ESSLLI2015 Advanced Course on Computational Models of Grounding in Dialogue

David Traum traum@ict.usc.edu

Lecture 4: Thursday August 13th, 2015

USC

ICT
INSTITUTE FOR CREATIVE TECHNOLOGIES

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
- Degrees of Grounding
- Incremental Grounding
- Applications of Grounding Analysis

REVIEW OF YESTERDAY

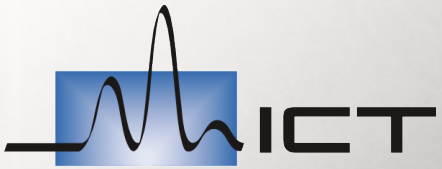
USC

ICT

INSTITUTE FOR CREATIVE TECHNOLOGIES

MISCOMMUNICATION: THE GOOD, THE BAD, AND THE UGLY

USC

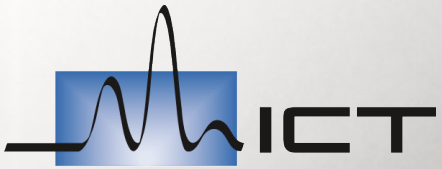


Miscommunication: Conclusions

- Miscommunication is Omnipresent
  - Perfect communication possible only in limited circumstances
- Some miscommunication not worth attention
- Ugly better than Bad
- Ugly -> Good

DECISION-THEORETIC MODELS OF GROUNDING

USC



Clark & Brennan '91: Costs of Grounding

- Formulation Costs
- Production Costs
- Reception Costs
- Understanding Costs
- Start-up Costs
- Delay Costs
- Asynchrony Costs
- Speaker Change Costs
- Display Costs
- Fault Costs
- Repair Costs

Traum & Dillenbourg ’96, ‘98

$$U(\alpha \rightarrow \mu) \propto \frac{GC(\mu) * (G_{\alpha}(\mu) - G(\mu))}{C(\alpha)}$$

- Utility of performing α to ground μ
- GC= Grounding criterion
- Current groundedness vs groundedness after alpha
- considering collaborative cost to both participants in dialogue of performing α.

Also consider utility of other actions for μ, and other effects of α, and other goals

Paek & Horvitz 2000: Conversation as Action Under Uncertainty

- **Quartet System: Bayesian model of grounding**
- **Tested with Bayesian Receptionist, and Presenter**
- **Value of Information (VOI) analysis**

Skantze 2007: Making Grounding Decisions

- **Grounding Decision Problem: which Types of Grounding moves to perform:**
- **E.g. in response to U: I can see a red building.**
  S (ACCEPT): *Ok, can you see a tree in front of you?*
  S (DISPLAY): *Ok, a red building, can you see a tree in front of you?*
  S (CLARIFY): *A red building?*

  S (REJECT): *What did you say?*
- **Factors:**

  1. Level of uncertainty
  2. Task-related costs and utility
  3. Cost of grounding actions

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

Clark & Brennan '91: Constraints on Grounding

- 1. *Copresence: A and B share the same physical environment.* In face-to- face conversation, the participants are usually in the same surroundings and can readily see and hear what each other is doing and looking at. In other media there is no such possibility.
- 2. *Visibility: A and B are visible to each other.* In face-to-face conversation, the participants can see each other, and in other media they cannot. They may also be able to see each other, as in video teleconferencing, without being able to see what each other is doing or looking at.
- 3. *Audibility: A and B communicate by speaking.* Face to face, on the telephone, and with some kinds of teleconferencing, participants can hear each other and take note of timing and intonation. In other media they cannot. An answering machine preserves intonation, but only some aspects of utterance tim- ing.

Clark & Brennan '91: Constraints on Grounding

- **4. *Cotemporality: B receives at roughly the same time as A produces.* In most conversations, an utterance is produced just about when it is received and understood, without delay. In media such as letters and electronic mail, this is not the case.**
- ***5. Simultaneity:A and B can send and receive at once and simultaneously.* Sometimes messages can be conveyed and received by both parties at once, as when a hearer smiles during a speaker's utterance. Simultaneous utterances are also allowed, for example, in the keyboard teleconferencing program called *talk,* where what both parties type appears letter by letter in two distinct halves of the screen. Other media are cotemporal but not simultaneous, such as the kind of keyboard teleconferencing that transmits characters only after the typist hits a carriage return.**

Clark & Brennan '91: Constraints on Grounding

- 6. ***Sequentiality: A's and B's turns cannot get out*** **of** ***sequence.*** **In face- to-face conversation, turns ordinarily form a sequence that does not include intervening turns from different conversations with other people. With email, answering machines, and letters, a message and its reply may be separated by any number of irrelevant messages or activities; interruptions do not have the same force.**
- 7. ***Reviewability: B can review A's messages.*** **Speech fades quickly, but in media such as email, letters, and recorded messages, an utterance stays behind as an artifact that can be reviewed later by either of the partners—or even by a third party. In keyboard teleconferencing, the last few utterances stay visible on the screen for awhile.**
- 8. ***Revisability: A can revise messagesfor B.*** **Some media, such as letters and email, allow a participant to revise an utterance privately before sending it to a partner. In face-to-face and telephone conversations, most self-repairs must be done publicly. Some kinds of keyboard teleconferencing fall in between; what a person types appears on the partner's screen only after every carriage return, rather than letter by letter.**

USC

ICT
INSTITUTE FOR CREATIVE TECHNOLOGIES

Clark & Brennan '91: Media constraints on Grounding

SEVEN MEDIA AND THEIR ASSOCIATED CONSTRAINTS

Medium	Constraints
Face-to-face	Copresence, visibility, audibility, cotemporality, simultaneity, sequentiality
Telephone	Audibility, cotemporality, simultaneity, sequentiality
Video teleconference	Visibility, audibility, cotemporality, simultaneity, sequentiality
Terminal teleconference	Cotemporality, sequentiality, reviewability
Answering machines	Audibility, reviewability
Electronic mail	Reviewability, revisability
Letters	Reviewability, revisability

Media and Activity factors in Grounding

- **Clark and Brennan 90**
  - Media influences amount and type of grounding
- **E.g., Traum & Heeman '96: Trains Domain, spoken language, no visual contact**

Category	% utterances
Explicit Ack	52%
Related	29%
Unrelated after Explicit	15%
Other Unrelated	3%
Uncertain	2%

Dillenbourg & Traum 96, 05
Multi-modal computer-mediated grounding

- **Collaborative dyadic interaction**
  - Mystery solving
- **Multiple (distant) modalities**
  - Moo (including 2-3 kinds of chat)
  - Shared Whiteboard
  - Private notebook (stored learned facts)
- **Extended interactions**
  - 45 min – 2 hrs

