ESSLLI2015 Advanced Course on
Computational Models of Grounding in Dialogue

David Traum traum@ict.usc.edu

Lecture 3: Wednesday August 12th, 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

REVIEW OF YESTERDAY

USC

ICT

INSTITUTE FOR CREATIVE TECHNOLOGIES

Brennan & Cahn 1999: Extensions to Clark and Schaefer

1. **All contribution graphs are private models from an individual's point of view**
   - C&S graph seen as composite final product
   - Incrementally constructed, utterance by utterance
2. **Task-specific heuristics for assessing evidence of understanding and grounding criterion**
3. **Principles for embedding contributions: only when not meeting grounding criterion**
4. **Addition of "Exchange" structure: propose and execute**
   - Remove unrooted medial contributions

Issues for Computational Theory of Speech Acts

- **When can an act be recognized**
  - as sincere and successful?
- **What are the effects of performance of an act**
  - On state of hearer and speaker
  - On state of dialogue
- **When should act be performed?**
- **How should act be performed?**

Dialogue Acts – Beyond standard Illocutionary acts

- Sinclair & Coulthard
- Bunt: Dialogue Acts
- Novick: Meta-locutionary acts
- Traum & Hinkelman: Conversation Acts

- Cover multiple dialogue phenomena
- Turn-taking
- Reference
- Grounding
- Discourse relations/ Adjacnecy pairs
- feedback

Multifunctionality

A: Henry, could you take us through these slides?

Turn Assign* to Henry; *Request

H: O..w..k..ay.. just ordering my notes

Turn Accept; Stalling; Accept Request; Inform

Dimensions of communication in dialogue:

- Turn Management
- Time Management
- Task performance
- .....

USC

ICT
INSTITUTE FOR CREATIVE TECHNOLOGIES

Types of Feedback (Allwood et al 92)

- **Levels:**
  - Contact
  - Perception
  - Understanding
  - Attitudinal Reaction
- **Signals types**
  - Request feedback
  - Prepare other
  - Provide
    - Positive
    - negative

Some Styles of Verbal Response

1	**Sys:**	Where do you want to go?
2	**User:**	Boston.
3a	**Sys:**	When would you like to go?
3b		Tell me more about your travel plans.
3c		When would you like to go to Boston?
3d		Do you want to go to Boston?
3e		Did you say Boston?
3f		Boston?
3g		Boston or Austin?
3h		Where?
3i		Please Repeat.

Computational Model (Traum 94)

- **Contribution recast as "DU" (Discourse Unit)**
  - (later "CGU") (Common Ground Unit)
- **Finite state network for CGU, tracking state of groundedness**
- **Set of Grounding acts to affect contents and state**
- **Interpretation and generation rules**

Grounding Model (Traum 94)

State	Entering Act	Preferred Exiting Act
S	———	$Initiate^{I}$
1	$Initiate^{I}$	Ack^{R}
2	$ReqRepair^{R}$	$Repair^{I}$
3	$Repair^{R}$	Ack^{I}
4	$ReqRepair^{I}$	$Repair^{R}$
F	$Ack^{\{I,R\}}$	$Initiate^{\{I,R\}}$ (next DU)
D	$Cancel^{\{I,R\}}$	$Initiate^{\{I,R\}}$ (next DU)

Label	Description
initiate	Begin new DU, content separate from previous uncompleted DUs
continue	same agent adds related content to open DU
acknowledge	Demonstrate or claim understanding of previous material by other agent
repair	Correct (potential) misunderstanding of DU content
Request Repair	Signal lack of understanding
Request Ack	Signal for other to acknowledge
cancel	Stop work on DU, leaving it ungrounded and ungroundable

Next Act	In State						
	S	1	2	3	4	F	D
initiate^{I}	1						
continue^{I}		1			4		
continue^{R}			2	3			
repair^{I}		1	1	1	4	1	
repair^{R}		3	2	3	3	3	
ReqRepair^{I}			4	4	4	4	
ReqRepair^{R}		2	2	2	2	2	
ack^{I}				F	1	F	
ack^{R}		F	F			F	
ReqAck^{I}		1				1	
ReqAck^{R}				3		3	
cancel^{I}		D	D	D	D	D	
cancel^{R}			1	1		D	

Problems with this Model
(later work addressing these issues)

- **Binary grounded/ungrounded decision**
  - No levels of "groundedness" (Roque 2009)
- **Leaves the unit size unspecified** (Visser, DeVault & Traum)
- **Confusability of grounding acts**
  - e.g. repetition = acknowledgment, repair, or request for repair? (Katagiri & Shimojima)
- **Only well-suited for spoken language grounding**
  - Different kinds and meanings of non-verbal feedback (Nakano et al 2003)
  - Less explicit signaling in computer-mediated chat (Dillenbourg & Traum)

