

ESLLI2015 Advanced Course on Computational Models of Grounding in Dialogue

David Traum traum@ict.usc.edu

Lecture 5: Friday August 14th, 2015

USC



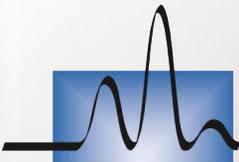
Outline of Course



- Preliminaries: representation, agency, communication
- Common Ground: How it is modeled and achieved
- Clark & Schaefer's Model of Grounding
- Computational Models of Grounding I: Brennan & Cahn
- Speech Acts and Dialogue Acts
- Multi-functionality of Utterances
- Feedback and Error-handling in Spoken Dialogue Systems
- Computational Models of Grounding II: Traum '94
- Miscommunication: The Good, the Bad, and the Ugly
- Decision-theoretic models of grounding
- Multi-modal Grounding
- **Multiparty Grounding**
- **Incremental Grounding**
- **Degrees of Grounding**
- **Applications of Grounding Analysis**

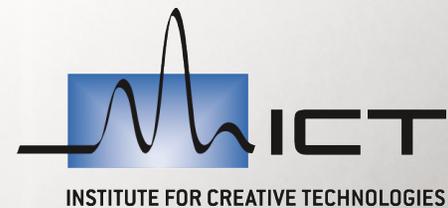
REVIEW OF YESTERDAY

USC

 M-ICT
INSTITUTE FOR CREATIVE TECHNOLOGIES

MULTI-MODAL GROUNDING

USC



Factors Affecting Grounding Behavior

▪ **Amount of grounding, type of act, content & realization of act, and model for groundedness depends on a number of factors including**

- Purposes & prior groundedness (Grounding Criterion)
- Available communication channels and resources
 - Costs and affordances: Clark and Brennan '90
 - Traum & Heeman '96: only 3-5% of utterances in spoken trains corpus had no grounding
 - Dillenbourg & Traum '96, 05: over 50% of utterances in typed MOO mystery solving dialogues had no grounding
- Content
 - Dillenbourg & Traum '96, 05
 - Sometimes shared situation model is better than explicit grounding model (for facts on shared whiteboard)

Multimodal Grounding: Key questions

- **What evidence signals can be performed in modality**
- **What affordances (constraints) does modality place on achieving/assuming common ground?**
- **Multifunctionality**
- **Within and cross-grounding**

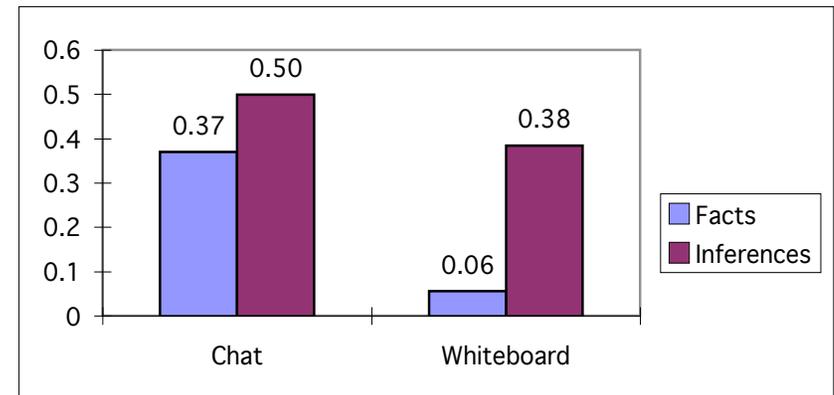
Dillenbourg & Traum 96, 05

Multi-modal computer-mediated grounding

- Grounding by category

Content of interactions	Acknowledgment Rate
Task knowledge	38%
Facts	26%
Inferences	46%
Task management	43%
Meta-Communication	55%
Technical problems	30%
<i>All categories</i>	<i>41%</i>

- Grounding by Category & Medium



Towards a Model of Face-to-Face Grounding

Yukiko Nakano (RISTEX)

Gabe Reinstein & Tom Stocky (Media Lab)

Justine Cassell (MIT Media Lab & Northwestern University)

Nonverbal Grounding: Nakano et. al. 2003

	Shift to	
	within UU	pause
Acknowledgement	$gMwN/gM$ (0.495)	gM/gM (0.888)
Answer	gP/gP (0.436)	gM/gM (0.667)
Info-req	gP/gM (0.38)	gP/gP (0.5)
Assertion	gP/gM (0.317)	gM/gM (0.418)

Speaker/Listener
 gP : gaze at partner
 gM : gaze at map
 $gMwN$: gaze at map & nod

UU: utterance unit (intonational)

Grounding Model for MACK

Target UU Type	Evidence Type	NV Pattern	Judgment of ground	Suggested next action
Assertion	positive	within: map pause: map /nod	grounded	go-ahead: 0.7 elaboration: 0.30
	negative	within: gaze pause: gaze	ungrounded	go-ahead: 0.27 elaboration: 0.73
Answer	positive	within: gaze pause: map	grounded	go-ahead: 0.83 elaboration: 0.17
	negative	pause: gaze	ungrounded	go-ahead: 0.22 elaboration: 0.78

MULTI-PARTY GROUNDING

USC



Multiparty Cases

- **Dyadic Exchanges within a larger group**
- **Multiple Addressees**
- **Multiple Conversations/floors**
 - Interactions

Participant Roles (Goffman 74, 81, Clark 96)

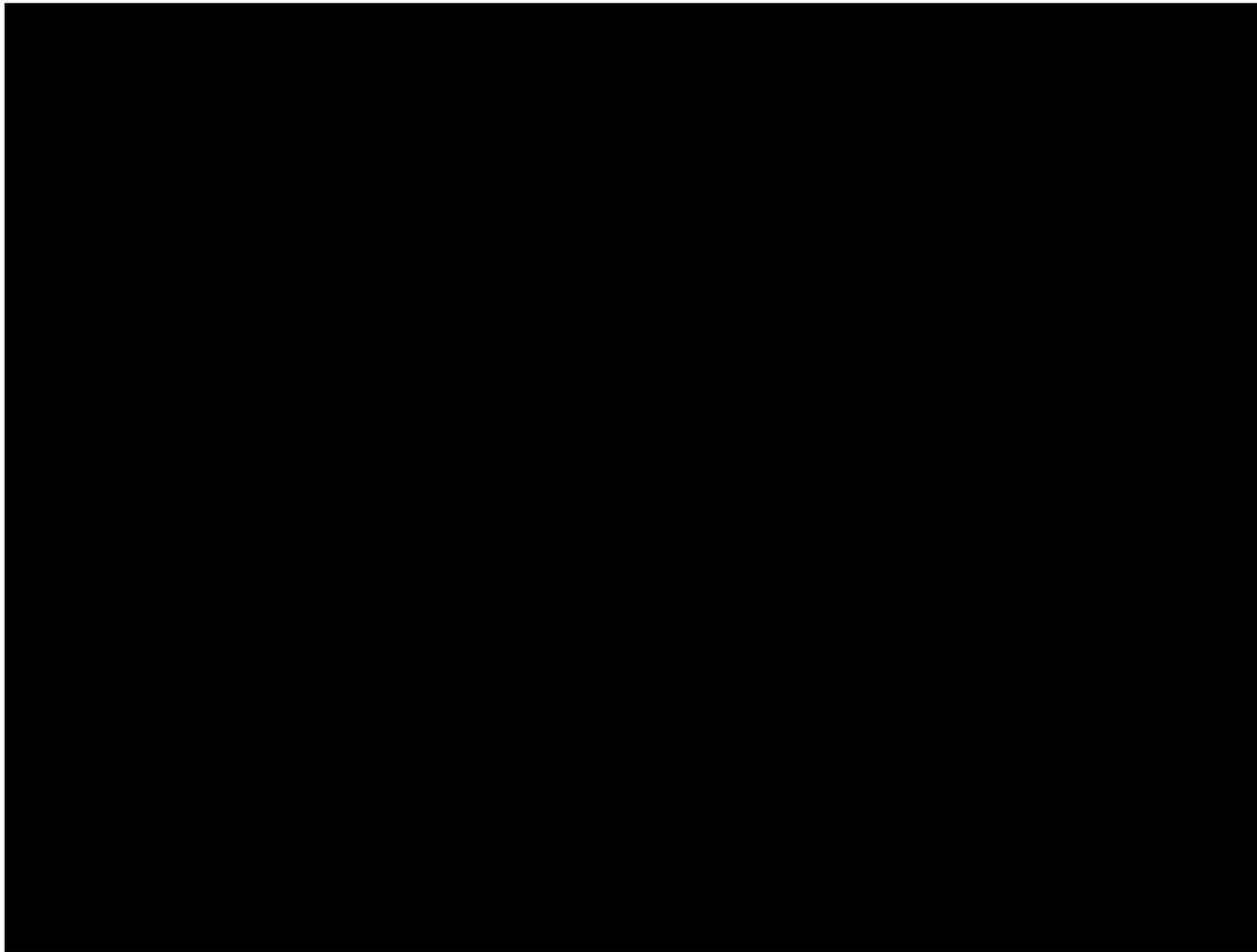
- ***Speaker* & *Hearer* are really complex composites**
 - Not individual roles
 - Different kinds of participant status
 - Different rights and responsibilities & actions