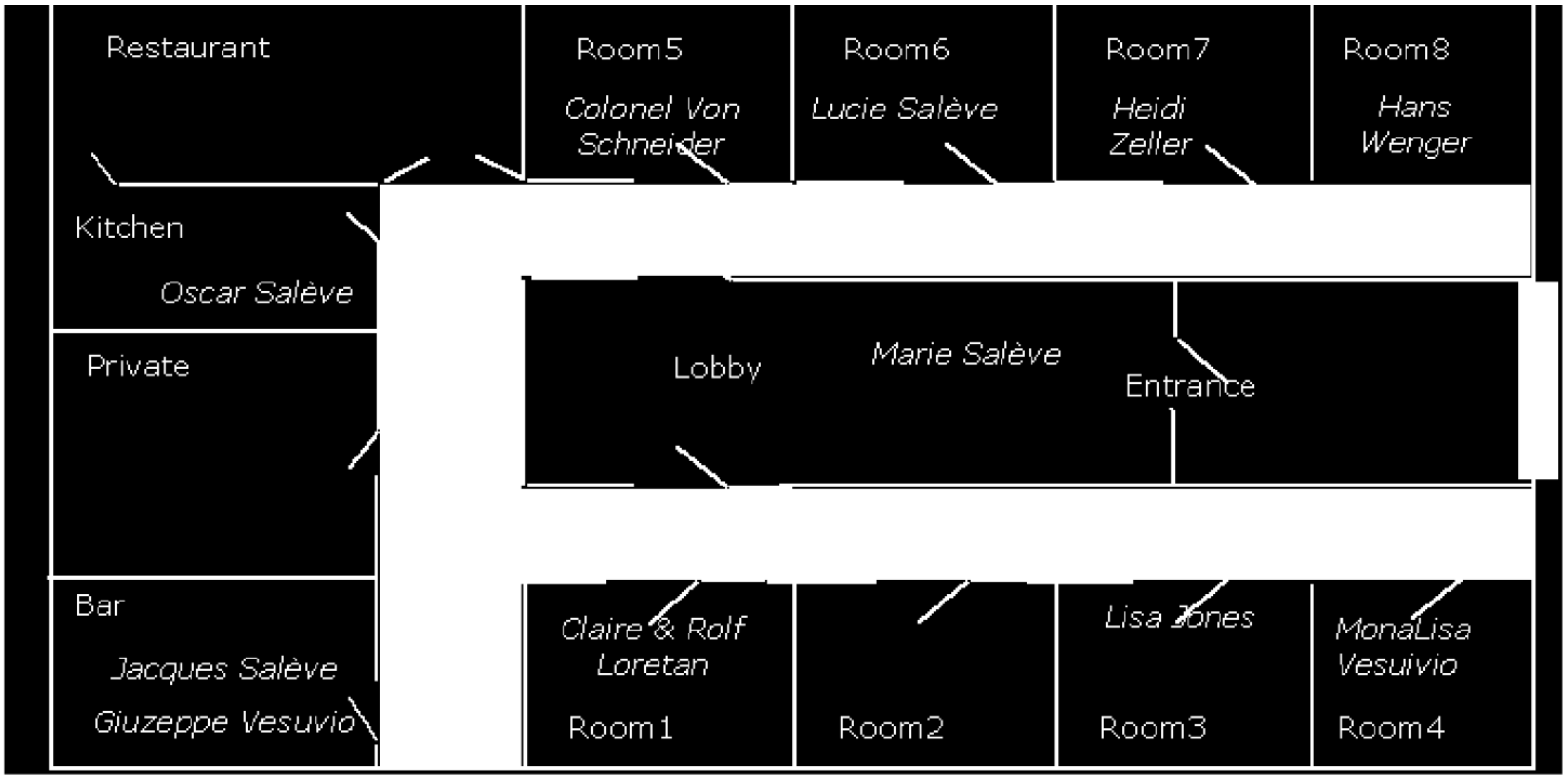
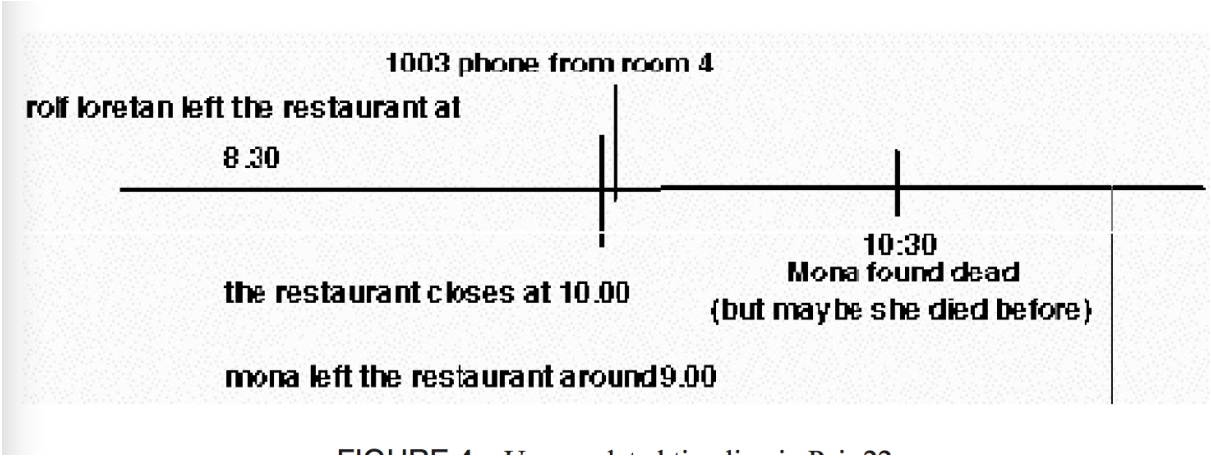


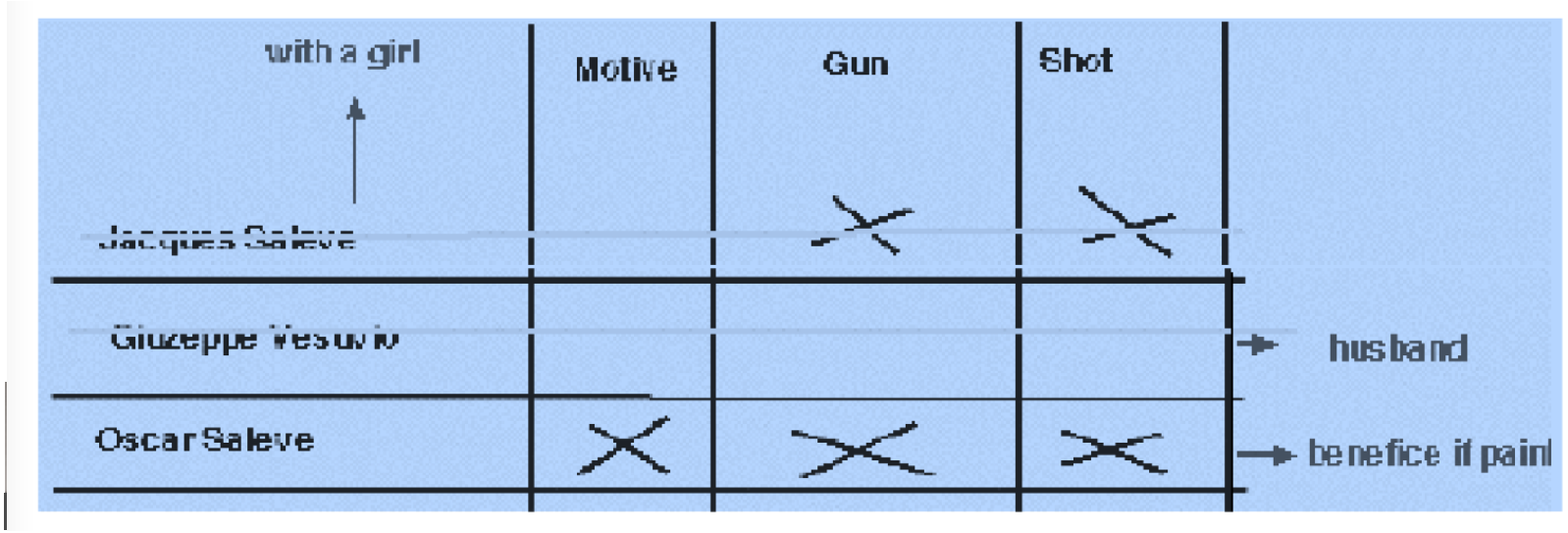
FIGURE 1 A map of the hotel participants received and had to explore.

