

Dialogue

Bill MacCartney and Christopher Potts

CS 244U: Natural language understanding

Mar 6



Overview

- As we discussed during our first class meeting, the Turing Test gives dialogue a special place in AI/NLU.
- Dialogue requires rich knowledge bases.
- Dialogue is always situated — many aspects of it are grounded in the immediate discourse situation.
- A realistic dialogue system must also master a wide range of challenging linguistic tasks, including:
 - acknowledging others' contributions;
 - managing the flow of information based on others' cues;
 - structuring utterances so as to engage properly with the preceding discourse;
 - managing the extra pragmatic inferences that others are likely to draw from its contributions.

Eliza

- 1 User: You are like my father in some ways.
- 2 ELIZA: What resemblance do you see
- 3 User: You are not very aggressive but I think you don't want me to notice that.
- 4 ELIZA: What makes you think I am not aggressive?
- 5 User: You don't argue with me.
- 6 ELIZA: Why do you think I don't argue with you?
- 7 User: You are afraid of me.
- 8 ELIZA: Does it please you to believe I am afraid of you?

Rewrite user's reply by (i) swapping 1st and 3rd person, (ii) interpolating stock phrases, and (iii) using scores to rank possible transformations.

HAL

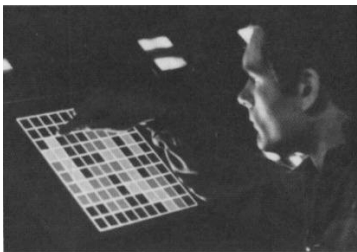
- In the 1967 Stanley Kubrick movie *2001: A Space Odyssey*, the spaceship's computer HAL can
 - display graphics;
 - play chess; and
 - conduct natural, open-domain conversations with humans.
- How well did the filmmakers do at predicting what computers would be captable in 2001?

(Slide idea from Andrew McCallum)

HAL

Graphics

HAL



Jurassic Park (1993)



(Slide idea from Andrew McCallum)

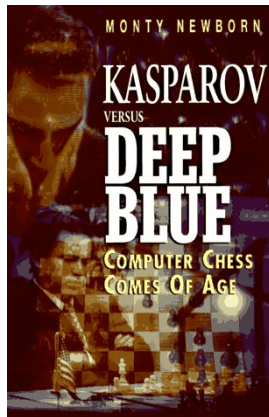
HAL

Chess

HAL



Deep Blue (1997)



(Slide idea from Andrew McCallum)

HAL

Dialogue

HAL

2012

David Bowman: Open the pod bay doors, HAL.

...

HAL: I'm sorry, Dave, I'm afraid I can't do that.

David: What are you talking about, HAL?

HAL: I know that you and Frank were planning to disconnect me, and I'm afraid that's something I cannot allow to happen.

(Slide idea from Andrew McCallum)

Siri



You: Any good burger joints around here?

Siri: I found a number of burger restaurants near you.

You: Hmm. How about tacos?

Apple: [Siri remembers that you asked about restaurants. so it will look for Mexican restaurants in the neighborhood. And Siri is proactive, so it will question you until it finds what you're looking for.]

(Slide from Marie de Marneffe)

Siri

Colbert: For the love of God, the cameras are on, give me something?

Siri: What kind of place are you looking for? Camera stores or churches?

[...]

Colbert: I don't want to search for anything! I want to write the show!

Siri: Searching the Web for "search for anything. I want to write the shuffle."



(Slide from Marie de Marneffe)

Plan and goals

Plan

- 1 The Switchboard Dialog Act Corpus
- 2 The importance of context; practical computational approximations of context.
- 3 The Cards Corpus: a new task-oriented dialogue corpus with rich context.
- 4 Collaborative reference in dialogue.