Display Act (Katagiri & Shimojima 2000)

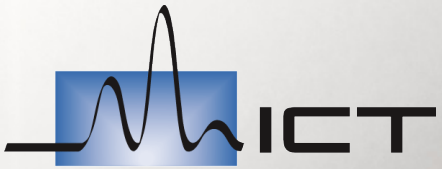
- **Problem for Clark & Shaefer 92 & Traum 94: display of responder's understanding might be acceptance/ acknowledgement, Repair, request repair**
- **Depends on initiator's determination of (in-)correctness and responder's projected certainty.**
- **Propose lower-level "display" act, that can be interpreted by initiator**

Generating Act (α)		Context	Generated Act (β)
Content	Result	Target	
"uh huh"		following p	acknowledgment
"what?"		following p	repair request
display p	High	following p	acknowledgment
display p'	High	following p	repair
display p	Neutral	following p	acknowledgment
display p'	Neutral	following p	repair request
display p	Low	following p	repair request
display p'	Low	following p	repair request

Table 1: Grounding acts generated by echoic responses.

MISCOMMUNICATION: THE GOOD, THE BAD, AND THE UGLY

USC



Definitions (1)

- Communication: performance + interpretation
  - A performs signaling action s in Context C to B, with intent that B recognizes meaning M
  - B observes A perform s' in Context C', and infers meaning M'
- Perfect Communication: M' = M
  - Transparent communication: s = s', C=C'
- Miscommunication: M' ≠M
- Negative Miscommunication (non-understanding):
  $\exists_m$ m ∈ M, m ∉ M'
- Positive Miscommunication (misunderstanding):
  $\exists_m$ m ∈ M', m ∉ M
- Partial Communication:
  $\exists_m$ m ∈ M, m ∉ M' , $\exists_{m'}$ m' ∈ M, m' ∈ M'

Searle ‘83 Intentionality (after Chisholm)

Suppose he is out driving thinking about how he is going to kill his uncle, and suppose his intention to kill his uncle makes him so nervous and excited that he accidentally runs over and kills a pedestrian who happens to be his uncle. Now in this case it is true to say that he killed his uncle and true to say that his intention to kill his uncle was (part of) the cause of his killing his uncle, but not true to say that he carried out his intention to kill his uncle or that his intention was satisfied; because he didn't kill his uncle *intentionally*.

Definitions (2)

- Types of Positive Miscommunication: m ∉ M'
  - Fortuitous Communication

    Intend(A,Comm(A,B,m))

    ¬Intend(A,Comm(A,B,C:s⇒m))
  - Unintended Communication

    ¬ ∃ Intend(A,Comm(A,B,m))

    Bel(A,m)
  - Naive Communication

    ¬ Bel(A,m)

    True(m)
  - True Misunderstanding

    ¬ Intend(A,Comm(A,B,m))

    ¬ Bel(A,m)

Sources of Miscommunication

- Polysemy, ambiguity, vagueness of s
- Different views of meaning interpretation: Bel(A,C:s⇒m), Bel(B,C:s⇒m')
- Different views of context: C ≠ C'
- Noisy channel s ≠s'

Processes to avoid or reduce Miscommunication

- Before communication
  - Adjust context
    - Adjust interpersonal relationships
    - Establish agreement on signal -> meaning conventions
    - Introduce/focus concepts
  - Prepare partner for communication
- During communication
  - Explanations, elaborations
  - Monitoring & 1st turn repair
- Just after communication
  - Grounding & repair

Repair (Schegloff)

- NTRI - signal of problematic understanding
- 3rd turn repair: A recognizes and deals with misunderstanding (indicated by B's response)
  - Initiation, agreement/acceptance, rejection, repair
- 4th turn repair: B recognizes and repairs misunderstanding (usually A's 3rd turn response indicates incompatibility with prior interp)
  - Change of state, id of trouble source

What to do when recognizing misunderstanding?