join sherlock
Hotel du Bout de Nappe: Lower Corridor
Obvious Exits: Lobby (to Lobby), UC (to Upper Corridor), B (to Bar), P (to Private Residence), R1 (to 1), R2 (to 2), R3 (to 3), and R4 (to 4).
Hotel Guest Room: 1
You see Rolf Loretan and Claire Loretan here.
Sherlock is here.
Obvious Exits: Out (to Lower Corridor).
Sherlock asks Claire Loretan about last night
Claire Loretan answers "I was in the restaurant with my husband and the Vesuvios. When the restaurant closed, I briefly went to my room and then joined the others in the bar."
Sherlock asks "Do you know when the bar has closed?"
wisper Did you notice that he is an insurance agent?
I don't understand that.
"what are doing?
You ask, "what are doing?"
ask rolf about the gun
hercule asks Rolf Loretan about the gun
Rolf Loretan answers "it looks like a military issue gun. Why don't you ask that Colonel?"
Sherlock says "Forget it. I thought it could help if we make a tab with the informations about where were th people at what time."
"Actually sounds a good idea.
You say, "Actually sounds a good idea. "
"I think we should find more information about the gun
You say, "I think we should find more information about the gun"

FIGURE 2 An excerpt from the MOO window.

Example Whiteboard constructions





Knowledge Categories

TABLE 4
Content Categories for Analyzing Interactions

Category	*Subcategory*	*Content and Examples*
Task knowledge	Facts	Utterances that contain information directly obtained from the Moo by the participants (e.g., "Rolf was a colleague of the victim"). These are often word-for-word repetitions of the answer given by a suspect
	Inferences	An utterance that involves some interpretation by the participant (e.g., "Helmut had no motive to kill").
Management		Utterances about how to proceed: How to collect information (which suspects, which rooms, which questions, …), how to organize data, how to prune the set of possible suspects, who does what in the pair, and so on. Utterances regarding spatial positions were generally related to strategy issues and were hence included in this category.
Metacommunication		Utterances about the interaction itself, such as discussing delay in acknowledgment (e.g., "Sorry I was busy with the whiteboard") or establishing conversational rules (e.g., "We should use a color coding").
Technical problems		Utterances where one participant asks his partner how to perform a particular action in the MOO (e.g., "I can't read my notebook").

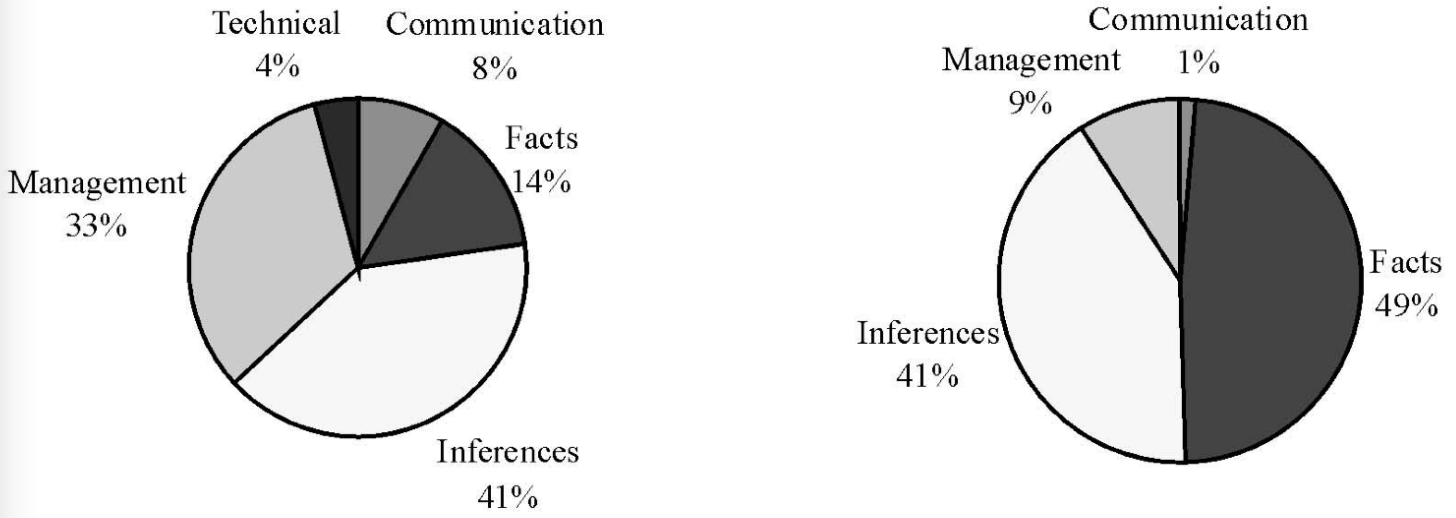


FIGURE 9 Classification of the content of interactions in MOO dialogues (left side) and on whiteboard notes (right side).

Dillenbourg & Traum

TABLE 2
Levels of Mutuality of Knowledge

If Agent A Wants to Communicate Information X to Agent B, A May Get Different Information/Feedback About the Extent to Which B Shares X:	
(Level 1) access: A can infer that B can (not) access X	For instance, in a virtual space, if A knows that B is in room 7 and that information X is available in room 7, then A knows that B can access X. If A knows that X is only available in Room 8, and B is not in room 8, A knows B can't access X.
(Level 2) perception: A can infer that B has (not) perceived X	For instance, if A writes a note on the whiteboard and B moves that note, A can infer that B has seen it (and probably read it). Lack of perception is harder to infer, except for cases of lack of access or behavior that is inconsistent with understanding, when understanding is simple given perception.
(Level 3) understanding: A can infer that B has (mis-)understood X	For instance, in a virtual space, if A says "let's ask *him* a few questions" and B moves to the room where "him" is located, then A can infer that B knows who has been referred to as "him." If B goes to the wrong room, or asks for repair, A can infer misunderstanding or lack of understanding.
(Level 4) agreement: A can infer that B (dis-)agrees on X	For instance, if A proposes B goes to room 7 and B goes there, A can infer that B agrees. If A writes a note on the whiteboard and B draws a red cross on the top this note,[a] A can infer that B disagrees.

Cross-grounding

TABLE 3
Frequency of Acknowledgment by Modality

Row is Acknowledged by Column	*Moo Actions*	*MOO Messages*	*Whiteboard*
MOO actions	2	10	0
MOO messages	42	1,025	34
Whiteboard	0	37	35

TABLE 5
Acknowledgment Rate in Different Content Categories

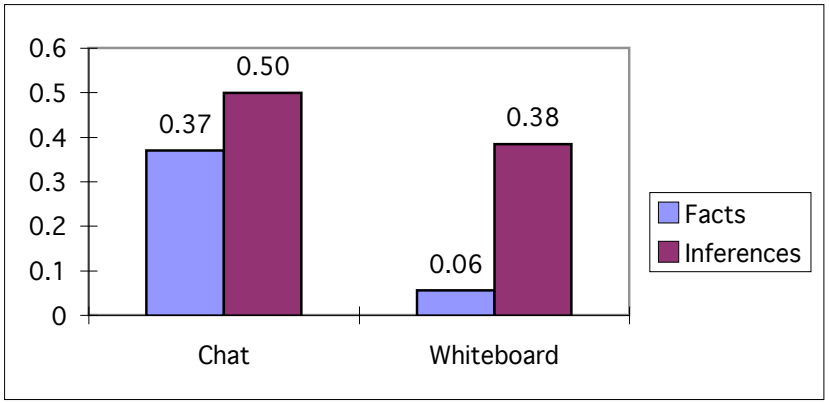
Content of Interactions	*Acknowledgment Rate (%)*
Task knowledge	38
Facts	26
Inferences	46
Task management	43
Metacommunication	55
Technical problems	30
All categories	41