Novick, Walton & Ward '96: Contribution Graphs in Multiparty Discourse

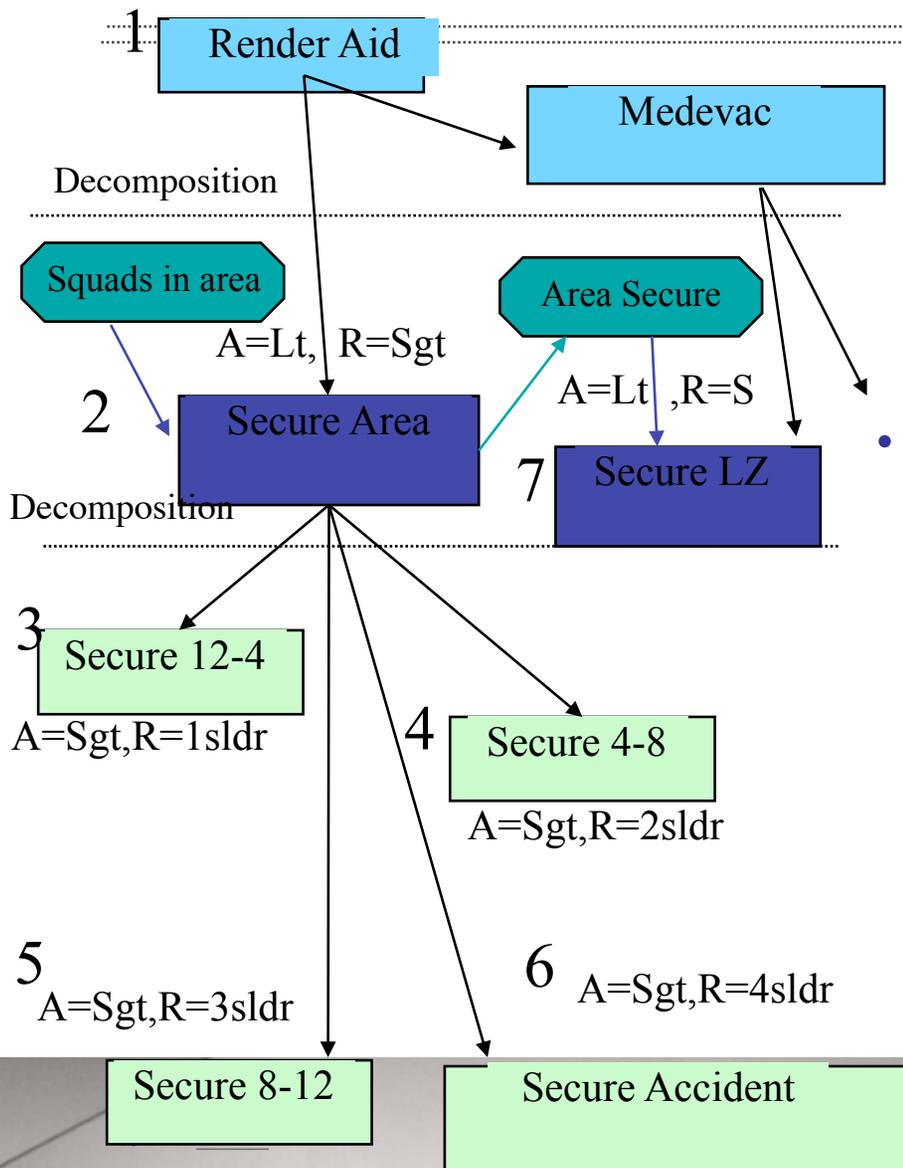
- **Assumptions:**

1. speaker need not ensure that non-addressees understand the presentation
2. a hearer may believe that she is an addressee even if she is not addressed directly by the speaker
3. hearer, even when she believes that she is an addressee, may present less-than-normally strong evidence of understanding if (a) other addressees present normally strong evidence and (b) the hearer believes the other addressees' understanding is sufficiently mutual.

MRE Team-Negotiation Example



Sgt's Negotiation Behavior



Focus=1

Lt: U9 "secure a landing zone"

Committed(Lt,7,sgt), 7 authorized, Obl(sgt,U9)

Sgt: U10 "sir first we should secure the assembly area"

Disparaged(sgt, 7,Lt), endorsed(sgt,2.Lt), grounded(U9)

Lt: U11 "secure the assembly area"

Committed(Lt,2,sgt), 2 authorized, Obl(sgt,U11),grounded(U10)

Sgt: U12 "understood sir"

Committed(sgt,2,Lt), grounded(U11), Push(2,focus)

Goal7: Announce(2,{1sldr,2sldr,3sldr,4sldr})

Goal8: Start-conversation(sgt, ,{1sldr,2sldr,...},2)

Goal8 -> Sgt: U21 "Squad leaders listen up!"

Goal7 -> Sgt: U22 "give me 360 degree security here"

Committed(sgt,2,{1sldr,2sldr,3sldr,4sldr})

Push(3, focus)

Goal9: authorize 3

Goal9 -> Sgt:U23 "1st squad take 12-4"

Committed(sgt,3, {1sldr,2sldr,3sldr,4sldr}), 3 authorized

Pop(3), Push(4)

Goal10: authorize 4

Goal10 -> Sgt: U24 "2nd squad take 4-8"

Committed(sgt,4,{1sldr,2sldr,3sldr,4sldr}), 4 authorized

Pop(4) ...

A10: Squads move

Grounded(U21-U26)

ends conversation about 2, Happened(2)

Push(7,Focus)

UTEP-ICT Cross-cultural multiparty multimodal dialog corpus (Herrera et al. 2010)

- Same three cultures (Arab, American, Mexican)
- Multiple groups (4 per culture)
- Multiple activities: collaborative/negotiative, narrative
 - Ex. Collaborative task: naming the toy



Grounding and Turn-Taking in Multimodal Multiparty Conversation



Interactive
Systems
Group

David Novick

Iván Gris

How do the mechanisms of grounding and turn-taking function across cultures in multiparty conversations?



Novick & Gris 2013

Research questions

1. Do grounding behaviors in multiparty conversations get cued in ways similar to those observed in dyadic conversation?
2. How do mechanisms of turn-transitions in multiparty conversations function?
3. Does presence of artifact in multiparty setting lead to changes in grounding behaviors?
4. How, if at all, do these behaviors differ between speakers of American English and of Mexican Spanish?

Methodology

UTEP-ICT Cross-Cultural Multiparty Multimodal Dialog Corpus

- Four-person groups performing narrative tasks, constructive tasks, artifact tasks & intercultural interaction tasks
- 20-second excerpts of conversations

Conversation analysis

- Transcribed speech
- Annotated gaze, nods & upper-body gestures
- Attempted to produce plausible explanation of how these actions served conversants in grounding (or not) each other's contributions to conversation and in taking conversational turns

American non-artifact



- Most listeners looking at speaker
- Lack of gaze shifts to listener to check for grounding or to offer turn
- Lack of nodding, and nods occur without eye contact
- Reinforcement gestures, mimicry
- Without artifact, conversants appeared to prefer overlapping for turn taking

American non-artifact

Start	End	A Verbal	A Nonverbal	B Verbal	B Nonverbal	C Verbal	C Nonverbal	D Verbal	D Nonverbal
04:15	05:10					"but they're driving like stupid"			
04:25	09:60		looks at C						
05:50	07:50				looks at C	"they're driving very slow and like"	looks at A		
06:00	10:80								looks at C
08:00	10:75					"they won't change lanes right"	looks away, mimics changing lane		
08:15	08:40				glances at A, looks back at C				

American artifact



- Conversants tend to gaze more at artifact than at each other
- Gaze shifts aim primarily at non-speakers
- Turn transition is not coordinated with gaze shift. Rather, gaze shift lags turn change
- Relative to non-artifact conversation, conversants use far fewer gestures

Mexican artifact



- Nods follow verbal consensus or agreement
- Nodding, even when no one is looking, can be the need to express agreement while not wanting to take the floor
- Repetition seems to act as an acknowledgment and an invitation for someone else to take the floor

Mexican artifact conversation

Start	End	A Verbal	A Nonverbal	B Verbal	B Nonverbal	C Verbal	C Nonverbal	D Verbal	D Nonverbal
19:00	19:50	"si blue punk"		"y luego"			nods		
19:50	20:00			"son los punk"					
20:00	20:50			"blue punk"			nods		
21:00	22:00	"si blue punk"							
22:00	23:00						nods	"ey, blue punk"	
23:00	24:00	"blue punk"							

Mexican non-artifact



- Conversants relatively more apt to gesture
- Two types of gestures
 - Agreement, which tends to be done in conjunction with verbal statement
 - Enacting of verbal description
- Overlap to elaborate on current topic

Conclusion

Do grounding behaviors such as nodding get cued in ways similar to those in dyadic conversation?

Multiparty conversants nod less frequently & are not cued by speaker's gaze shift.

How do the mechanisms of turn-transition function?

In multiparty conversation, conversants do not rely as much on gaze as turn cue. They overlap possibly because conversant could not engage speaker's gaze.

How, if at all, do these behaviors differ across cultures?

Differences likely reflect natural variation in conversation rather than clear cultural differences.

Comparison across cultures revealed similarities.