- Ignore
  - Could involve undesirable commitments
  - May make communication worse
- Repair
  - Could slow down conversation, make less fluent
  - Could indicate more importance than merited (grounding criterion)
  - Undesirable social consequences
    - Loss of face for speaker: unable to communicate properly
    - Loss of face for addressee: unable to interpret properly
- Re-introduce
  - May be difficult
  - May have same consequences as Ignore & repair

Miscommunication: The ugly

- Non-fluent communications
  - Non-understanding
  - Repair
  - Tendency to “get stuck”, impasse
    - Give up on attempted repair
    - Repetitive and Cyclical repair

Example: Monty Python and the Holy Grail

- FATHER: Guards! Make sure the Prince doesn't leave this room until I come and get him.
- GUARD #1: Not to leave the room even if you come and get him.
- GUARD #2: Hic!
- FATHER: No, no. Until I come and get him.
- GUARD #1: Until you come and get him, we're not to enter the room.
- FATHER: No, no, no. You stay in the room and make sure he doesn't leave.
- GUARD #1: And you'll come and get him.
- GUARD #2: Hic!
- FATHER: Right.
- GUARD #1: We don't need to do anything, apart from just stop him entering the room.
- FATHER: No, no. Leaving the room.
- GUARD #1: Leaving the room. Yes. [sniff]
- FATHER: All right?
- GUARD #1: Right.
- GUARD #2: Hic!
- FATHER: Right.

- FATHER: All right?
- GUARD #1: Right. We'll stay here until you get back.
- GUARD #2: Hic!
- FATHER: And, uh, make sure he doesn't leave.
- GUARD #1: What?
- FATHER: Make sure 'e doesn't leave.
- GUARD #1: The Prince?
- FATHER: Yes. Make sure he doesn't leave.
- GUARD #1: Oh, yes, of course.
- GUARD #2: Hic!
- GUARD #1: Ah. I thought you meant him. You know, it seemed a bit daft me havin' to guard him when he's a guard.
- FATHER: Is that clear?
- GUARD #2: Hic!
- GUARD #1: Oh, quite clear. No problems.
- FATHER: Right. Where are you going?
- GUARD #1: We're coming with you.
- FATHER: No, no. I want you to stay here and make sure 'e doesn't leave.
- GUARD #1: Oh, I see. Right.

- GUARD #1: Oh, if-- if-- if-- uhh-- if-- if-- w-- ehh-- i-- if-- if we--
- FATHER: Yes? What is it?
- GUARD #1: Oh, I-- if-- i-- oh--
- FATHER: Look, it's quite simple.
- GUARD #1: Uh...
- FATHER: You just stay here, and make sure 'e doesn't leave the room. All right?
- GUARD #2: Hic!
- FATHER: Right.
- GUARD #1: Oh, I remember. Uhh, can he leave the room with us?
- FATHER: N-- no no. No. You just keep him in here, and make sure he--
- GUARD #1: Oh, yes. We'll keep him in here, obviously. But if he had to leave and we were with him--
- FATHER: No, no, no, no. Just keep him in here--
- GUARD #1: Until you, or anyone else--
- FATHER: No, not anyone else. Just me.
- GUARD #1: Just you.
- GUARD #2: Hic!
- FATHER: Get back.
- GUARD #1: Get back.

Miscommunication: The bad

- Undiscovered Misunderstanding
  - Insufficient grounding
  - Grounded misunderstanding
  - Leads to unfulfilled expectations
    - Lack of subsequent coordination
    - Magnified effects
      - Tragedy: e.g. Romeo & Juliet
      - Wars started without irreconcilable conflict

Miscommunication: The good

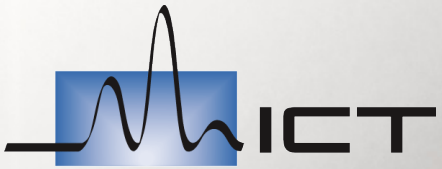
- Recognition of Miscommunication is where thinking starts
  - Diagnosis
  - Planning
  - Reaction
- Awareness of Misunderstanding can lead to awareness of other ways of thinking
  - Ethics (Levinas, Martinovski) - can treat other as different from self, with own value
  - Collaborative learning (Dillenbourg, Chi) - one must reflect on own knowledge to repair
    - Knowledge construction
    - Co-construction

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



Grounding Issues

- **How is a particular grounding act realized?**
- **How important is the grounding?**
  - How useful will it be to the system?
- **What criteria are needed?**
- **How well will a particular act ground its intended content?**
- **And what is the opportunity cost of performing this act?**
  - Is it worth it?