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%
All categories	*41%*

- **Grounding by Category & Medium**



USC
ICT
INSTITUTE FOR CREATIVE TECHNOLOGIES

Impact of grounding rate on repetition

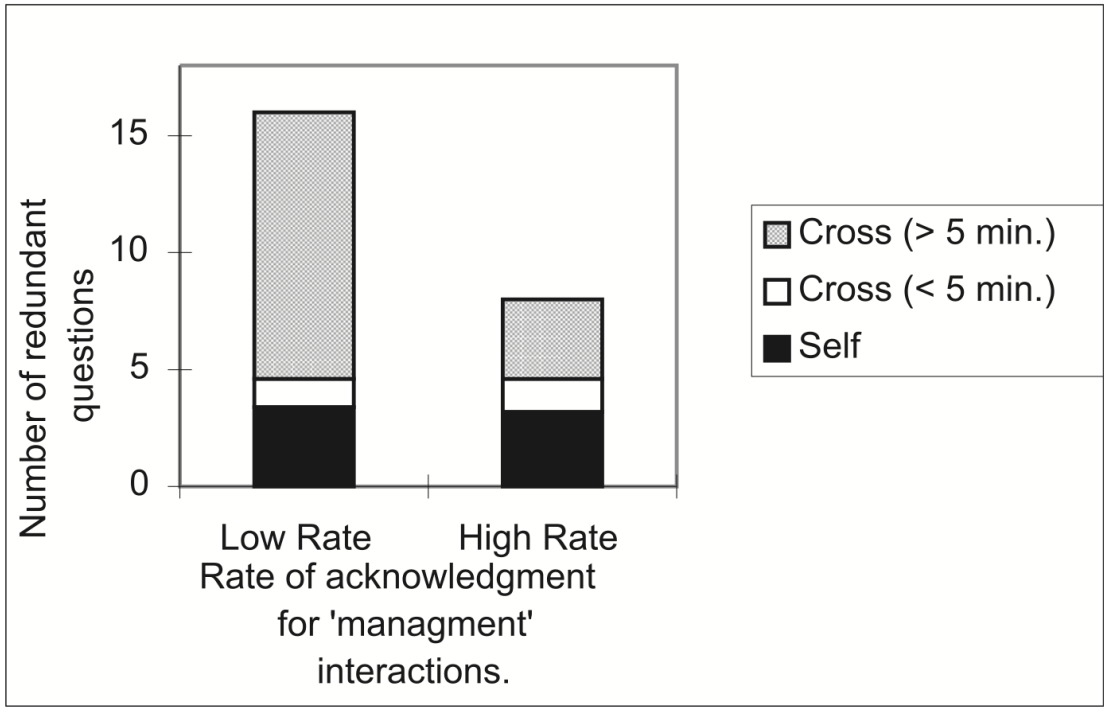


FIGURE 3 Comparison between the number of redundant questions asked by the low acknowledgers (on task management interactions) and high acknowledgers.

USC

ICT
INSTITUTE FOR CREATIVE TECHNOLOGIES

Towards a Model of Face-to-Face Grounding

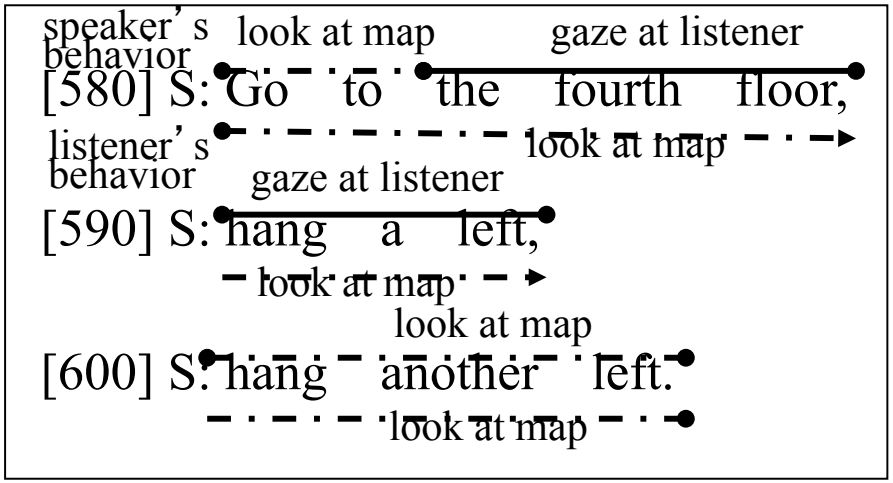
Yukiko Nakano (RISTEX)

Gabe Reinstein & Tom Stocky (Media Lab)

Justine Cassell (MIT Media Lab & Northwestern University)

Questions

S is giving a direction while sharing a map with H



- How do people use nonverbal behaviors to ground information in face-to-face?
- How can a model of face-to-face grounding be applied to human-computer interaction?

USC

ICT
INSTITUTE FOR CREATIVE TECHNOLOGIES

Empirical Study

- **Task: Students from 2 universities give directions to one another**
- **Design**
  - Face-to-face condition (F2F): two subjects share a map drawn by the direction-giver sitting between them.
  - Shared reference condition (SR): L-shaped screen lets them share a map, but not see the other's face and body.

USC

ICT
INSTITUTE FOR CREATIVE TECHNOLOGIES

Data Coding

- **Coding verbal behaviors**
  - Unit of analysis: Utterance Unit (corresponds to single intonational phrase. [Nakatani & Traum 1999])
  - Categories of UUs: using part of DAMSL coding scheme
    - Acknowledgement
    - Answer
    - Information-request (Info-req)
    - Assertion
- **Coding nonverbal behaviors**
    - Gaze at Partner (gP)
    - Gaze at Map (gM)
    - Gaze Elsewhere (gE)
    - Head Nod (Nod)

Coding of NV Status of Dyad

Combinations of NVs		Listener's behavior			
		gP	gM	gMwN	gE
Speaker's behavior	gP	gP/gP	gP/gM	gP/gMwN	gP/gE
	gM	gM/gP	gM/gM	gM/gMwN	gM/gE
	gMwN	gMwN/gP	gMwN/gM	gMwN/gMwN	gMwN/gE
	gE	gE/gP	gE/gM	gE/gMwN	gE/gE

Nonverbal status shift within and between a UU were counted, and used as nonverbal data.

Results: Effect of Access to Body

- **In non-F2F, speakers present information in smaller chunks and take more time. In F2F, more information is conveyed in one UU, which takes less time.**
- **The number of NV shifts in non-F2F is less than half of F2F**
- **Therefore, access to interlocutor's body affects conversation, suggesting that nonverbal behaviors used as communicative signals.**

Mean length of conversation (min)	F2F (3.24) < SR (3.78)	p<.07
Mean length of Utterance Unit (UU) (words)	F2F (5.26) >SR (4.43)	p<.01
The number of NV shifts	F2F (887) > SR (425)	p<.01

Results: NVs as Communicative Signal

- **Correlation between verbal and nonverbal behaviors**

	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)

- Usage of nonverbal signals is different depending on type of conversational action.
- Therefore, these are used as positive evidence of understanding in F2F conversation.