Yamashita et al 2008: Difficulties in Establishing Common Ground in Multiparty Groups using Machine Translation

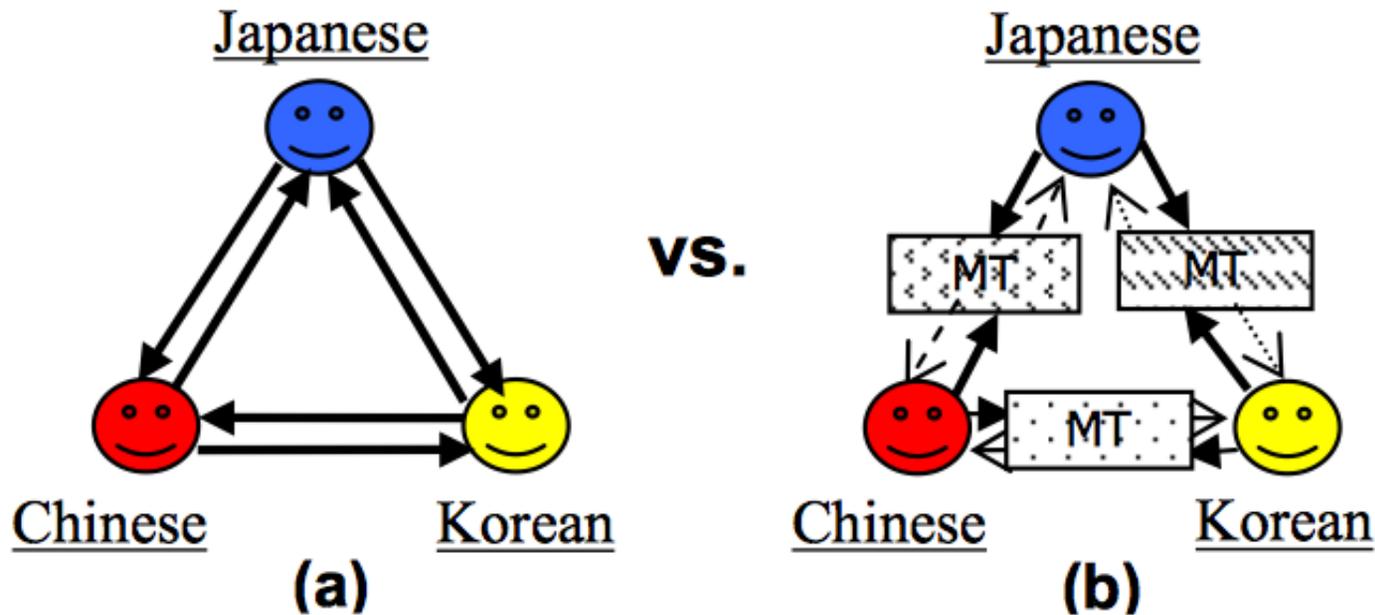


Figure 1 Three members communicating: (a) in their shared second language (English) or (b) in their native languages using machine translation software.

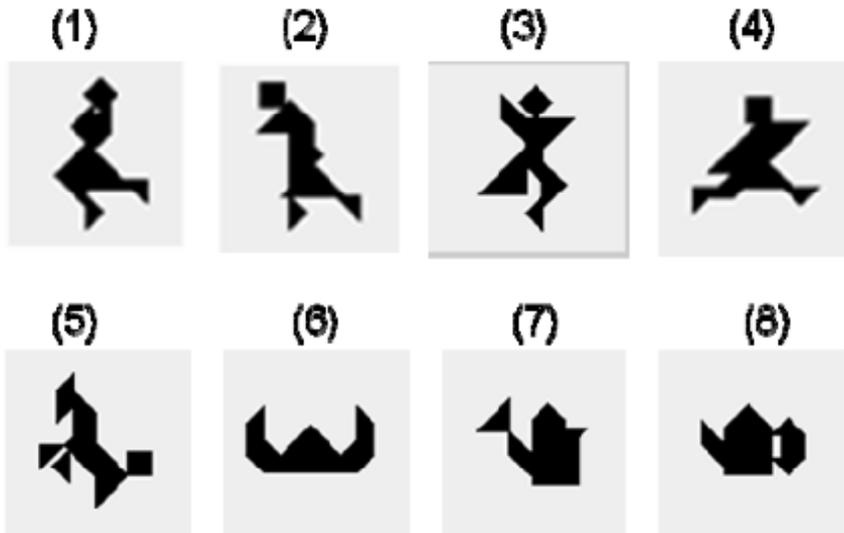
Conditions for efficient grounding

1. **they must share the same conversational content with others**
 2. **they must be aware that they are sharing the conversational content with others**
 3. **they must be able to distinguish between information they do and do not share with others**
- **All violated by MT-mediated conversation**

Hypotheses

- ***H1 (Efficiency of Mutual Acceptance Process):*** Participants will more efficiently identify a referent when using English rather than machine translation.
- ***H2 (Abbreviation of Referring Expressions over Trials):*** Participants will abbreviate their referring expressions more when using English than when using machine translation.
- ***H3 (Improvements in Making Appropriate References):*** Participants are less able to improve their efficiency of formulating appropriate references when using machine translation than when using English.

Example Interaction



e 3. Eight tangram figures used in the exper

Excerpt 3. Directors not being able to abbreviate their referring expressions (conversation is translated into English). Underline&Boldface indicates the originator of each message.

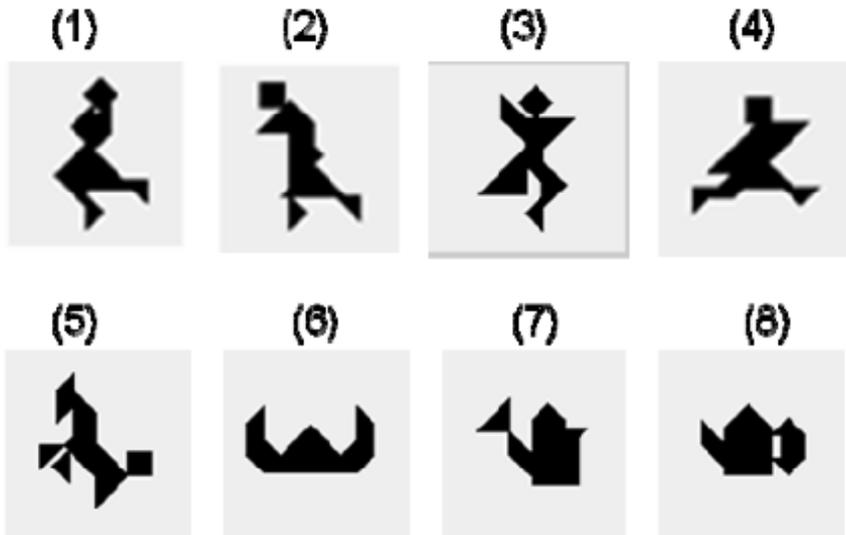
Japanese Screen	Korean Screen	Chinese Screen
<1st trial> Director: Japanese		
<u>J: Number 2 is a horse.</u>	J: Number 2 is a horse.	C: Number 2 is a horse.
<2nd trial> Director: Korean		
K: Number 4 is	<u>K: Number 4 is a person standing upside down.</u>	K: 4 times
--- (snip) ---		
<u>J: Mr. B. Which number is the animal?</u>	J: Mr. B. Which number is the animal?	J: Mr. B. Which number is the animal?
K: Animal?	<u>K: Animal?</u>	K: Animal?
--- (snip) ---		
<u>J: Which number is the creature with a square tail?</u>	J: Which number is the creature by which a tail is a square?	J: A tail, what number is a square creature?
C: An animal will be 8 days.	C: An animal is 8 days.	<u>C: Animal is number 8.</u>
K: I wouldn't know what to say, but something like an animal is 4 times most.	<u>K: I don't know what you are saying but the most animal like thing is number 4.</u>	K: Something like whatever animal says, is it wasteful, an unclear one is 4 times most.
<3rd trial> Director: Chinese		
C: It seems to be an animal.	C: It seems to be an animal.	<u>C: It looks like an animal.</u>
C: Horse	C: Horse	<u>C: Horse</u>
<4th trial> Director: Japanese		
<u>J: Horse. Animal.</u>	J: Horse. Animal.	J: Horse. Animal.
<u>J: Tail is square.</u>	J: A tail is square.	J: A tail is square.
<5th trial> Director: Korean		
K: It's an animal	<u>K: It's an animal.</u>	K: It's an animal.
K: It seems to be a word which raised its foreleg.	<u>K: It's a shape of a horse raising its front legs.</u>	K: A word is the design which entered a foreleg.
<6th trial> Director: Chinese		
C: Animal, it seems to be a horse.	C: Animal, it seems to be a horse.	<u>C: Animal, seems to be a horse.</u>
C: There is a square on the right side.	C: There is a square on the right side.	<u>C: There is a square on the right side.</u>

Excerpt 2. Director not being able to coordinate his utterance toward the slow Matcher (translated into English). Underline&Boldface indicates the originator of each message.

	Japanese Screen	Korean Screen	Chinese Screen
<2nd trial> Director: Korean			
1	K: Looks like a pitcher.	<u>K: The shape of a pitcher.</u>	K: It's a financial aid person electron, an arm is done.
2	C: Sorry, not well understood.	C: Sorry. Not well understood.	<u>C: Sorry, I don't understand.</u>
3	K: The third one is swept when watering flowers.	<u>K: The third one is used when watering flowers.</u>	K: When giving water to a flower, the third is used.
4	<u>J: A sprinkler?</u>	J: A sprinkler?	J: Is this a sprinkler?
5	K: Yes.	<u>K: Yes.</u>	K: Yes.
6	C: The mouth was big.	C: The mouth became big.	<u>C: Its spout is big.</u>
7	K: The mouth is big.	<u>K: The mouth is big.</u>	K: The mouth is big.
8	<u>J: Is the mouth triangle?</u>	J: Is the mouth triangle?	J: Is the mouth triangle?
9	C: Got it, no problem.	C: Got it. No problem.	<u>C: Got it. No problem.</u>
10	K: Do you understand?	<u>K: Do you understand?</u>	K: Do you understand?
11	K: OK.	<u>K: OK.</u>	K: OK.
12	K: The mouth is triangle.	<u>K: The mouth is triangle.</u>	K: Mouth is triangle.
13	<u>J: I got it!</u>	J: I got it!	J: I got it!
<3rd trial> Director: Chinese			
14	C: A sprinkler.	C: A sprinkler.	<u>C: A sprinkler.</u>
15	C: Water was given and it was consumed.	C: Water was given and it was consumed.	<u>C: We use it for watering flowers.</u>
16	K: I got it.	<u>K: I got it.</u>	K: I got it.
17	C: The mouth is big.	C: The mouth is big.	<u>C: The spout is big.</u>
18	K: Yes, yes.	<u>K: Sure, sure.</u>	K: Nene.
19	K: It has a right triangle mouth, right?	<u>K: It has a right triangle mouth.</u>	K: You had a mouth of a right triangle, right?
20	<u>J: Sorry,</u>	J: Sorry.	J: Sorry.
21	<u>J: I got it.</u>	J: I got it.	J: I got it.