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

Paek & Horvitz 2000: Quartet system Clark's Levels of Analysis:

lowest

- **Channel Level: attempt to open communication channel with some behavior**
- **Signal Level: behavior is intended as a signal**
- **Intention Level: understanding of semantic content occurs**

highest

- **Conversation Level: a joint activity is proposed and responded to**

Paek & Horvitz: System Design

- **Two modules:**
  - **maintenance**
  - **intention**
- **Conversation Control**
  - exchanges info between the modules
  - determines grounding state
  - weighs costs and benefits
  - evaluates module performance & reliability

Signal & Channel level

Intention level

Conversation level

USC

ICT
INSTITUTE FOR CREATIVE TECHNOLOGIES

Paek & Horvitz: Bayesian Maintenance Module

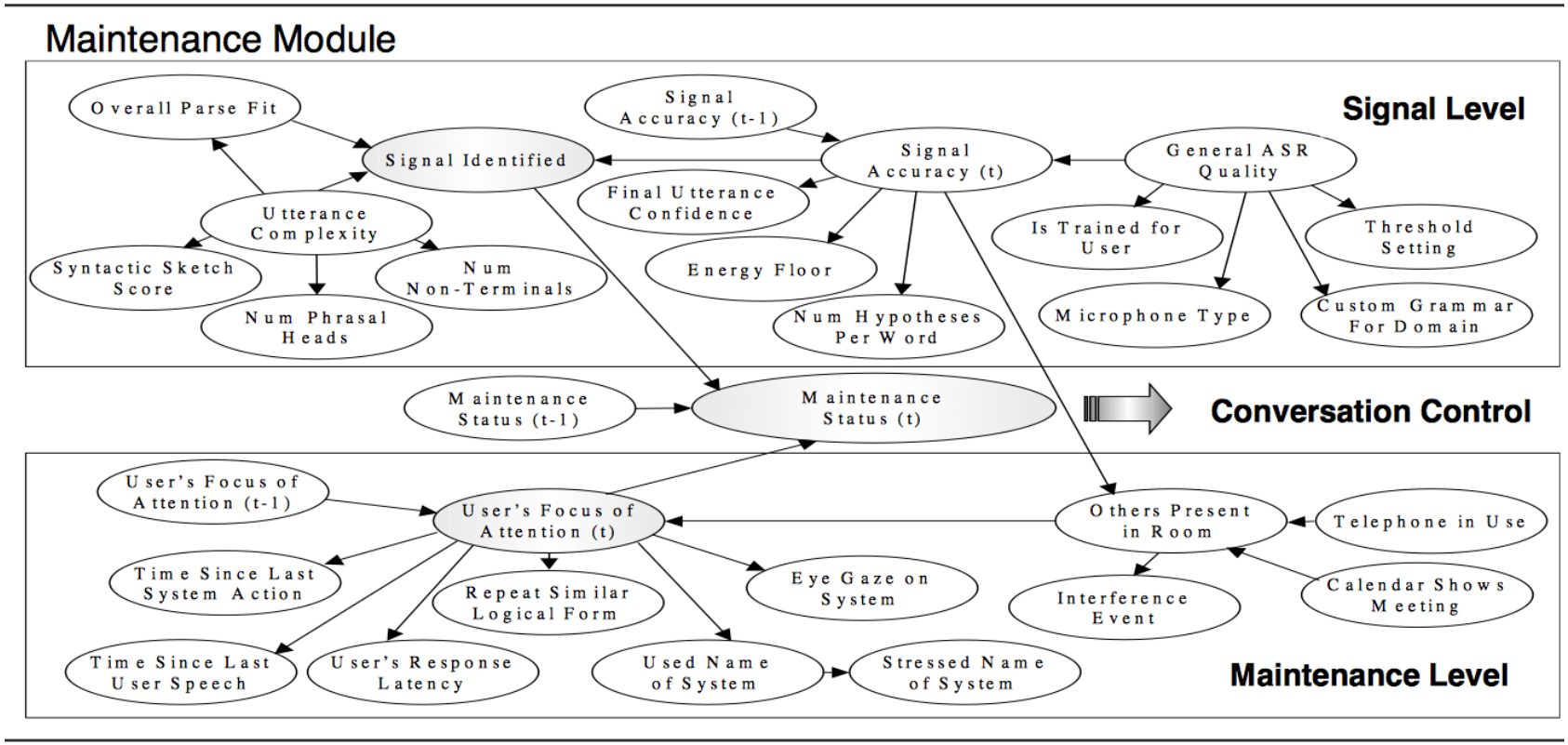
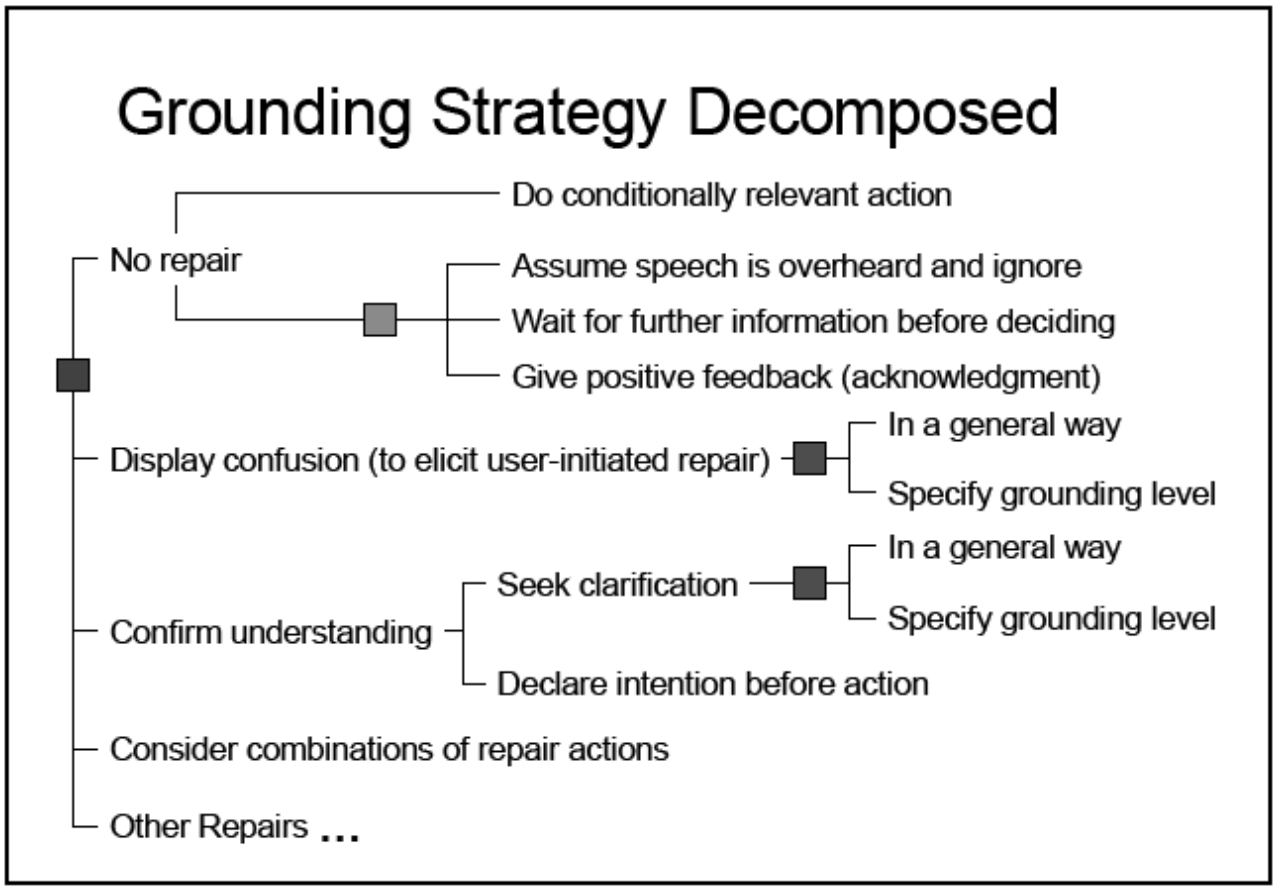