Results: Function of NV Signals

- **Correlation between speaker and listener behavior**

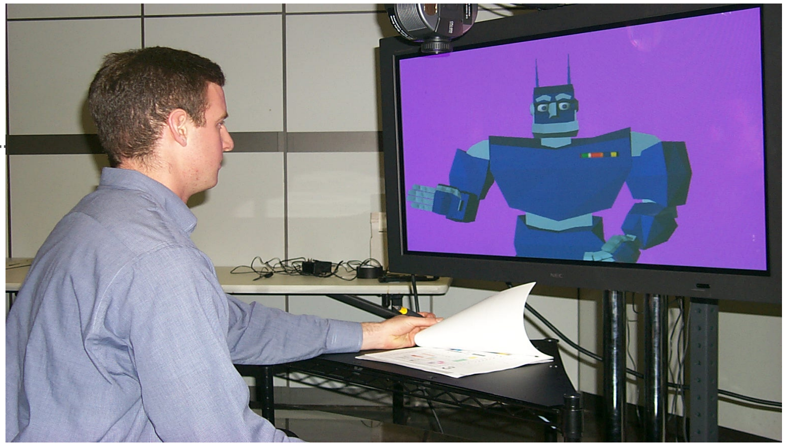
  How listener's nonverbal signals affects speaker's following action

```
[U1] S: And then, you'll go down this little
        corridor.
[U2-a] S: It's not very long. (elaboration)
[U2-b] S: Then, take a right. (go-ahead)
```

- In Assertion, when listener keeps gazing at speaker, speaker's next UU is an elaboration of previous UU 73% of the time.
- When listener keeps gazing at map, only 30% of next UU is elaboration.
- Therefore, speakers interpret listeners' continuous gaze as evidence of not-understanding, and add more explanation for ungrounded message.

System: MACK

Media lab Autonomous Conversational Kiosk

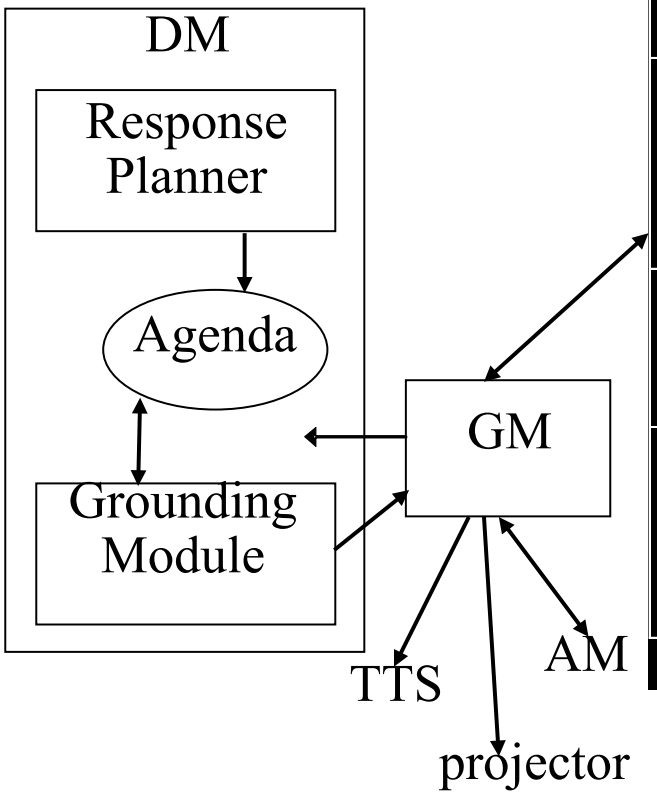


- **Appearance**
  - *Life-sized animated robot*
- **Knowledge base**
  - *Media Lab's projects, research groups, and directions about how to find them.*
- **Input**
  - *Speech*
  - *Pen gestures on shared paper map*

USC

ICT INSTITUTE FOR CREATIVE TECHNOLOGIES

Generating Nonverbal Signals



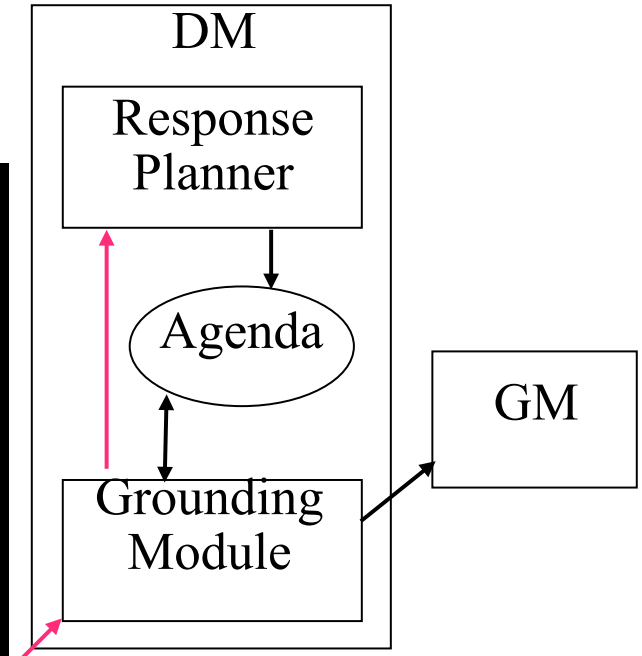
Mack's nonverbal signals

UU type	probability	MACK nonverbal behavior	
		within UU	pause
Assertion	0.66	keep *gM*	*gM*
	0.14	shift *gM* to *gP*	*gP*
	0.11	shift *gM* to *gP*	*gM*
	0.09	keep *gP*	*gP*
Answer	0.45	keep *gM*	*gM*
	0.36	keep *gP*	*gP*
	0.18	shift *gM* to *gP*	*gP*
Elaboration	0.47	keep *gP*	*gP*
	0.2	keep *gP*	*gM*
	0.2	keep *gM*	*gM*
	0.13	shift *gM* to *gP*	*gP*

Determining Next Action

Decision table for the next action

Target UU Type	Grounding judgment	Suggested next action
Assertion	grounded	go-ahead: 0.7 elaboration: 0.30
	ungrounded	go-ahead: 0.27 elaboration:0.73
Answer	grounded	go-ahead: 0.83 elaboration: 0.17
	ungrounded	go-ahead: 0.22 elaboration: 0.78