Excerpt 3. Directors not being able to abbreviate their referring expressions (conversation is translated into English). Underline&Boldface indicates the originator of each message.

Japanese Screen	Korean Screen	Chinese Screen
<1st trial> Director: Japanese		
J: Number 2 is a horse.	J: Number 2 is a horse.	C: Number 2 is a horse.
<2nd trial> Director: Korean		
K: Number 4 is	K: Number 4 is a person standing upside down.	K: 4 times
--- (snip) ---		
J: Mr. B. Which number is the animal?	J: Mr. B. Which number is the animal?	J: Mr. B. Which number is the animal?
K: Animal?	K: Animal?	K: Animal?
--- (snip) ---		
J: Which number is the creature with a square tail?	J: Which number is the creature by which a tail is a square?	J: A tail, what number is a square creature?
C: An animal will be 8 days.	C: An animal is 8 days.	C: Animal is number 8.
K: I wouldn't know what to say, but something like an animal is 4 times most.	K: I don't know what you are saying but the most animal like thing is number 4.	K: Something like whatever animal says, is it wasteful, an unclear one is 4 times most.
<3rd trial> Director: Chinese		
C: It seems to be an animal.	C: It seems to be an animal.	C: It looks like an animal.
C: Horse	C: Horse	C: Horse
<4th trial> Director: Japanese		
J: Horse. Animal.	J: Horse. Animal.	J: Horse. Animal.
J: Tail is square.	J: A tail is square.	J: A tail is square.
<5th trial> Director: Korean		
K: It's an animal	K: It's an animal.	K: It's an animal.
K: It seems to be a word which raised its foreleg.	K: It's a shape of a horse raising its front legs.	K: A word is the design which entered a foreleg.
<6th trial> Director: Chinese		
C: Animal, it seems to be a horse.	C: Animal, it seems to be a horse.	C: Animal, seems to be a horse.
C: There is a square on the right side.	C: There is a square on the right side.	C: There is a square on the right side.



e 3. Eight tangram figures used in the exper

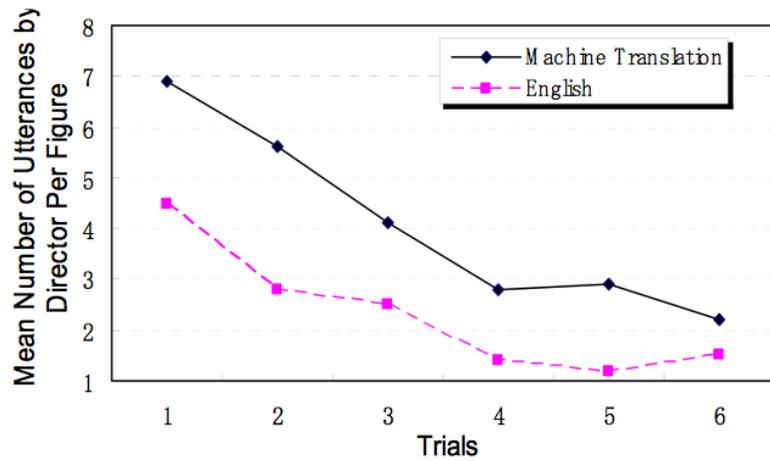


Figure 4. Mean number of utterances by a Director per figure.

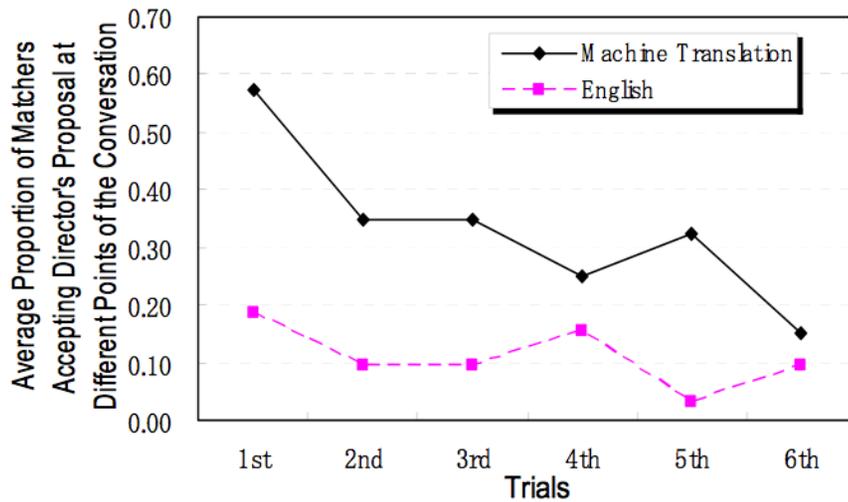


Figure 5. Average proportion of Matchers identifying a figure at different points in the conversation.

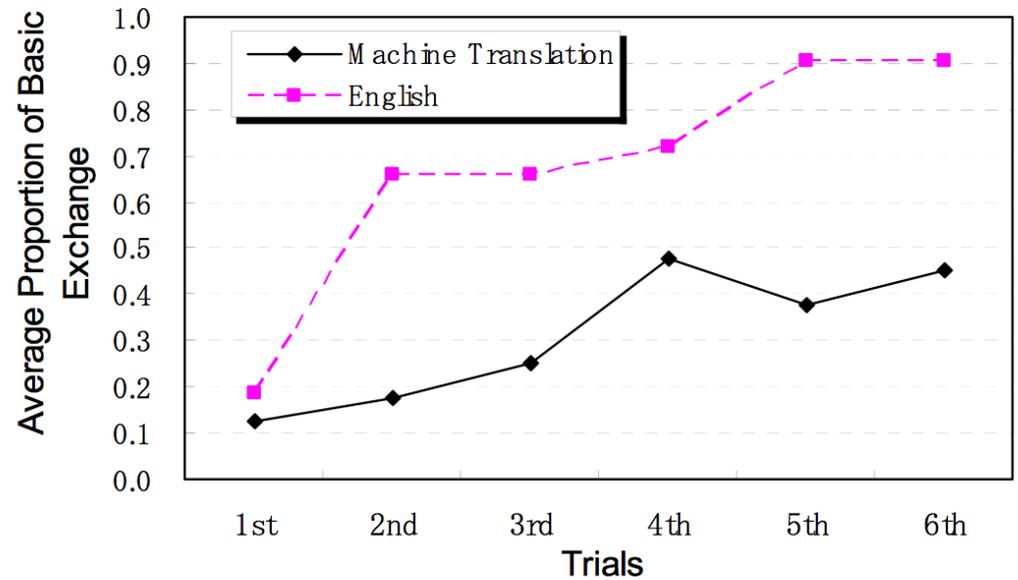


Figure 6. Average Proportion of Basic Exchange.

INCREMENTAL GROUNDING

USC



Incremental Grounding

Thomas Visser , David Traum, David DeVault, Riëks op den Akker A Model for Incremental Grounding in Spoken Dialogue Systems in *Journal of Multimodal User Interfaces*, March 2014, Volume 8, Issue 1, 61-73



- How to model what has and hasn't been added to common ground
 - Multiple, phased (and continuous) signals

Turn-taking, Grounding, and Feedback in Complex Human Dialogue



Example: AMI dialogue: collaborative completion

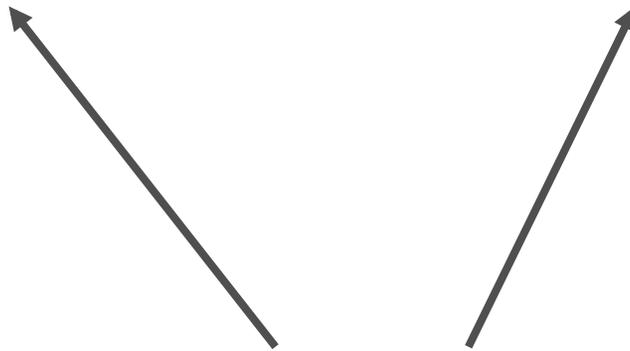


- C: We could just go with um
- D*: Yeah
- A: Normal coloured buttons
- B: Well do you want colour differentiation here?
- A: No that's not the button we're talking about
- B*: O yeah sorry yeah
- A: The buttons only refer to the pad so
- B: Right so
- A: Should we take that off
- B: aww
- A: Hey it's back to the original. Um so then these just become normal coloured buttons
- C: Hmhm
- A: So that might be some some way of cutting the cost
- C: Hmhm
- B: Okay



We could just go with um

normal coloured buttons



Explicit / Implicit

Example AMI dialogue: “incorrect” completion



- B: That would probably not be in keeping with the um the
- C*: *laugh* Technology
- B: fashion statement and such, yeah.
- C*: Yeah.

AMI: completion, accepted



- B: However I've got a couple of worries about that
The power required , um and the ability to
- D: the cost
- B: cost It seems like for an embedded system ...

Incremental Processing



- Goal: Can we make Virtual Human dialogue more human-like by enabling simultaneous listening and reactive/predictive behavior?
- Approach:
 - Incremental ASR & NLU
 - Analyze speech signal fully as its being created
 - Predict ultimate utterance meaning and timing
 - Create reaction strategies
 - backchannel feedback
 - Collaborative completions
 - Early responses
 - Co-construction of utterances/meaning

Incremental NLU: Increasing Interactivity



- Issues:
 - Can we decrease reaction time?
 - Can we predict what speaker means before they finish?
 - Can we use this knowledge in dialogue behavior?
- Approach
 - Analyze accuracy and confidence of prediction of NLU on partial input
 - Use NLU prediction for cognitive reaction decisions
- Results :
 - Prediction model
 - Sagae et al NAACL 2009
 - Maxf confidence (DeVault et al Sigdial 2009)
 - Completion demo
 - Sagae et al NAACL 2010 demo session
 - Additional metrics
 - DeVault et al, Interspeech 2011
 - Predicting Explicit sub-frame
 - DeVault & Traum Interspeech 2013

Standard Pipeline: no NLU until speech finished





Incremental Natural Language Understanding

- User utterance:

i have orders to move this clinic to a camp near the U.S. base

- NLU incrementally interprets ASR output:

i have

i have orders

i have orders to move this clinic

i have orders to move this clinic to a camp near the

- Research questions:
 - How well do our virtual humans understand at each point?
 - How confident can they be in their interpretations?
 - Can they sometimes guess how the user's utterance will end?