Goals

- There is no way to cover dialogue in single day/year.
- The closer we got to this class meeting, the more overwhelmed I felt!
- In the end, I decided to focus on a few representative areas where I think it's realistic to expect major scientific gains in the near term.
- For a more comprehensive review, see Jurafsky and Martin 2009:§24 and the references therein.
- My goal is not to show you how to develop full dialogue systems, but rather to highlight some important scientific ideas and to make progress in important sub-parts of that task.

The Switchboard Dialog Act Corpus (SwDA)

- The SwDA extends the Switchboard-1 Telephone Speech Corpus, Release 2, with turn/utterance-level dialog-act tags.
- The tags summarize syntactic, semantic, and pragmatic information about the associated turn.
- It is freely available:

<http://www.stanford.edu/~jurafsky/ws97/>

- The SwDA is not inherently linked to the Penn Treebank 3 parses of Switchboard, and it is far from straightforward to align the two resources (Calhoun et al. 2010).
- In addition, the SwDA is not distributed with the Switchboard's tables of metadata about the conversations and their participants.
- This summer, I created a CSV version of the corpus that pools all of this information to the best of my ability, thereby allowing study of the correlations among dialog tags, conversational metadata, and full syntactic structures:

<http://compprag.christopherpotts.net/swda.html>

DAMSL tags for the Switchboard Dialog Act Corpus

There are over 200 tags in the SwDA, most used only a few times. It is more common to work with a collapsed version involving just 44 tags.

	name	act tag	example	train count	full count
1	Statement-non-opinion	sd	Me, I'm in the legal department.	72824	75145
2	Acknowledge (Backchannel)	b	Uh-huh.	37096	38298
3	Statement-opinion	sv	I think it's great	25197	26428
4	Agree/Accept	aa	That's exactly it.	10820	11133
5	Abandoned or Turn-Exit	%	So, -	10569	15550
6	Appreciation	ba	I can imagine.	4633	4765
7	Yes-No-Question	qy	Do you have to have any special training?	4624	4727
8	Non-verbal	x	[Laughter], [Throat_clearing]	3548	3630
9	Yes answers	ny	Yes.	2934	3034
10	Conventional-closing	fc	Well, it's been nice talking to you.	2486	2582
11	Uninterpretable	%	But, uh, yeah	2158	15550
12	Wh-Question	qw	Well, how old are you?	1911	1979
13	No answers	nn	No.	1340	1377
14	Response Acknowledgement	bk	Oh, okay.	1277	1306
15	Hedge	h	I don't know if I'm making any sense or not.	1182	1226
16	Declarative Yes-No-Question	qy`d	So you can afford to get a house?	1174	1219
17	Other fo_o_fw_by_bc		Well give me a break, you know.	1074	883
18	Backchannel in question form	bh	Is that right?	1019	1053
19	Quotation	`q	You can't be pregnant and have cats	934	983
20	Summarize/reformulate	bf	Oh, you mean you switched schools for the kids.	919	952
21	Affirmative non-yes answers	na	It is.	836	847
22	Action-directive	ad	Why don't you go first	719	746

DAMSL tags for the Switchboard Dialog Act Corpus

There are over 200 tags in the SwDA, most used only a few times. It is more common to work with a collapsed version involving just 44 tags.

	name	act tag	example	train count	full count
23	Collaborative Completion	~2	Who aren't contributing.	699	723
24	Repeat-phrase	b^m	Oh, fajitas	660	688
25	Open-Question	qo	How about you?	632	656
26	Rhetorical-Questions	qh	Who would steal a newspaper?	557	575
27	Hold before answer/agreement	^h	I'm drawing a blank.	540	556
28	Reject	ar	Well, no	338	346
29	Negative non-no answers	ng	Uh, not a whole lot.	292	302
30	Signal-non-understanding	br	Excuse me?	288	298
31	Other answers	no	I don't know	279	286
32	Conventional-opening	fp	How are you?	220	225
33	Or-Clause	qrr	or is it more of a company?	207	209
34	Dispreferred answers	arp.nd	Well, not so much that.	205	207
35	3rd-party-talk	t3	My goodness, Diane, get down from there.	115	117
36	Offers, Options, Commits	oo_co_cc	I'll have to check that out	109	110
37	Self-talk	t1	What's the word I'm looking for	102	103
38	Downplayer	bd	That's all right.	100	103
39	Maybe/Accept-part	aap_am	Something like that	98	105
40	Tag-Question	^g	Right?	93	92
41	Declarative Wh-Question	qw^d	You are what kind of buff?	80	80
42	Apology	fa	I'm sorry.	76	79
43	Thanking	ft	Hey thanks a lot	67	78