Figure 2. Portions of the temporal Bayesian networks used in the Quartet Maintenance Module.

USC

ICT
INSTITUTE FOR CREATIVE TECHNOLOGIES

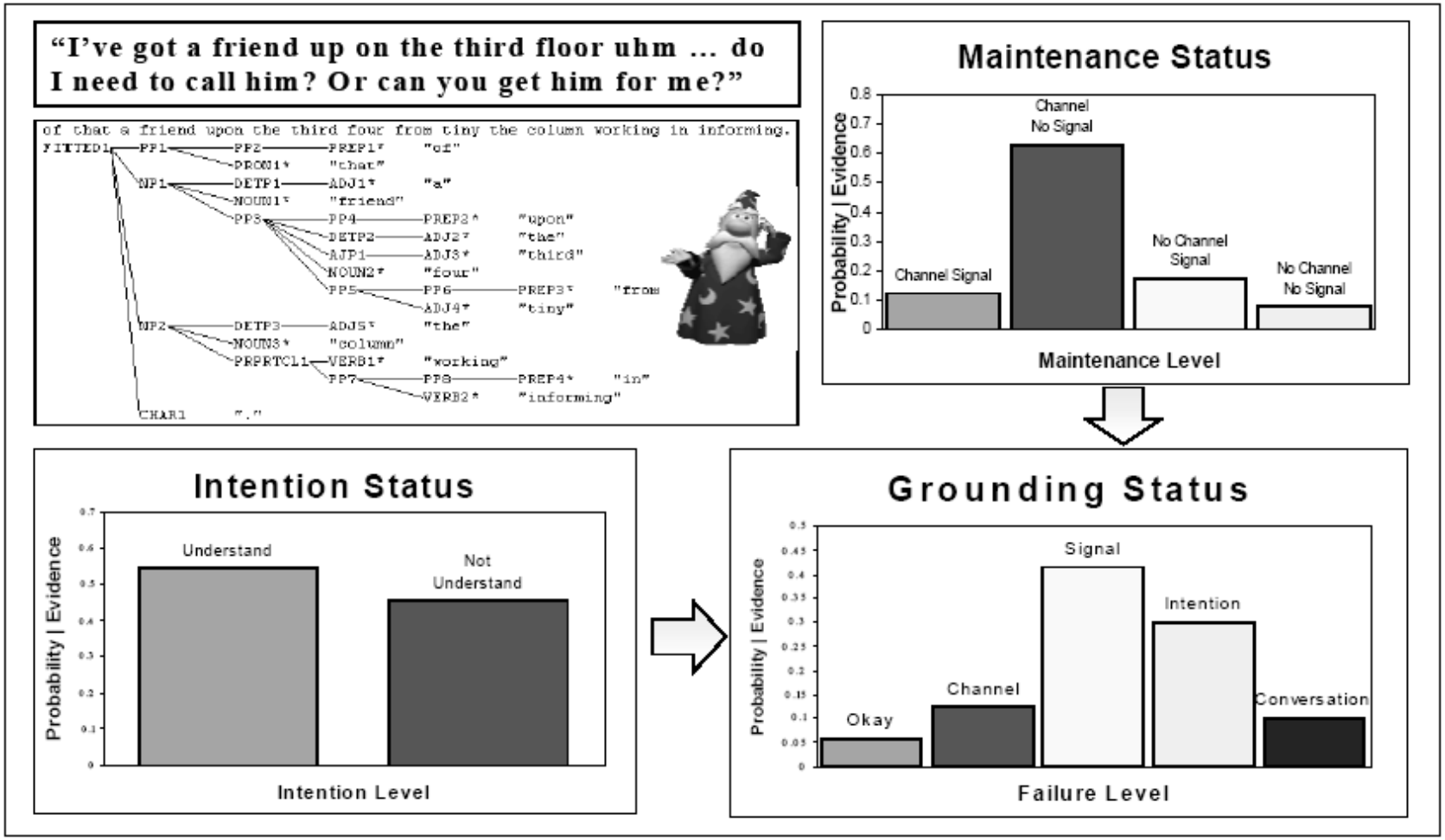
Grounding Strategies



USC

ICT
INSTITUTE FOR CREATIVE TECHNOLOGIES

Signal Failure



USC

ICT
INSTITUTE FOR CREATIVE TECHNOLOGIES

Updating on the basis of evidence

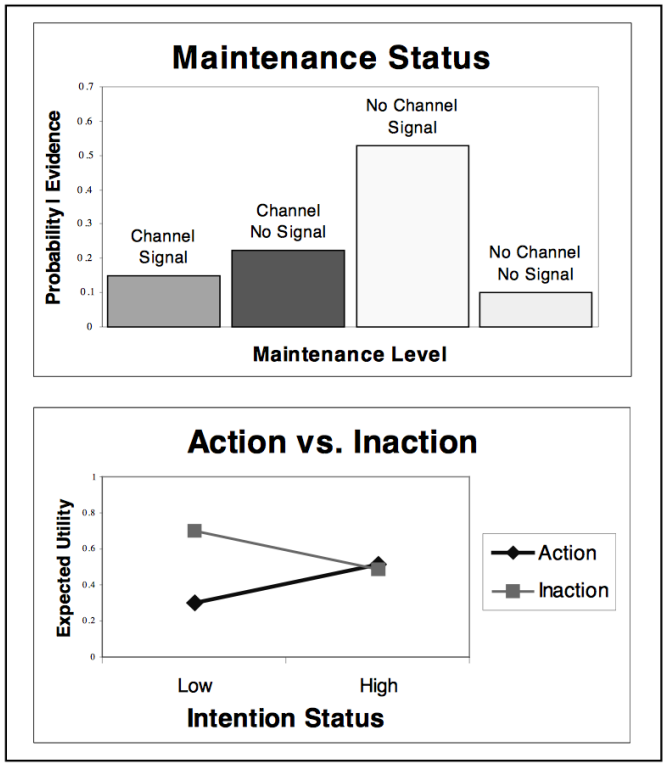


Figure 8. Change in expected utility as Presenter tries to distinguish overheard from intended speech.