Incremental NLU: Updated NLU Predictions while listening





Comparing different kinds of incremental NLU

full user utterance: *we are prepared to give you guys generators for electricity*

partial user utterance: *we apparently give you*

explicit incremental NLU

predictive incremental NLU

full utterance NLU

- <s>.mood declarative
- <s>.sem.agent captain-kirk
- <s>.sem.modal.intention will
- <s>.sem.event deliver
- <s>.sem.speechact.type promise

(Sagae et al. 2009, DeVault et al. 2009, DeVault et al. 2011a,b)

- <s>.mood declarative
- <s>.sem.agent captain-kirk
- <s>.sem.modal.intention will
- <s>.sem.event deliver
- <s>.sem.speechact.type promise
- <s>.sem.theme medical-supplies

partial utterance meaning:

full utterance meaning:

Learning a Prediction model: MaxF (DeVault, Sagae, Traum, SIGDIAL 2009)



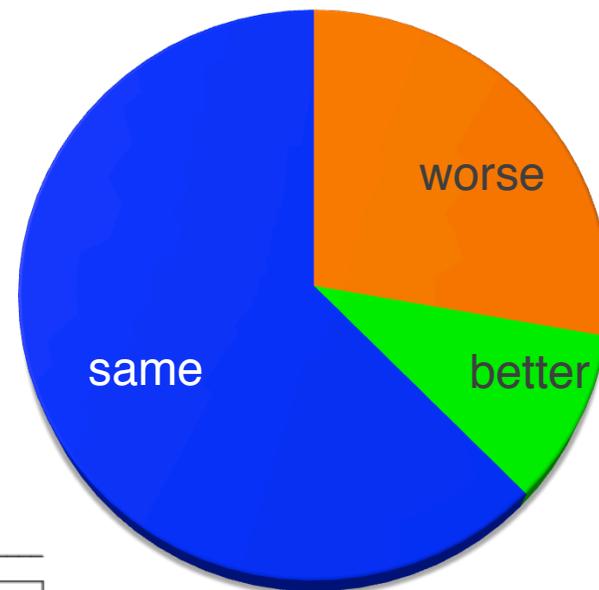
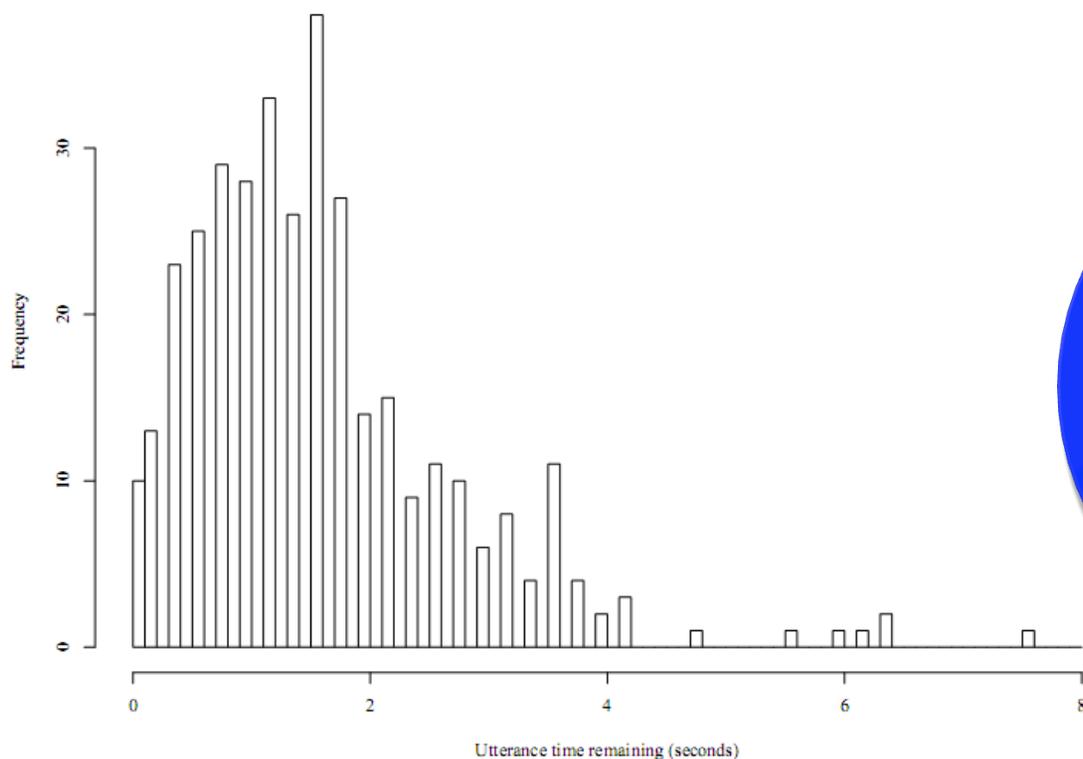
- Train decision tree
 - available features:
 - Temporal features: elapsed time (200 ms increments)
 - ASR features: partial length
 - NLU features: entropy, maximum probability, output frame
 - Training data:
 - 449 utterances with 6068 partials
- Results
 - Trained model predicts early maximum F-score in 79.2% of utterances
 - decision accuracy
 - Precision = 0.88
 - Recall = 0.52
 - F-score = 0.65



Impact of Prediction model:

- Faster Understanding (Mean time saved: 1.6sec)

Change in Performance (mean -0.14 F-score)



Automated Completions (Sagae et al NAACL 2010)





Meta NLU (DeVault et al, Interspeech2011):

- Confidence indicators for each partial utterance
 - F_t = estimated f-score of current hypothesis
 - F_L = estimated f-score of final hypothesis
- Derived metrics

Metric	Definition	Metric	Definition	Metric	Definition
$High_t$:	$F_t \geq \frac{1}{2}$	$WillBeHigh_t$:	$F_L \geq \frac{1}{2}$	$PF1_t$:	$Correct_t \vee (Incorrect_t \wedge WillBeCorrect_t)$
$Correct_t$:	$F_t = 1$	$WillBeCorrect_t$:	$F_L = 1$	$PF2_t$:	$High_t \vee (Low_t \wedge WillBeHigh_t)$
$Incorrect_t$:	$F_t < 1$	$WillBeIncorrect_t$:	$F_L < 1$	$PF3_t$:	$High_t \vee (Low_t \wedge \neg MAXF_t)$
Low_t :	$F_t < \frac{1}{2}$	$WillBeLow_t$:	$F_L < \frac{1}{2}$		
		$MAXF_t$:	$F_t \geq F_L$		



Explicit NLU: use probability information to identify explicit meaning (DeVault & Traum, NAACL-HLT 2013)



Speech: we are prepared to give you guys generators for electricity

Partial ASR result: we apparently give you

explicit NLU probability **frame element**

↑	.856	<s>.mood declarative
	.824	<s>.sem.agent captain-kirk
	.663	<s>.sem.event deliver
↓	.457	<s>.sem.modal.intention will
	.412	<s>.sem.speechact.type promise
	.321	<s>.sem.theme power-generators

The “explicit subframe” produced by our method

We evaluate how well the explicit subframe tracks the user’s explicit incremental meaning during the utterance.



Explicit Incremental Dialogue Visualization

- Explicit & Predicted
- Current and predicted understanding levels
- Probability distributions

WE CAN GIVE YOU

we could give you three hundred fifty dollars

s.mood declarative
s.sem.agent we
s.sem.destination you
s.sem.event give
s.sem.modal.possibility can
s.sem.speechact.type offer
s.sem.theme threehundredfifty
s.sem.type event

Subframe:

s.mood declarative
s.sem.agent we
s.sem.destination you
s.sem.event give
s.sem.modal.possibility can
s.sem.speechact.type offer
s.sem.type event

Will Understand

Now Understanding

|can you say that again
|can you say that again harmony
|can you say that again utah
|do you agree to be the sheriff
|do you agree to be the sheriff utah
|do you approve of utah being the sheriff
do you approve of utah being the sheriff harmony
|do you disapprove of utah being the sheriff
|do you disapprove of utah being the sheriff harmony
goodbye
harmony
|hello
hello harmony
hello utah
i accept that
|i promise not to tell anyone your secret
i reject that
|it's nice to talk to you
|it's nice to talk to you harmony
|it's nice to talk to you utah and harmony
okay
sorry
thank you
thank you ma'am
thank you sir
thanks guys
utah
|we could arrest the outlaws
|we could give you guns
|we could give you one hundred fifty dollars
|we could give you three hundred fifty dollars
|we could give you two hundred dollars
|we do not want you to pay the outlaws two hundred dollars
|we promise to kill the outlaws
|we want the town to be safe