Act-tag and syntactic category alignment

A quick experiment: to what extent are dialog act tags and clause-types aligned?

1 Request act

- a. Take these pills twice a day.
- b. You should take these twice a day.
- c. Could you please take these twice a day?

2 Question act

- a. Is today Tuesday?
- b. It's Tuesday, right?
- c. I need to confirm that it's Tuesday.

3 Imperative form

- a. Take these pills twice a day.
- b. Have a seat.
- c. Get well soon.

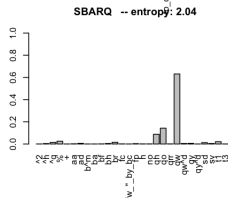
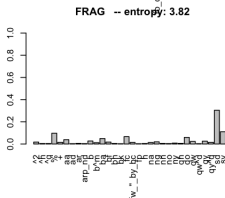
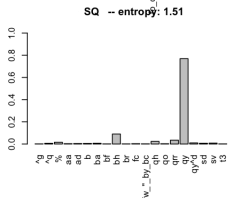
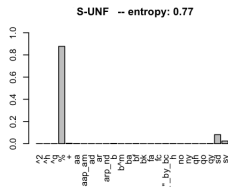
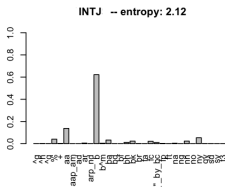
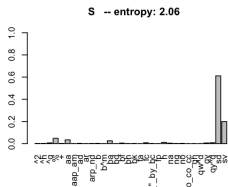
4 Interrogative

- a. Is today Tuesday?
- b. Is he ever tall!
- c. Can you pass the salt?

Act-tag and syntactic category alignment

A quick experiment: to what extent are dialog act tags and clause-types aligned?

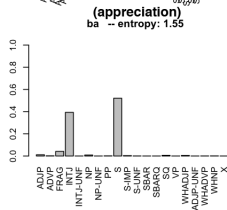
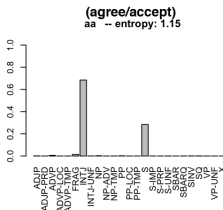
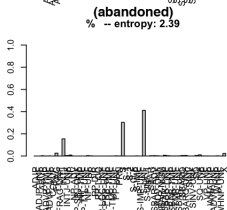
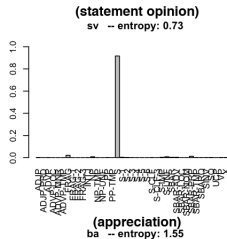
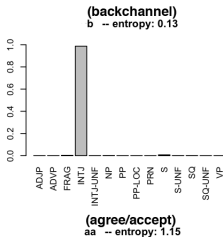
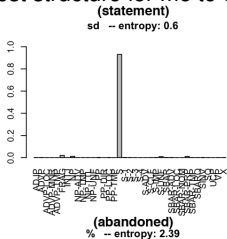
The hearer's perspective: given that I heard a syntactic structure with root label L , what are the speaker's possible intended dialog acts?



Act-tag and syntactic category alignment

A quick experiment: to what extent are dialog act tags and clause-types aligned?

The **speaker's perspective**: given that I want to convey dialog act D , what is the best structure for me to choose?