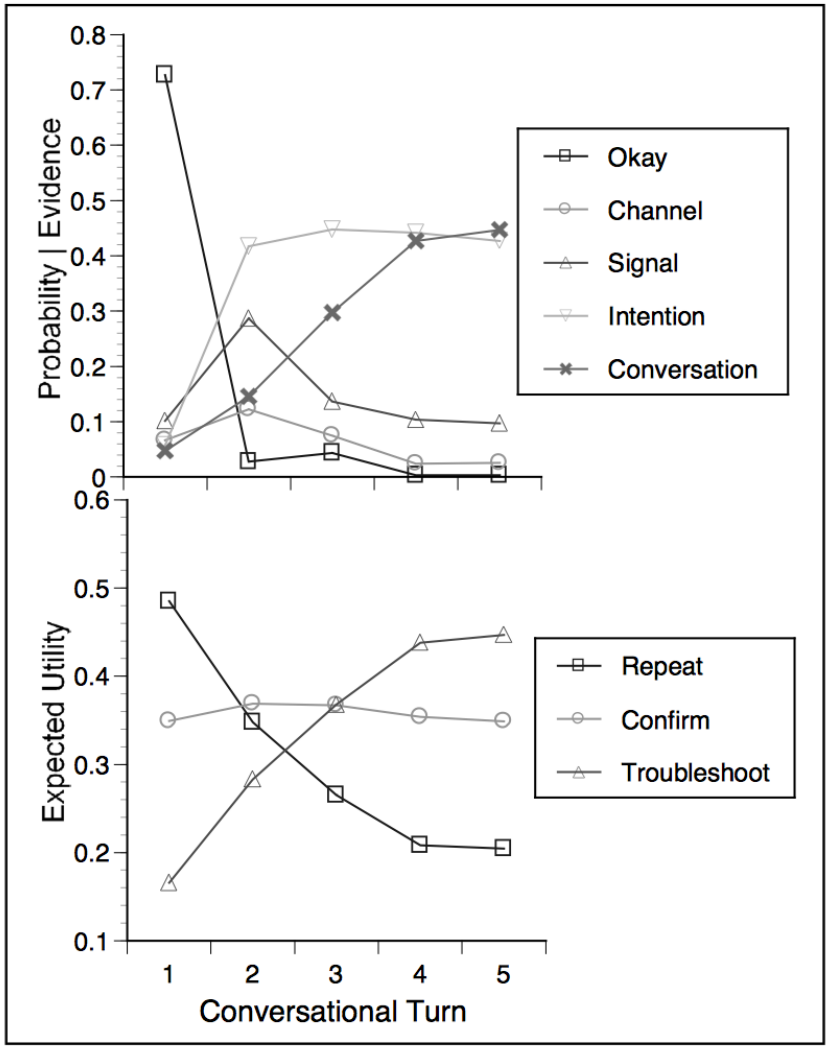


Figure 9. Two graphs demonstrating how grounding strategies adapt to the Grounding Status.

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

Skantze 2007: Model
Estimate utilities from data (from Higgins system)

- **Cost function for grounding act, given correctness**

Table 1: Costs for different grounding actions, given the correctness of the recognition (COR=Correct, INC=Incorrect).

Action,Hyp	Costs
ACCEPT,COR	No cost
ACCEPT,INC	The number of extra syllables the misunderstanding adds to the dialogue (*SylMis*).
DISPLAY,COR	Grounding dialogue *(SylDispCor)*.
DISPLAY,INC	Grounding dialogue (*SylDispInc*). Risk that the user does not correct the system *(P(Fail|Disp,Inc))* times the consequences of a misunderstanding (*SylMis*).
CLARIFY,COR	Grounding dialogue *(SylClarCor)*. Risk that the user does not confirm the system *(P(Fail|Clar,Cor))* times the syllables for recovering the rejected concept *(SylRec)*.
CLARIFY,INC	Grounding dialogue *(SylClarInc)*
REJECT,COR	The number of syllables it takes to receive new information of the same value as the rejected concept *(SylRec)*.
REJECT,INC	No cost

- **Probability measure for ASR/NLU hypothesis being correct/incorrect**
  - (derived from ASR confidence score)
- **Derived Costs for Acts**

Action	Expected cost
ACCEPT	*P(incorrect)* x *SylMis*
DISPLAY	*P(correct)* x *SylDispCor* + *P(incorrect)* x *(SylDispInc + P(Fail|Disp,Inc)* x *SylMis)*
CLARIFY	*P(correct)* x *(SylClarCor + P(Fail|Clar,Cor)* x *SylRec)* + *P(incorrect)* x *SylClarInc*
REJECT	*P(correct)* x *SylRec*