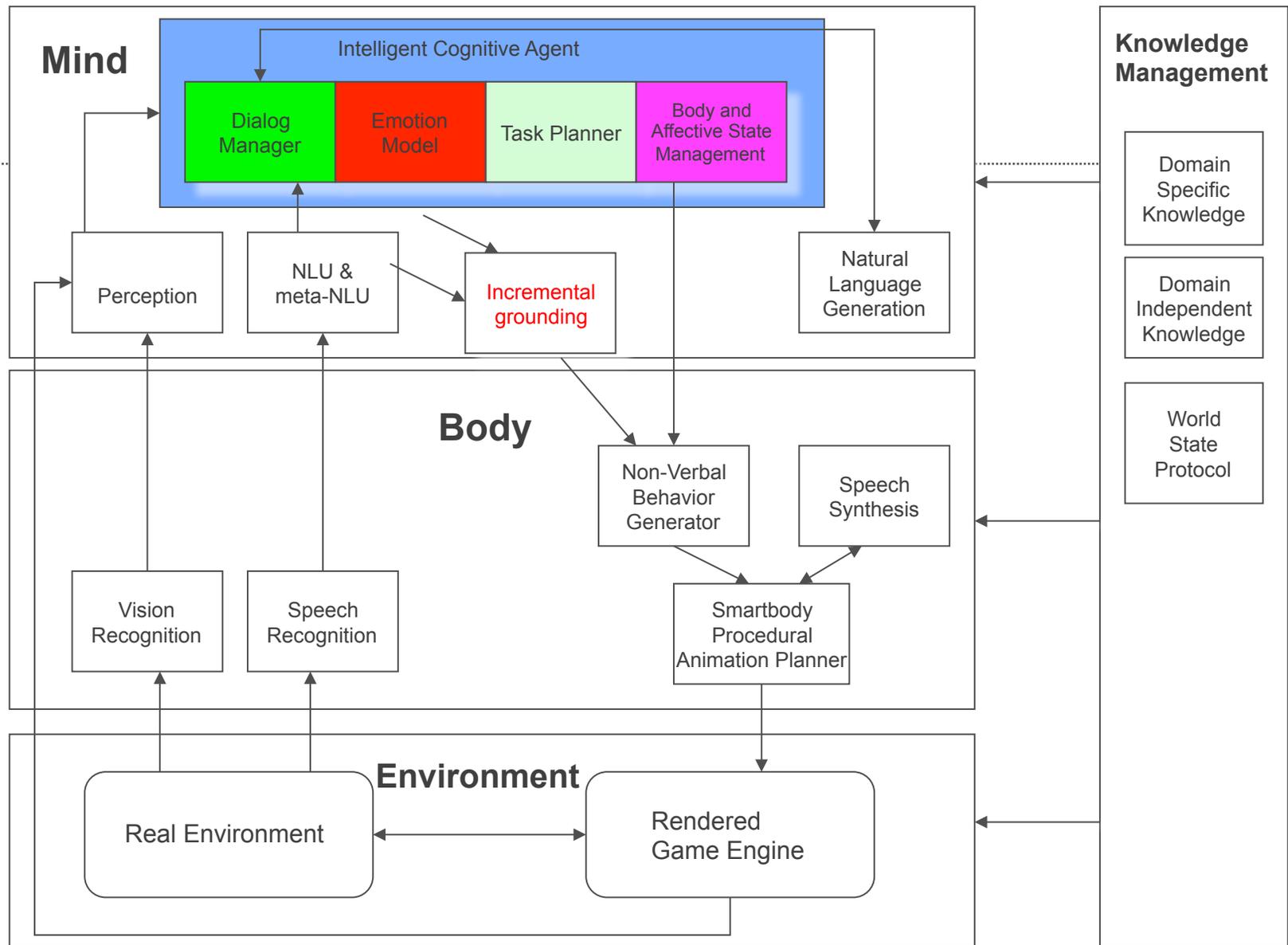
|< < 8 of 17 > >

Predicting Explicit Subframes

DeVault & Traum , Naacl 2013

- Predicted & Explicit sub-Frame
- Probability distribution across hypotheses
- Estimated F-score & confidence







Inputs to incremental grounding

- Stream of hypotheses
 - Explicit part
 - Predicted part
 - Confidence scores

Full frame

s.addressee	utah
s.mood	declarative
s.sem.type	event
s.sem.agent	you
s.sem.event	providePublicServices
s.sem.modal.desire	want
s.sem.modal.holder	we
s.sem.speechact.type	statement
s.sem.theme	sheriff-job

Explicit sub-frame

High, Low, Correct, Incorrect,
 WillBeCorrect, WillBeIncorrect,
 EF,
 Explicit sub-frame threshold

Confidence metrics



“Utah, we can give you two hundred dollars”

#	Incremental ASR results	Confidence
0	UTAH	medium
1	UTAH	high
2	UTAH WHAT	high
3	UTAH WHAT WE CAN	medium
4	UTAH WHAT WE CAN GET YOU	low
5	UTAH WHAT WE CAN GET YOU	low
6	UTAH WHAT WE CAN GIVE YOU TWO	high
7	UTAH WHAT WE CAN GIVE YOU TWO HUNDRED DOLLARS	correct



“Utah, we can give you two hundred dollars”

#	Incremental ASR results	Confidence	NLU Frame (Explicit Predicted)	
0	UTAH	medium	s.Addressee	utah
1	UTAH	high		
2	UTAH WHAT	high		
3	UTAH WHAT WE CAN	medium		
4	UTAH WHAT WE CAN GET YOU	low		
5	UTAH WHAT WE CAN GET YOU	low		
6	UTAH WHAT WE CAN GIVE YOU TWO	high		
7	UTAH WHAT WE CAN GIVE YOU TWO HUNDRED DOLLARS	correct		



“Utah, we can give you two hundred dollars”

#	Incremental ASR results	Confidence	NLU Frame (Explicit Predicted)	
0	UTAH	medium	s.Addressee	utah
1	UTAH	high		
2	UTAH WHAT	high		
3	UTAH WHAT WE CAN	medium		
4	UTAH WHAT WE CAN GET YOU	low		
5	UTAH WHAT WE CAN GET YOU	low		
6	UTAH WHAT WE CAN GIVE YOU TWO	high		
7	UTAH WHAT WE CAN GIVE YOU TWO HUNDRED DOLLARS	correct		



“Utah, we can give you two hundred dollars”

#	Incremental ASR results	Confidence	NLU Frame (Explicit Predicted)
0	UTAH	medium	
1	UTAH	high	s.Addressee utah s.sem.speechact.type no-ack
2	UTAH WHAT	high	
3	UTAH WHAT WE CAN	medium	
4	UTAH WHAT WE CAN GET YOU	low	
5	UTAH WHAT WE CAN GET YOU	low	
6	UTAH WHAT WE CAN GIVE YOU TWO	high	
7	UTAH WHAT WE CAN GIVE YOU TWO HUNDRED DOLLARS	correct	



“Utah, we can give you two hundred dollars”

#	Incremental ASR results	Confidence	NLU Frame (Explicit Predicted)	
0	UTAH	medium	s.Addressee	utah
1	UTAH	high	s.sem.speechact.type	no-ack
2	UTAH WHAT	high		
3	UTAH WHAT WE CAN	medium		
4	UTAH WHAT WE CAN GET YOU	low		
5	UTAH WHAT WE CAN GET YOU	low		
6	UTAH WHAT WE CAN GIVE YOU TWO	high		
7	UTAH WHAT WE CAN GIVE YOU TWO HUNDRED DOLLARS	correct		



“Utah, we can give you two hundred dollars”

#	Incremental ASR results	Confidence	NLU Frame (Explicit Predicted)	
0	UTAH	medium	s.Addressee	utah
1	UTAH	high	s.mood	declarative
2	UTAH WHAT	high	s.sem.type	event
3	UTAH WHAT WE CAN	medium	s.sem.speechact.type	statement
4	UTAH WHAT WE CAN GET YOU	low	s.sem.agent	you
5	UTAH WHAT WE CAN GET YOU	low	s.sem.event	providePublicServices
6	UTAH WHAT WE CAN GIVE YOU TWO	high	s.sem.modal.desire	want
7	UTAH WHAT WE CAN GIVE YOU TWO HUNDRED	correct	s.sem.modal.holder	we
			s.sem.theme	sheriff-job



“Utah, we can give you two hundred dollars”

#	Incremental ASR results	Confidence	NLU Frame (Explicit Predicted)	
0	UTAH	medium	s.Addressee	utah
1	UTAH	high	s.mood	declarative
2	UTAH WHAT	high	s.sem.type	event
3	UTAH WHAT WE CAN	medium	s.sem.speechact.type	statement
4	UTAH WHAT WE CAN GET YOU	low	s.sem.agent	you
5	UTAH WHAT WE CAN GIVE YOU TWO	high	s.sem.event	providePublicServices
6	UTAH WHAT WE CAN GIVE YOU TWO	high	s.sem.modal.desire	want
7	UTAH WHAT WE CAN GIVE YOU TWO HUNDRED	correct	s.sem.modal.holder	we
			s.sem.theme	sheriff-job



“Utah, we can give you two hundred dollars”

#	Incremental ASR results	Confidence	NLU Frame (Explicit Predicted)	
0	UTAH	medium		
1	UTAH	high	s.Addressee	utah
2	UTAH WHAT	high	s.mood	declarative
3	UTAH WHAT WE CAN	medium	s.sem.type	event
4	UTAH WHAT WE CAN GET YOU	low	s.sem.agent	we
5	UTAH WHAT WE CAN GET YOU	low	s.sem.destination	you
6	UTAH WHAT WE CAN GIVE YOU TWO	high	s.sem.modal.possibility	can
7	UTAH WHAT WE CAN GIVE YOU TWO HUNDRED DOLLARS	correct	s.sem.speechact.type	offer
			s.sem.event	give
			s.sem.theme	twohundred



“Utah, we can give you two hundred dollars”