Modeling act sequences

- Modeling act sequences could be an important step towards realistic interpretation and production.
- Shriberg et al. (1998); Stolcke et al. (2000) uses acoustic features to predict general dialog act labels, using the SwDA. Their model is a decision-tree classifier.
- Other classifiers might also be appropriate; the natural assumption here is that the classifications decisions are made on a by-utterance basis, with no inspection of neighboring utterances (Bangalore et al. 2006; Kumar Rangarajan Sridhar et al. 2009).
- Dialog act prediction can also be viewed as a sequence modeling problem akin to POS tagging, and thus Hidden Markov Models and Conditional Random Fields models are often used. Such models incorporate earlier and/or later tags to make classification decisions.

On the SwDA for dialogue research

Advantages

- Richly annotated.
- Includes speech data.
- Includes sociolinguistic metadata.
- Long conversations, and lots of them.
- Participants did not typically know each other before the conversation, so most of their common ground is general knowledge.

On the SwDA for dialogue research

Advantages

- Richly annotated.
- Includes speech data.
- Includes sociolinguistic metadata.
- Long conversations, and lots of them.
- Participants did not typically know each other before the conversation, so most of their common ground is general knowledge.

Disadvantages

- Open-domain, unfocussed (participants do not stick closely to their topics).
- Virtually no hope of modeling the context or grounding the language in the world or in action.

Context

- Dialogue understanding requires deep contextual understanding.
- Neither you nor your computer can understand a conversation unless you have a lot of background information about who is talking to whom, where the conversation is taking place, why the conversation is taking place, . . .
- There is always uncertainty about the context, so there is no hope of modeling it fully, but we can aim to model aspects of it that are helpful for the tasks at hand.

Indexicality

Natural language is full of expressions that get their meanings directly from the context:

- 1 I'll be there in an hour.
- 2 Did you turn off the stove? (Partee 1973)
- 3 Where are we?
- 4 Where can we find good Chinese food?
- 5 I want to go here (speaker points at a spot on a map).
- 6 a local bar
- 7 a foreign language

I teased Siri above for its template matching, but I bet it has robust and sophisticated methods for interpreting indexicals in terms of information provided by its sensors and measuring devices.

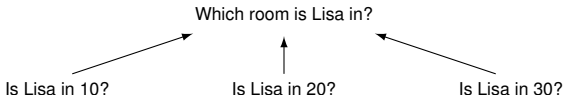
Discourse plans and goals

Context: Homer calls a hotel.

Homer: Is Lisa Simpson in Room 10?

Clerk A: She's in room 20.

Clerk B: #No.



(Perrault and Allen 1980; Groenendijk and Stokhof 1982; Ginzburg 1996; Roberts 1996; Büring 1999; Clark and Parikh 2007)

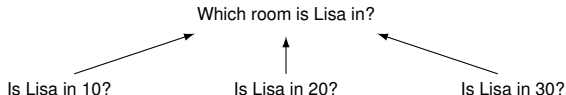
Discourse plans and goals

Context: Homer calls a hotel.

Homer: Is Lisa Simpson in Room 10?

Clerk A: She's in room 20.

Clerk B: #No.



A: Do you know what time it is?

B1: It's 4:35.

B2: No.

B3: #Yes.

A: Can you pass the salt?

A: Can you reach the salt?

A: Would you mind passing the salt?

B3: #Yes.

(Perrault and Allen 1980; Groenendijk and Stokhof 1982; Ginzburg 1996; Roberts 1996; Büring 1999; Clark and Parikh 2007)

Discourse plans and goals

Where are you from?

- *Connecticut.* (Issue: birthplaces)
- *The U.S.* (Issue: nationalities)
- *Stanford.* (Issue: affiliations)
- *Planet earth.* (Issue: intergalactic meetings)

(Perrault and Allen 1980; Groenendijk and Stokhof 1982; Ginzburg 1996; Roberts 1996; Büring 1999; Clark and Parikh 2007)

Discourse plans and goals

Where can we buy supplies?