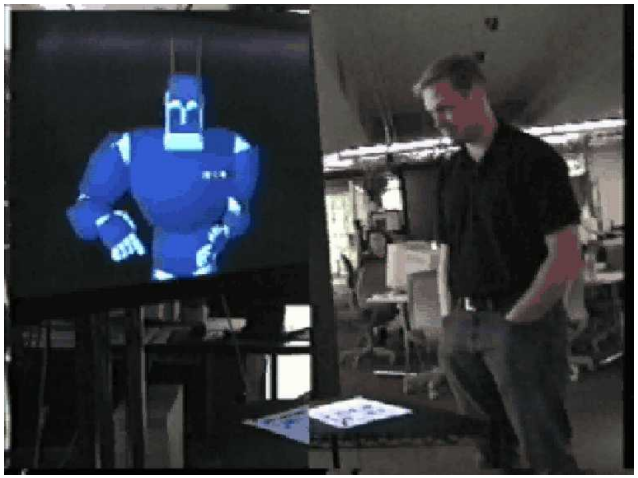


USC

ICT
INSTITUTE FOR CREATIVE TECHNOLOGIES

Preliminary Evaluation

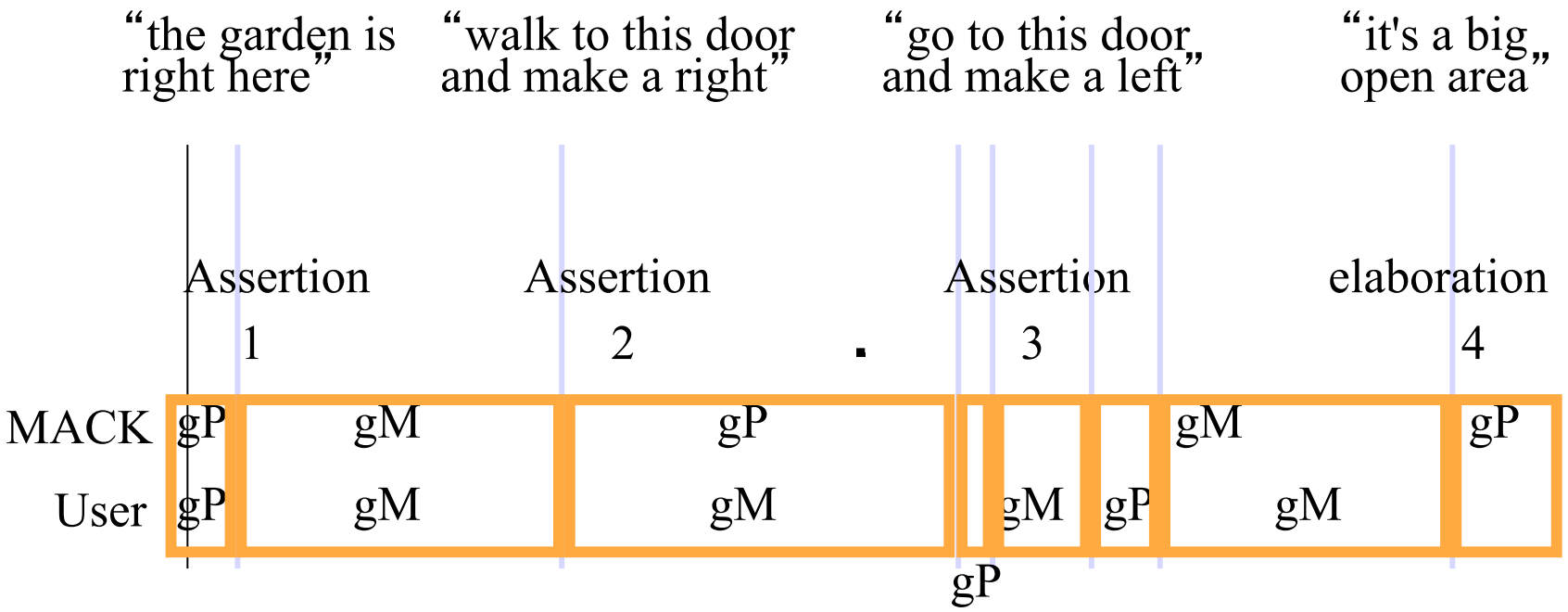
- **Do human users interact with MACK as we expect?**
  - Wizard of Oz setting
  - Naïve users
  - Two versions of MACK



(a) MACK-with-grounding

(b) MACK-without-grounding (neither recognize user's nonverbal signals nor display nonverbal signals of grounding)

Human and MACK-with-grounding



-The NV transition patterns in MACK-with-grounding condition are strikingly similar to those in our empirical study of human-human communication.
-In Without-Grounding condition, user broke these conventions: neither nodded nor spoke.

USC
ICT
INSTITUTE FOR CREATIVE TECHNOLOGIES

Conclusion

- **Empirical results demonstrate nonverbal behaviors used as positive/negative evidence of understanding.**
- **Usage of NV different depending on type of verbal action.**
- **Based on these results, face-to-face grounding mechanism for ECA.**
- **Preliminary evaluation supports model, and shows MACK's potential for interacting with a human user using human-human conversational protocols.**

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

Speaker sub-roles

- Roles
  - Composer
  - Performer
  - Responsible Agent
  - Ratified/unratified
- Examples of split roles
  - Author/performer
  - Speechwriter/politician
  - Foreign language speaker/interpreter
  - Copywriter/spokesman/owner

Hearer sub-roles

- **Roles**
  - Addressee (spoken directly to)
  - Side participant (ratified)
  - Bystander (tolerated)
  - Eavesdropper (unknown)
- **Issues: who gets/has/does/is**
  - Signals from speaker
  - Obligations to speaker
  - Right to become speaker
  - Speaker intend to hear (or intends not to hear)
  - Message designed for
  - Speaker awareness
  - Attention of participants

Activity-oriented talk (Goffman)

- **Main Activity -ratified speakers & addressees**
  - "Off the record" (among speakers, not meant for ratified listeners)
- **Byplay - ratified addresses & side participants**
  - Borderplay (Brandt) - addressees & other ratified
- **Sideplay - unratified overhearers**
- **Crossplay - ratified & unratified**

Speaker -> Addressee signals

- Vocatives & semantic indications
- Message tailored for understanding
- Body orientation
- Gaze
- Gesture
- Mirroring

Addressee -> Speaker signals

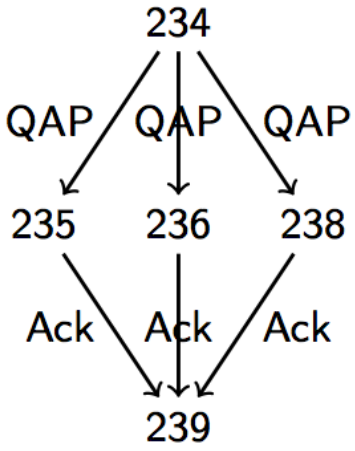
- **Attention**
  - Gaze
  - Posture/orientation
  - mirroring
- **Uptake**
  - Nods, head shakes
  - Facial expressions
  - Eyebrow flashes
- **Turn-taking**
  - Feedback
  - Hands in gesture space
  - gaze

Grounding

- **Two-party**
  - existing models, e.g. Traum 94
  - Signals of understanding from addressee needed for grounding
- **Multi-party**
  - signals from whom? One participant? All?

Settler's of Catan trading dialogue (from Nicholas Asher)

234	gotwood4sheep	anyone got wheat for a sheep ?
235	inca	sorry, not me
236	CheshireCatGrin	nope. you seem to have lots of sheep !
237	gotwood4sheep	yup baaa
238	dmm	i think i'd rather hang on to my wheat i'm afraid
239	gotwood4sheep	kk I'll take my chances then...



USC

ICT
INSTITUTE FOR CREATIVE TECHNOLOGIES

Issues in Multiparty (multi-conversation) Grounding

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.

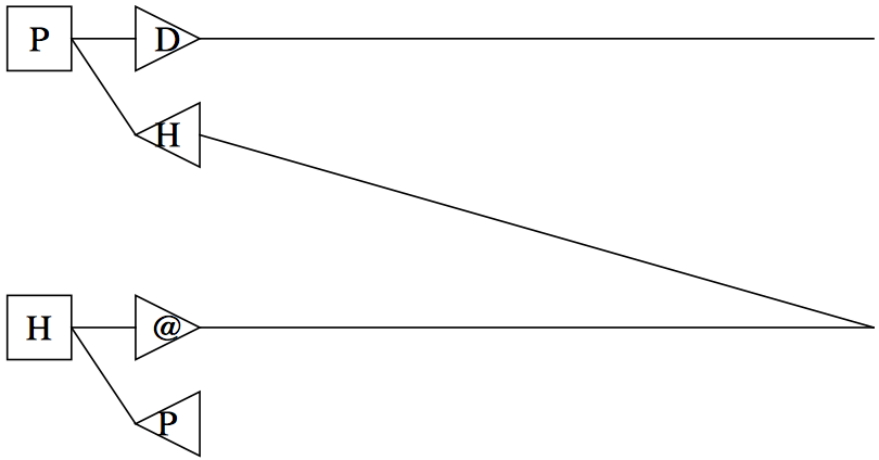
Novick, Walton & Ward '96:

- **Contribution:** is an action by a speaker that has content intended to be conveyed to at least one hearer and that assists some subset of the conversants in establishing mutual belief.
- **Primary Evidence** is evidence *e*' presented by hearer B_i where she believes that she was an intended addressee of A's. That is, B_i believes that A requires evidence from her to believe that they mutually understand *u*.
- **Secondary Evidence** is evidence *e*' presented by hearer B_i when she believes that she was not an intended addressee of A's and/or she believes that A does not require primary evidence of understanding.