#	Incremental ASR results	Confidence	NLU Frame (Explicit Predicted)	
0	UTAH	medium		
1	UTAH	high	s.Addressee	utah
2	UTAH WHAT	high	s.mood	declarative
			s.sem.type	event
3	UTAH WHAT WE CAN	medium	s.sem.agent	we
			s.sem.destination	you
4	UTAH WHAT WE CAN GET YOU	low	s.sem.modal.possibility	can
			s.sem.speechact.type	offer
5	UTAH WHAT WE CAN GET YOU	low	s.sem.event	give
			s.sem.theme	twohundred
6	UTAH WHAT WE CAN GIVE YOU TWO	high		
7	UTAH WHAT WE CAN GIVE YOU TWO HUNDRED DOLLARS	correct		



Incremental Grounding

- First non-empty explicit sub-frame **initiates** new CGU
- Consecutive compatible explicit sub-frames **continue** the CGU
- An incompatible explicit sub-frame requires **repair** of existing CGU's: old elements are removed
- If, after the repair, a CGU is empty, it is **cancelled**
- An acknowledging nod or verbal backchannel **acknowledges** the open CGU
- First explicit sub-frame with unacknowledged elements **initiates** new CGU

Recognition conditions for incremental grounding acts



- **Initiate:**
 - New utterance that is not another act on an existing unit
 - Utterance continuation where previous part is already grounded
- **Continue:**
 - Open cgu on same topic with speaker as initiator, current utterance is not a repair

incremental Acknowledgement



- Understanding Backchannel (e.g. head nod or “yeah”)
 - Grounds explicit part of utterance, any continuation by original speaker is now a new initiate
- Correct completion
 - Grounds explicit+predicted
- Incorrect completion
 - Grounds explicit part, initiates completion itself



incremental grounding acts

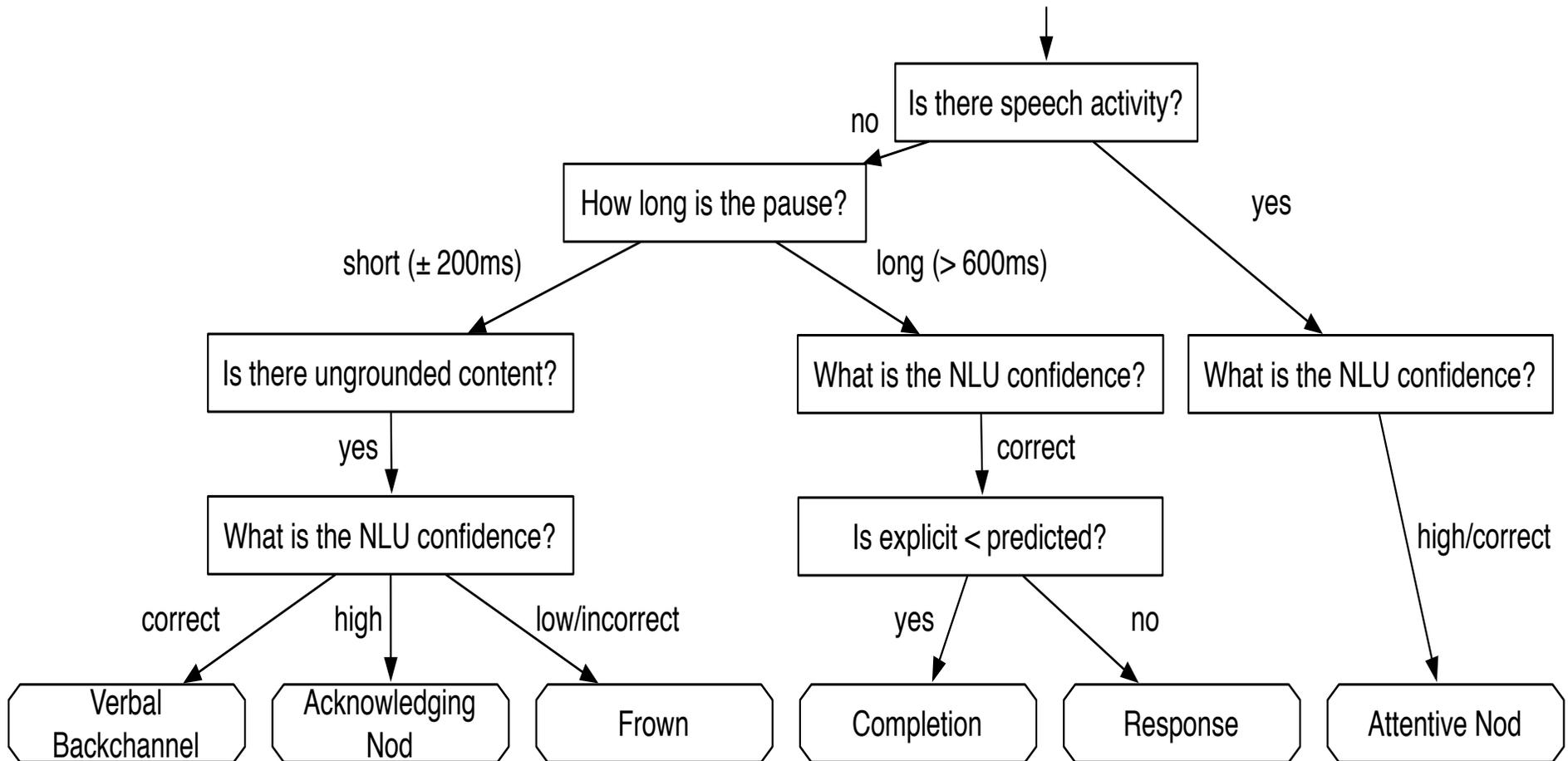
- Request for Repair
 - Signal of not understanding or specific problems
- Repair
 - Remove content from CGU, replace with explicit content from repair (cancel cgu if nothing left)



Grounding Behaviors:

- Acknowledging head nod: ack
- Attentive head nod: not acknowledging, but inviting continuation
- Frown: request for repair
- Verbal backchannel: ack of explicit
- Completion: attempted ack of explicit+predicted
- Response: acknowledge explicit + predicted (if any)

Behavior Policy





“Utah, we can give you two hundred dollars”

#	Incremental ASR results	Confidence
0	UTAH	medium
1	UTAH	high
	U Acknowledging nod	
2	UTAH WHAT	high
3	UTAH WHAT WE CAN	medium
4	UTAH WHAT WE CAN GET YOU	low
5	UTAH WHAT WE CAN GET YOU	low
	U frown	
6	UTAH WHAT WE CAN GIVE YOU TWO	high
	U Attentive nod	
7	UTAH ...TWO HUNDRED DOLLARS	correct

U Response

“Utah, we can give you two hundred dollars”



#	Incremental ASR results	Confidence	Grounding act
0	UTAH	medium	initiate ₁
1	UTAH	high	initiate ₁
U	Acknowledging nod		ack ₁
2	UTAH WHAT	high	initiate ₂
3	UTAH WHAT WE CAN	medium	initiate ₂
4	UTAH WHAT WE CAN GET YOU	low	initiate ₂
5	UTAH WHAT WE CAN GET YOU	low	initiate ₂
U	frown		reqrepair ₂
6	UTAH WHAT WE CAN GIVE YOU TWO	high	repair ₂
U	Attentive nod		
7	UTAH ...TWO HUNDRED DOLLARS	correct	Repair ₂
U	Response		Ack ₂ , init ₃

“Utah, we can give you two hundred dollars”



Incremental ASR results Confidence Grounding act

#	Incremental ASR results	Confidence	Grounding act	CGU	State	Contents
0	UTAH	medium	initiate ₁			
1	UTAH	high	initiate ₁	1	1	s.Addressee utah
	U Acknowledging nod		ack ₁			
2	UTAH WHAT	high	initiate ₂			
3	UTAH WHAT WE CAN	medium	initiate ₂			
4	UTAH WHAT WE CAN GET YOU	low	initiate ₂			
5	UTAH WHAT WE CAN GET YOU	low	initiate ₂			
	U frown		reqrepair ₂			
6	UTAH WHAT WE CAN GIVE YOU TWO	high	repair ₂			
	U Attentive nod					
7	UTAH ...TWO HUNDRED DOLLARS	correct	Repair ₂			
	U Response		Ack ₂ , init ₃			

“Utah, we can give you two hundred dollars”



#	Incremental ASR results	Confidence	Grounding act
---	-------------------------	------------	---------------

0	UTAH	medium	initiate ₁
1	UTAH	high	initiate ₁
	U Acknowledging nod		ack ₁
2	UTAH WHAT	high	initiate ₂
3	UTAH WHAT WE CAN	medium	initiate ₂
4	UTAH WHAT WE CAN GET YOU	low	initiate ₂
5	UTAH WHAT WE CAN GET YOU	low	initiate ₂
	U frown		reqrepair ₂
6	UTAH WHAT WE CAN GIVE YOU TWO	high	repair ₂
	U Attentive nod		
7	UTAH ...TWO HUNDRED DOLLARS	correct	Repair ₂
	U Response		Ack ₂ , init ₃

CGU	State	Contents
1	1	s.Addressee utah

“Utah, we can give you two hundred dollars”



#	Incremental ASR results	Confidence	Grounding act			
				CGU	State	Contents
0	UTAH	medium	initiate ₁			
1	UTAH	high	initiate ₁	1	F	s.Addressee utah
	U Acknowledging nod		ack ₁			
2	UTAH WHAT	high	initiate ₂			
3	UTAH WHAT WE CAN	medium	initiate ₂			
4	UTAH WHAT WE CAN GET YOU	low	initiate ₂			
5	UTAH WHAT WE CAN GET YOU	low	initiate ₂			
	U frown		reqrepair ₂			
6	UTAH WHAT WE CAN GIVE YOU TWO	high	repair ₂			
	U Attentive nod					
7	UTAH ...TWO HUNDRED DOLLARS	correct	Repair ₂			
	U Response		Ack ₂ , init ₃			

“Utah, we can give you two hundred dollars”