(Perrault and Allen 1980; Groenendijk and Stokhof 1982; Ginzburg 1996; Roberts 1996; Büring 1999; Clark and Parikh 2007)

Discourse plans and goals

Where can we buy supplies?

Mention-all

- **Context:** We're writing a comprehensive guide to the area.
- **Resolvedness condition:** An exhaustive listing of the (reasonable) shopping places.

(Perrault and Allen 1980; Groenendijk and Stokhof 1982; Ginzburg 1996; Roberts 1996; Büring 1999; Clark and Parikh 2007)

Discourse plans and goals

Where can we buy supplies?

Mention-all

- **Context:** We're writing a comprehensive guide to the area.
- **Resolvedness condition:** An exhaustive listing of the (reasonable) shopping places.

Mention-some

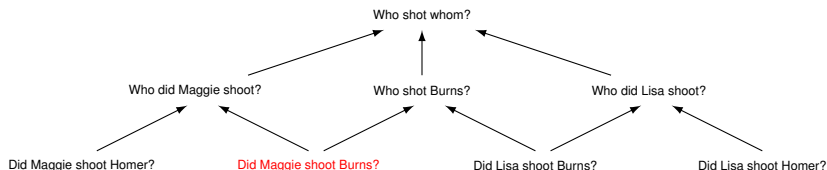
- **Context:** We're low on food and water.
- **Resolvedness condition:** Mentioning the best (closest, safest, etc.) place, or a few good options.

(Perrault and Allen 1980; Groenendijk and Stokhof 1982; Ginzburg 1996; Roberts 1996; Büring 1999; Clark and Parikh 2007)

Discourse plans and goals

Homer: Did Maggie shoot Burns?

Wiggum: She did.

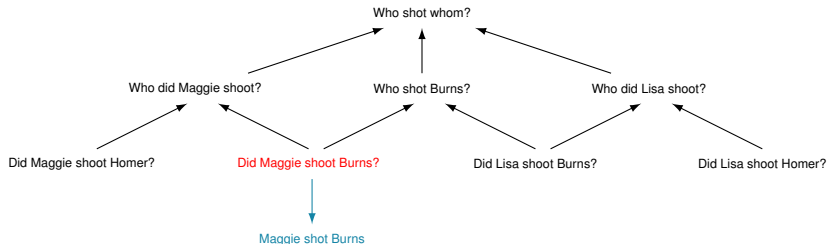


(Perrault and Allen 1980; Groenendijk and Stokhof 1982; Ginzburg 1996; Roberts 1996; Büring 1999; Clark and Parikh 2007)

Discourse plans and goals

Homer: Did Maggie shoot Burns?

Wiggum: She did. **Maggie shot Burns**

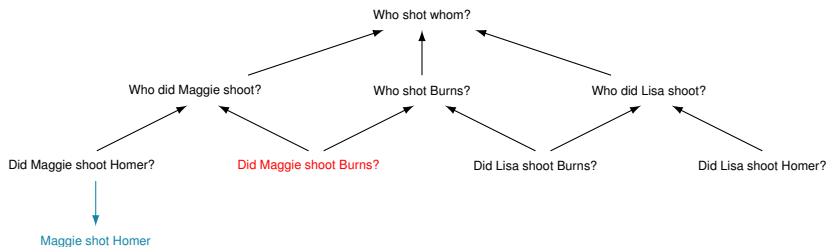


(Perrault and Allen 1980; Groenendijk and Stokhof 1982; Ginzburg 1996; Roberts 1996; Büring 1999; Clark and Parikh 2007)

Discourse plans and goals

Homer: Did Maggie shoot Burns?