Novick, Walton & Ward '96:

- **Presentation Phase**
  - A presents utterance *u* for some subset of B_1, ..., B_n to consider based on the assumption that if that same subset of hearers collectively gives enough primary evidence *e*, he can believe that they understand what he meant by *u*.
- **Acceptance Phase**
  - For all hearers 1 <= *i* <= *n*, B_i accepts utterance *u* by giving either primary or secondary evidence that she understands what A means by *u*. She does so on the assumption that if A registers the evidence, he will believe that A understands.

Novick, Walton & Ward '96:



P: We don't want to harm the people either. That is my concern. We can't harm these young people (inaudible). They were doing things for the best interests of their country—that is all.

H: Well, we don't have any question here of some guy stashing money in his pocket.

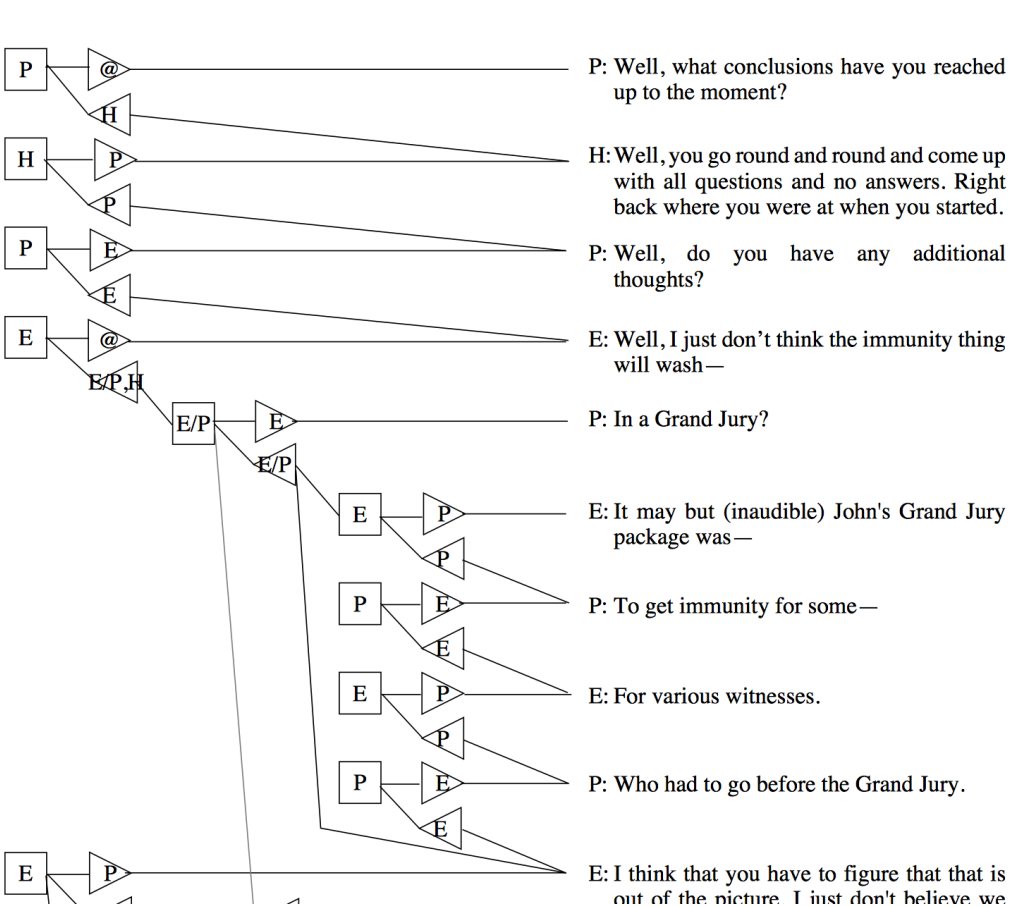
Figure A1. Basic Annotations.

The first utterance is a single contribution by the President, as indicated by the letter P in the first contribution symbol. This contribution is directed toward Dean, as indicated by the letter D in the right-facing triangle which follows the contribution box. This contribution is acknowledged by Haldeman, who responds with a relevant next contribution. Although Haldeman's contribution acknowledges the President's contribution, it is directed toward the entire group; this is indicated by the @ in presentation symbol.

USC

ICT
INSTITUTE FOR CREATIVE TECHNOLOGIES

Novick, Walton & Ward ‘96: Example



USC

INSTITUTE FOR CREATIVE TECHNOLOGIES

MRE (Austin) Multi-party grounding model

- **Implemented:**
  - Multiparty conversation, single addressee
  - Components:
    - State
    - Initiator
    - Responder
    - Contents
- **Multi-addressee**
  - Any addressee acknowledgement grounds
  - Split into multiple single speaker-addressee units
- **Cross-grounding**

Dialogue Acts

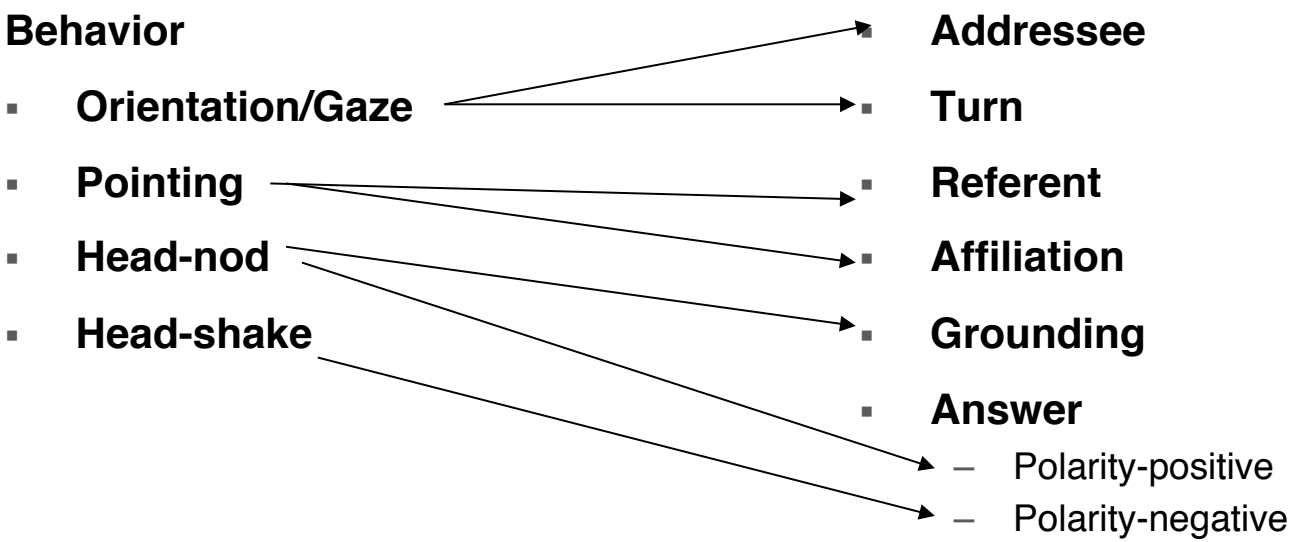
- **Forward**
  - (A1 ^action info-req ^actor <speaker> ^addressee <adr>* ^content <Q> ^type csa)
  - (A2 ^action assert ^actor <speaker> ^addressee <adr>* ^content <P> ^type csa)
- **Backward**
  - (A3 ^action answer ^actor <speaker> ^addressee <adr>* ^answer <SA> ^question <Q> ^type backward)
  - (A4 ^action clarify-parameter ^cand <cand>* ^context <SA> ^parameter <role> ^type backward)
- **Grounding**
  - (A8 ^action initiate ^actor <speaker> ^responder <adr>* ^cgu <cgu> ^content <SA2>* ^conversation <CON>* ^type grounding)
  - (A7 ^action acknowledge ^actor <speaker> ^cgu <cgu> ^content <SA>* ^conversation <CON> ^type grounding)
  - (A5 ^action repair ^actor <speaker> ^cgu <cgu> ^content <SA>* ^conversation <CON> ^parameter <role> ^type grounding ^value <filler>)
  - (A6 ^action request-repair ^actor <speaker> ^cgu <cgu> ^content <SA2>* ^conversation <CON> ^type grounding)

