cease execution of current action (temporarily)
  3. Cancel current action
  4. Cancel future action
Possible Responses to “stop”

A. perform a “stopping” action, to terminate current velocity
B. halt current execution of an action (and later do something unrelated)
C. pause current execution of an action (and resume the action later)
D. pause execution and resume a slightly altered action after a correction has been specified
E. ignore the command as redundant with what has already been done (or already planned to do)
F. explain or request clarification when the command seems inappropriate or unclear
G. refrain from repeating a previous or current action (that might or might not currently be planned to do again)
Temporal Positioning of “Stop” in instruction-execution-grounding sequence

1. As part of the initial instruction, prior to beginning execution
2. During the grounding of the instruction; for example, when the operator has specified part but not all of the instruction sequence, or if the addressee is clarifying, questioning, or negotiating some aspects, such as a termination point or manner
3. After the instruction has been given and accepted, but before execution has begun
4. During execution, when part has been performed and part remains unperformed
5. After execution, but before the conversational partners ground the fact that execution has (successfully or unsuccessfully) terminated
6. After it has been agreed that the action has terminated (seems unrelated to the instruction, perhaps like the first case, relating to a new instruction)
Temporal Positioning of “Stop” in In Botlanguage Multi-floor dialogue

1. As part of the initial instruction, prior to beginning execution
2. During the grounding of the instruction;
3. After the instruction has been given, but before execution has begun
   3.1 After the instruction has been given by commander but before it has been translated by DM to the RN
   3.2 After the instruction has been translated by DM, but before it has been carried out by RN
4. During execution, when part has been performed and part remains unperformed
5. After execution, but before the conversational partners ground the fact that execution has (successfully or unsuccessfully) terminated
   5.1 After the RN finished executing, but before the RN has acknowledged completion
   5.2 After the RN has acknowledged completion but before the DM has translated the completion.
6. After it has been agreed that the action has terminated (seems unrelated to the instruction, perhaps like the first case, relating to a new instruction)
### Example: unsuccessful execution

<table>
<thead>
<tr>
<th>Floor 1</th>
<th>Floor 2</th>
</tr>
</thead>
<tbody>
<tr>
<td># User</td>
<td>DM &gt; User</td>
</tr>
<tr>
<td>82</td>
<td>um go &lt;pause. 33&gt;go straight</td>
</tr>
<tr>
<td>83</td>
<td></td>
</tr>
<tr>
<td>84</td>
<td>five feet</td>
</tr>
<tr>
<td>85</td>
<td></td>
</tr>
<tr>
<td>86</td>
<td></td>
</tr>
<tr>
<td>87</td>
<td>okay stop</td>
</tr>
<tr>
<td>88</td>
<td></td>
</tr>
<tr>
<td>89</td>
<td></td>
</tr>
<tr>
<td>90</td>
<td></td>
</tr>
<tr>
<td>91</td>
<td>go east, go east five feet</td>
</tr>
</tbody>
</table>

Table 3: Evidence from unsuccessful execution of the original instruction (line 89) and re-use of structure of the original instruction indicates “stop” precedes a *change* repair strategy.
<table>
<thead>
<tr>
<th>Time</th>
<th>Action</th>
<th>Coordinates</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>25x</td>
<td>robot turn forty five degrees right and continue to second doorway</td>
<td>279.71, 286.29</td>
<td>8</td>
</tr>
<tr>
<td>26</td>
<td>turn 45 right,</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>26</td>
<td>continue</td>
<td></td>
<td>25</td>
</tr>
<tr>
<td>27A</td>
<td>move forward to second doorway</td>
<td>298.25, 304.15</td>
<td>8</td>
</tr>
<tr>
<td>27B</td>
<td>executing...</td>
<td></td>
<td>26*</td>
</tr>
<tr>
<td>28</td>
<td>robot stop</td>
<td>314.21, 322.35</td>
<td>8</td>
</tr>
<tr>
<td>30</td>
<td>stop</td>
<td></td>
<td>26*</td>
</tr>
<tr>
<td>31</td>
<td>robot take a photo</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>User</td>
<td>DM &gt;</td>
<td>DM &gt; RN</td>
</tr>
<tr>
<td>---</td>
<td>---------------------------------------------------------------------</td>
<td>------</td>
<td>---------</td>
</tr>
<tr>
<td>68</td>
<td>keep moving until you see your next shovel</td>
<td>User</td>
<td></td>
</tr>
<tr>
<td>69</td>
<td>I think you are more familiar with shovels than I am.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>move forward &lt;pause .41&gt; ten feet</td>
<td></td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>processing...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>I will move forward as far as I can, ok?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>73</td>
<td>stop</td>
<td></td>
<td></td>
</tr>
<tr>
<td>74</td>
<td>done</td>
<td></td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>do you see any shovels near you?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>76</td>
<td>done</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4: The failed execution of the command underway (line #76) when stop is issued, as well as the speaker’s multiple attempts at producing an instruction that reflects their intent, provide evidence that “stop” (line #74) is an edit marker, providing the opportunity for a *fresh start* repair in line 77.
<table>
<thead>
<tr>
<th>#</th>
<th>User</th>
<th>Floor 1</th>
<th>Floor 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>move</td>
<td>DM &gt; User</td>
<td>DM &gt; RN</td>
</tr>
<tr>
<td>69</td>
<td>forward</td>
<td>move forward ten feet</td>
<td>move 10 feet</td>
</tr>
<tr>
<td>70</td>
<td></td>
<td>executing...</td>
<td>done</td>
</tr>
<tr>
<td>71</td>
<td>stop</td>
<td>done</td>
<td></td>
</tr>
<tr>
<td>72</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>73</td>
<td>take a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>74</td>
<td>picture</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Timing</td>
<td>Repair Status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------</td>
<td>--------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>No evidence of repair - halt motion</td>
<td>Evidence of repair - change strategy</td>
<td>Evidence of repair - fresh start strategy</td>
</tr>
<tr>
<td>Original instruction</td>
<td>38</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>During grounding</td>
<td>2</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>After grounding, before execution</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>During execution</td>
<td>59</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>After execution, before grounding termination</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>After grounding termination</td>
<td>12</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 2: Corpus counts of “stop”, as characterized along the dimensions of the timing of issuance and the status as a potential edit marker, signaling repair to come, either change or fresh start repair strategies.
Conclusions:

• Many aspects to discourse/dialogue structure
  – Types of phenomena
  – Types of structural relationships

• Scaling of Interactional Richness:
  -> Multifloor Dialogue

• Some aspects of structure only revealed in richer situations

• Most phenomena apply in richer situations, but some generalizations may not (e.g. global intentional structure)

• Still more work to do in all areas, but especially multi-party/multi-floor