Wiggum: She did. **Maggie shot Homer**

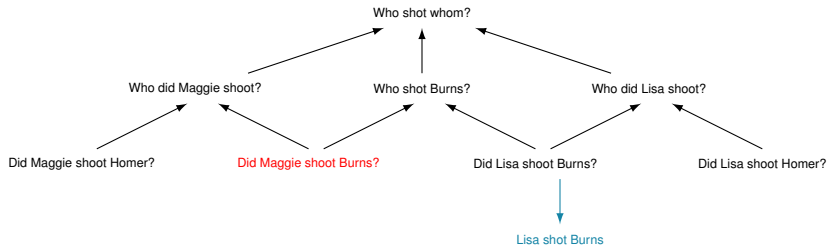


(Perrault and Allen 1980; Groenendijk and Stokhof 1982; Ginzburg 1996; Roberts 1996; Büring 1999; Clark and Parikh 2007)

Discourse plans and goals

Homer: Did Maggie shoot Burns?

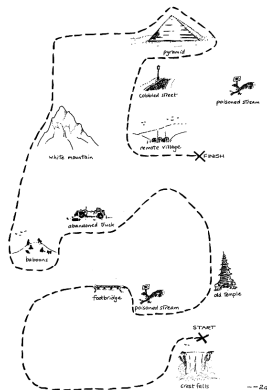
Wiggum: She did. **Lisa shot Burns**



(Perrault and Allen 1980; Groenendijk and Stokhof 1982; Ginzburg 1996; Roberts 1996; Büring 1999; Clark and Parikh 2007)

The Edinburgh Map Corpus (Thompson et al. 1993)

One participant tells the other how to reproduce a path through a map; the context and discourse goals are fully specified.



- g right it starts directly above the crest falls if you go to the left of your page just to the edge of the crest falls
- f mmhmm
- g come south due south to the bottom of the page
- f mmhmm
- g go to the left of the page to about an inch from the end
- f over the banana tree
- g i suppose so yeah eh
- f mmhmm
- g go north to the level of the footbridge
- f mmhmm
- g go up and go across the footbridge and stop exactl– right at the end edge of the footbridge
- f above the footbridge
- g o– over the footbridge
- f mm
- g and stop right at the end of it
- g there is a poisoned stream on mine but which you don't have
- f right
- g okay
- g and so what you do is you come south immediately south at the edge of the footbridge oh for about for about an inch so that you're about an inch above the starting cross not actually above it you know but eh along and above it

For transcripts, audio, maps, etc.: <http://groups.inf.ed.ac.uk/maptask/>

The TRIPS and TRAINS corpora

<http://www.cs.rochester.edu/research/cisd/projects/trips/>

- TRIPS and TRAINS offer text and audio of task-oriented dialogues in rich contexts.
- TRAINS is similar to the Map Task, except that the task is to define efficient routes for trains to travel between cities.
- TRIPS also involves path-planning, except with additional challenges — different kinds of cargo and vehicles, more real-world constraints.
- These corpora are predecessors for PLOW (Allen et al. 2007), which uses TRIPS technology and which also seeks to create agents that collaboratively assist with planning and scheduling.



Example TRAINS domain



Example TRIPS-911 domain

PLOW: webpage structure as context

For the PLOW system, the context is the webpage:

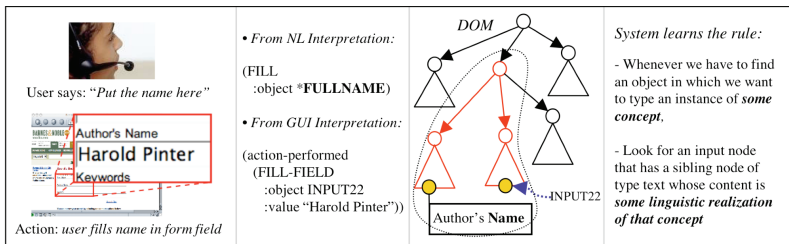


Figure 4: Learning to find and fill a text field

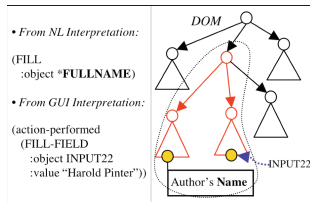
- Project homepage:
<http://www.cs.rochester.edu/research/cisd/projects/plow/>
- Language processing with the TRIPS parser:
<http://www.cs.rochester.edu/research/cisd/projects/trips/parser/cgi/web-parser-xml.cgi>