#	Incremental ASR results	Confidence	Grounding act			
				CGU	State	Contents
0	UTAH	medium	initiate ₁			
1	UTAH	high	initiate ₁	1	F	s.Addressee utah
	U Acknowledging nod		ack ₁	2	1	s.sem.speechact,type no-ack
2	UTAH WHAT	high	initiate ₂			
3	UTAH WHAT WE CAN	medium	initiate ₂			
4	UTAH WHAT WE CAN GET YOU	low	initiate ₂			
5	UTAH WHAT WE CAN GET YOU	low	initiate ₂			
	U frown		reqrepair ₂			
6	UTAH WHAT WE CAN GIVE YOU TWO	high	repair ₂			
	U Attentive nod					
7	UTAH ...TWO HUNDRED DOLLARS	correct	Repair ₂			

“Utah, we can give you two hundred dollars”



#	Incremental ASR results	Confidence	Grounding act	CGU State Contents		
				CGU	State	Contents
0	UTAH	medium	initiate ₁			
1	UTAH	high	initiate ₁	1	F	s.Addressee utah
	U Acknowledging nod		ack ₁	2	1	s.sem.speechact,type no-ack
2	UTAH WHAT	high	initiate ₂			
3	UTAH WHAT WE CAN	medium	initiate ₂			
4	UTAH WHAT WE CAN GET YOU	low	initiate ₂			
5	UTAH WHAT WE CAN GET YOU	low	initiate ₂			
	U frown		reqrepair ₂			
6	UTAH WHAT WE CAN GIVE YOU TWO	high	repair ₂			
	U Attentive nod					
7	UTAH ...TWO HUNDRED DOLLARS	correct	Repair ₂			

“Utah, we can give you two hundred dollars”



#	Incremental ASR results	Confidence	Grounding act	CGU	State	Contents
0	UTAH	medium	initiate ₁			
1	UTAH	high	initiate ₁	1	F	s.Addressee utah
	U Acknowledging nod		ack ₁			
2	UTAH WHAT	high	initiate ₂	2	1	s.mood declarative s.sem.type event s.sem.speechact.type statement
3	UTAH WHAT WE CAN	medium	initiate ₂			
4	UTAH WHAT WE CAN GET YOU	low	initiate ₂			
5	UTAH WHAT WE CAN GET YOU	low	initiate ₂			
	U frown		reqrepair ₂			
6	UTAH WHAT WE CAN GIVE YOU TWO	high	repair ₂			
	U Attentive nod					
7	UTAH ...TWO HUNDRED DOLLARS	correct	Repair ₂			
	U Response		Ack ₂ , init ₃			

“Utah, we can give you two hundred dollars”



#	Incremental ASR results	Confidence	Grounding act	CGU	State	Contents
0	UTAH	medium	initiate ₁			
1	UTAH	high	initiate ₁	1	F	s.Addressee utah
	U Acknowledging nod		ack ₁			
2	UTAH WHAT	high	initiate ₂	2	1	s.mood declarative s.sem.type event s.sem.speechact.type statement
3	UTAH WHAT WE CAN	medium	initiate ₂			
4	UTAH WHAT WE CAN GET YOU	low	initiate ₂			
5	UTAH WHAT WE CAN GET YOU	low	initiate ₂			
	U frown		reqrepair ₂			
6	UTAH WHAT WE CAN GIVE YOU TWO	high	repair ₂			
	U Attentive nod					
7	UTAH ...TWO HUNDRED DOLLARS	correct	Repair ₂			
	U Response		Ack ₂ , init ₃			

“Utah, we can give you two hundred dollars”



#	Incremental ASR results	Confidence	Grounding act	CGU	State	Contents
0	UTAH	medium	initiate ₁			
1	UTAH	high	initiate ₁	1	F	s.Addressee utah
	U Acknowledging nod		ack ₁			
2	UTAH WHAT	high	initiate ₂	2	2	s.mood declarative s.sem.type event s.sem.speechact.type statement
3	UTAH WHAT WE CAN	medium	initiate ₂			
4	UTAH WHAT WE CAN GET YOU	low	initiate ₂			
5	UTAH WHAT WE CAN GET YOU	low	initiate ₂			
	U frown		reqrepair ₂			
6	UTAH WHAT WE CAN GIVE YOU TWO	high	repair ₂			
	U Attentive nod					
7	UTAH ...TWO HUNDRED DOLLARS	correct	Repair ₂			
	U Response		Ack ₂ , init ₃			

“Utah, we can give you two hundred dollars”



#	Incremental ASR results	Confidence	Grounding act
---	-------------------------	------------	---------------

0	UTAH	medium	initiate ₁
1	UTAH	high	initiate ₁
	U Acknowledging nod		ack ₁
2	UTAH WHAT	high	initiate ₂
3	UTAH WHAT WE CAN	medium	initiate ₂
4	UTAH WHAT WE CAN GET YOU	low	initiate ₂
5	UTAH WHAT WE CAN GET YOU	low	initiate ₂
	U frown		reqrepair ₂
6	UTAH WHAT WE CAN GIVE YOU TWO	high	repair ₂
	U Attentive nod		
7	UTAH ...TWO HUNDRED DOLLARS	correct	Repair ₂
	U Response		Ack ₂ init ₃

CGU	State	Contents
1	F	s.Addressee utah
2	1	s.mood declarative s.sem.type event s.sem.agent we s.sem.destination you s.sem.modal.possibility can s.sem.speechact.type offer s.sem.event give s.sem.theme twohundred

“Utah, we can give you two hundred dollars”



#	Incremental ASR results	Confidence	Grounding act	CGU	State	Contents
0	UTAH	medium	initiate ₁			
1	UTAH	high	initiate ₁	1	F	s.Addressee utah
	U Acknowledging nod		ack ₁			
2	UTAH WHAT	high	initiate ₂	2	1	s.mood declarative s.sem.type event
3	UTAH WHAT WE CAN	medium	initiate ₂			s.sem.agent we s.sem.destination you
4	UTAH WHAT WE CAN GET YOU	low	initiate ₂			s.sem.modal.possibility can
5	UTAH WHAT WE CAN GET YOU	low	initiate ₂			s.sem.speechact.type offer
	U frown		reqrepair ₂			s.sem.event give s.sem.theme twohundred
6	UTAH WHAT WE CAN GIVE YOU TWO	high	repair ₂			
	U Attentive nod					
7	UTAH ...TWO HUNDRED DOLLARS	correct	Repair ₂			
	U Response		Ack ₂ , init ₃			

“Utah, we can give you two hundred dollars”



#	Incremental ASR results	Confidence	Grounding act	CGU	State	Contents
0	UTAH	medium	initiate ₁			
1	UTAH	high	initiate ₁	1	F	s.Addressee utah
	U Acknowledging nod		ack ₁			
2	UTAH WHAT	high	initiate ₂	2	1	s.mood declarative s.sem.type event s.sem.agent we s.sem.destination you s.sem.modal.possibility can s.sem.speechact.type offer
3	UTAH WHAT WE CAN	medium	initiate ₂			
4	UTAH WHAT WE CAN GET YOU	low	initiate ₂			
5	UTAH WHAT WE CAN GET YOU	low	initiate ₂			
	U frown		reqrepair ₂			s.sem.event give s.sem.theme twohundred
6	UTAH WHAT WE CAN GIVE YOU TWO	high	repair ₂			
	U Attentive nod					
7	UTAH ...TWO HUNDRED DOLLARS	correct	Repair ₂			
	U Response		Ack ₂ init ₃			

“Utah, we can give you two hundred dollars”



Incremental ASR results Confidence Grounding act

0	UTAH	medium	initiate ₁
1	UTAH	high	initiate ₁
	U Acknowledging nod		ack ₁
2	UTAH WHAT	high	initiate ₂
3	UTAH WHAT WE CAN	medium	initiate ₂
4	UTAH WHAT WE CAN GET YOU	low	initiate ₂
5	UTAH WHAT WE CAN GET YOU	low	initiate ₂
	U frown		reqrepair ₂
6	UTAH WHAT WE CAN GIVE YOU TWO	high	repair ₂
	U Attentive nod		
7	UTAH ...TWO HUNDRED DOLLARS	correct	Repair ₂

CGU	State	Contents
1	F	s.Addressee utah
2	F	s.mood declarative s.sem.type event s.sem.agent we s.sem.destination you s.sem.modal.possibility can s.sem.speechact.type offer s.sem.event give s.sem.theme twohundred
3	1	...

U Response Ack₂, init₃



Current & Future work

- Evaluation with Users
- Full integration with Dialogue manager/NVBG:
 - grounding, turn-taking, obligation, & intention for utterance planning, grounding + non-grounding motivations
- Degrees of Grounding (Clark & Shaefer, Roque & Traum 2008, 2009)
- Two views of incremental grounding
 - Revising views on current utterance meaning
 - Change model only on system action or completion
 - Revising state created by previous update
 - More “internal” repairs



Outline of Course

- Preliminaries: representation, agency, communication
- Common Ground: How it is modeled and achieved
- Clark & Schaefer's Model of Grounding
- Computational Models of Grounding I: Brennan & Cahn
- Speech Acts and Dialogue Acts
- Multi-functionality of Utterances
- Feedback and Error-handling in Spoken Dialogue Systems
- Computational Models of Grounding II: Traum '94
- Miscommunication: The Good, the Bad, and the Ugly
- Decision-theoretic models of grounding
- Multi-modal Grounding
- Multiparty Grounding
- Incremental Grounding
- Degrees of Grounding
- Applications of Grounding Analysis



Open Problems

- Better models of grounding
 - Covering more (combinations of) phenomena
 - Multiparty,
 - multimodal
 - degrees of grounding,
 - multiple genres,
 - cultural and social differences
 - Automated recognition of acts
 - Decision models
 - Combinations with other dialogue goals
- Applications
 - More human-like virtual characters
 - Collaborative learning
 - Discourse analysis
 - Medical/diagnosis
 - ?

Thank You!



- Questions?