Conversation Object

- **^active-participant +&**
- **^overhearer +& ;#^participant is union of active-participant, overhearer**
- **^mode ;# face-to-face, radio,**
- **^last-utterance ;#stack of utterances part of conversation**
- **^dialogue-history + & ;# speech-input objects**
- **^last-mentions ;#history list of mentioned concepts and recency**
- **^initiative ;# one of the active-participants**
- **^turn ;# one of the active participants**
- **^purpose**
- **^QUD**

- **^grounding +&;#set of cgus**
  - ^initiator ;# one of active participants
  - ^state ;# grounding state: S,F,D,1-4
  - ^dialogue-history + & ;#core speech acts
  - ^obligation + & ;#see social state for details
  - ^commitment + &
  - ^conditional + &
  - ^negotiation-stance + & ;#negotiation objects

Non-verbal Behavior & communicative functions



Ex 1: Use of Orientation for Addressee Recognition

- **Without Vision**
  - Use explicit naming
  - Use context of previous speaker/ addressee
  - Can't tell sometimes
- **Example video**

- **With Vision**
  - Use gaze/orientation to disambiguate addressee
- **Example Video**

USC

INSTITUTE FOR CREATIVE TECHNOLOGIES

MRE Multi-party Addressee Identification Algorithm

1. **If utterance specifies addressee**
   - Vocative
   - not expecting short answer or clarification of person type

   ⇒ Addressee = specified addressee
2. **Else if speaker facing someone**

   ⇨ Addressee = faced participant
3. **Else If current utterance speaker is same as previous utterance speaker**

   ⇒ Addressee = previous addressee
4. **Else If previous speaker≠ current speaker**

   ⇒ Addressee = previous speaker
5. **Else if (active) conversational participant in same conversation**

   ⇒ Addressee = participant
6. **Else ?**

Head Nods & Shakes

- **Treated as “yes” or “no”**
- **Can realize multiple dialogue acts**
  - Answer to a YNQ
  - Grounding (ack)

Ex 2: Show (dis)agreement with head gesture

- **Without Vision**
  - Need explicit verbal utterance
- **With Vision**
  - Use head nod for yes or head shake for no.
  - No need to take the turn

USC

INSTITUTE FOR CREATIVE TECHNOLOGIES

Turn-taking

- **Turn-assigned with specific signals**
  - Question
- **Turn-kept with other signals**
  - Filled pause
- **Underspecified in some cases**
  - Assertion
  - Use of context
    - Initiative holder keeps/takes turn
  - Gaze at end of utterance determines hold/assign turn

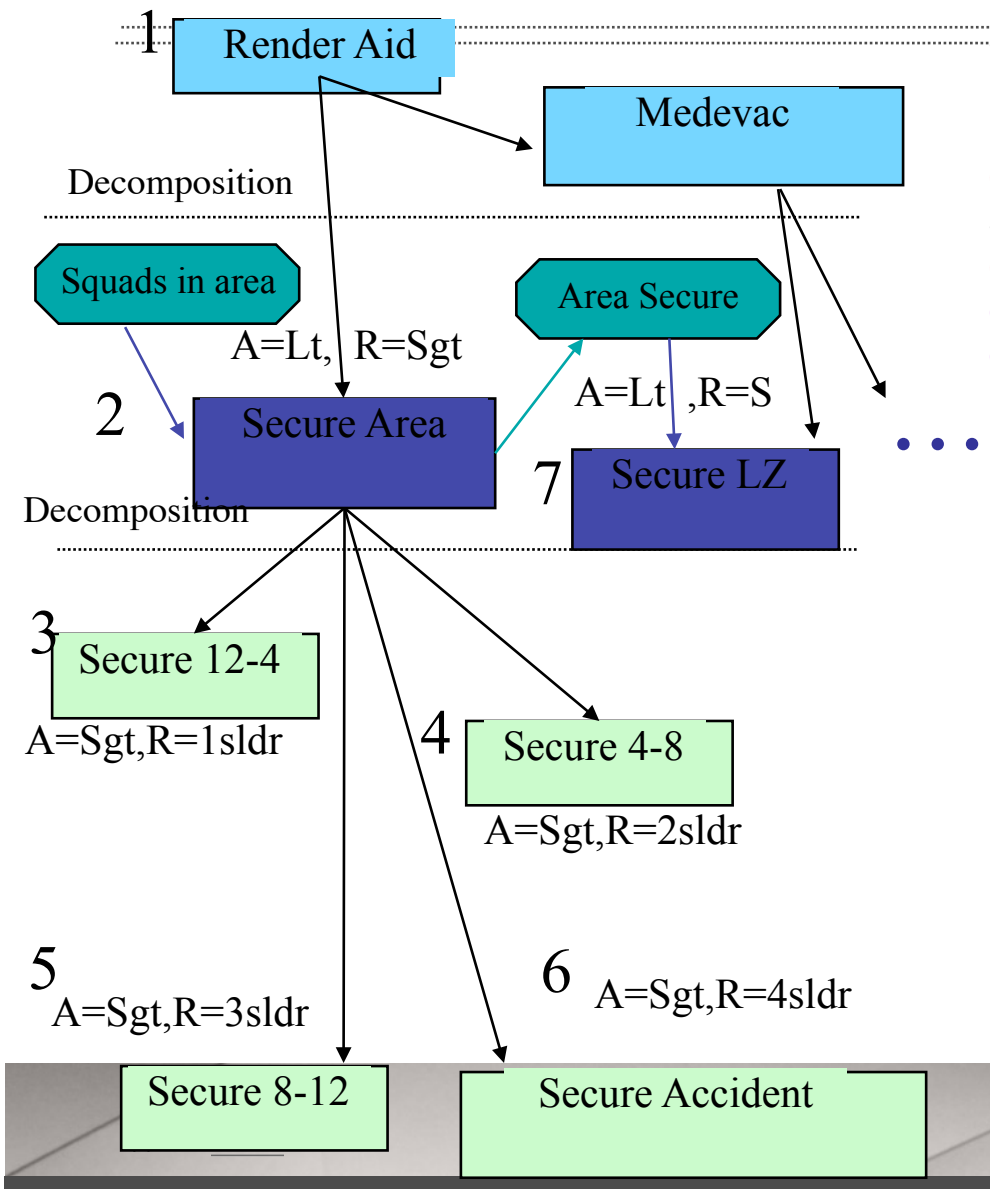
Ex 3: Turn-taking and addressee selection

- **Without Vision**
  - For neutral utterance, Assume previous speaker is current addressee
  - Wait for current speaker's turn to time out (in case he will say more)

- **With Vision**
  - Use orientation to pick addressee
  - Turn given to addressee with neutral utterance
  - No delay needed

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)