Skantze – Example Cost functions

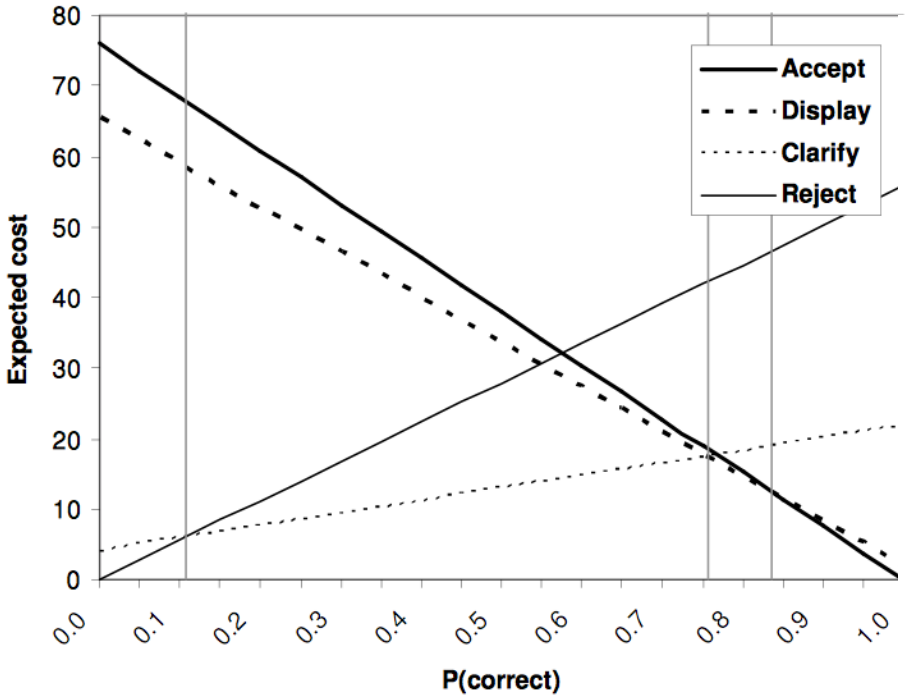


Figure 1: Cost functions and confidence thresholds for grounding the concept MAILBOX after "I can see a mailbox".

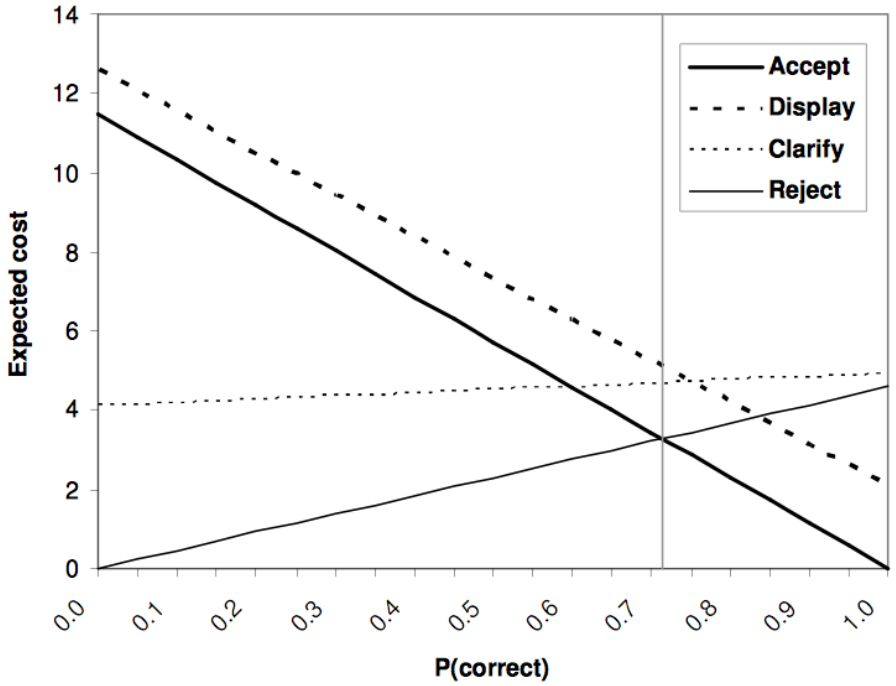


Figure 2: Cost functions and confidence thresholds for grounding the concept TWO after "I can see a two storey building".

USC

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)

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