Learning new rules and generalizations

- Learning rules of the form ‘If A, then B, else C’ is a challenge because the latent variable A is generally not observed. Rather, one sees only B or C.
- In an interactive, instructional setting, one needn’t rely entirely on abduction or probabilistic inference: users generally state the needed rules during their interactions.

Language-based principles

- 1 The user's actions ground the parsed language.



- 2 The DOM structure grounds the user's indexicals and other referential devices.
 - Put the name here. (user clicks the mouse on the DOM element)
 - This is the ISBN number. (user highlights some text)
 - Find another tab. (user has selected a tab)
- 3 Indefinites mark new information; definites refer to established information:
 - *A man walked in. He/The man looked tired.*
 - *an address* ⇒ new input parameter
 - *the address* ⇒ existing input parameter

Interaction and error correction

- PLOW is tested with human users in real scenarios. (It has been used by the US Military Health System to set up doctor's appointments.)
- Thus, PLOW tries to immediately apply the rules it infers, so that the user will correct it. This helps with:
 - finding the right level of generalization; and
 - overcoming noise in the context (from poor HTML mark-up)

Evaluation

16 independent evaluators trained to work with PLOW and three other systems.

Phase 1

- 1 The evaluators used the systems to teach some of the predefined tasks.
- 2 The system then performed those tasks with different input parameters.

Phase 2

- 1 The evaluators used the systems to teach some of the tasks at right.
- 2 PLOW received the highest average score of all systems.
- 3 Evaluators had free choice of which system to use. 13 chose PLOW for at least one task, and PLOW was used for 30 of the 55 tasks constructed during evaluation.

1. What <businesses> are within <distance> of <address>?
2. Get directions for <integer> number of restaurants within <distance> of <address>.
3. Find articles related to <topic> written for <project>.
4. Which <project> had the greatest travel expenses between <start date> and <end date>?
5. What is the most expensive purchase approved between <start date> and <end date>?
6. For what reason did <person> travel for <project> between <start date> and <end date>?
7. Find <ground-transport, parking> information for <airport>.
8. Who should have been notified that <person> was out of the office between <start date> and <end date>?
9. Summarize all travel and purchase costs for <project> between <date> and <date> by expense category
10. Which projects exceeded the current government maximum allowable expense for travel costs?

Figure 1: Previously unseen tasks used in the evaluation

The Cards Corpus

<http://CardsCorpus.christopherpotts.net/>

Included:

- The transcripts in CSV format
- Python classes for working with the transcripts
- Examples of the Python classes in action
- R code for reading in the corpus as a data frame

Also, a search function with HTML/Javascript game visualizations:

<http://CardsCorpus.christopherpotts.net/search/>

Amazon Mechanical Turk HIT (Human Intelligence Task)

- **Title:** Collaborative Search Game with Chat
- **Description:** Two-player collaborative video game involving dialogue/chat with other Turkers.
- **Payment:** \$1.00, and up to \$0.50 cents for rich, collaborative problem-solving using meaningful dialogue.
- **Restrictions:** US IP addresses; at least 95% approval rating

HIT set-up

Instructions

1. You'll first be presented with an experimental consent form; the transcripts from these games will be used to study collaborative problem solving.
2. Once you click through, you'll go into "Wait for Participant" mode until another player shows up. **It might take a while for another player to show up! We suggest opening a new window and doing other things until someone arrives**
3. When play begins, you'll be presented with a task description: you and your partner are tasked with finding six consecutive cards of the same suit.
4. When you've solved the game, click the "Task Complete" button to register that the task is complete. You can continue playing and chatting at that point, or you can click **Complete HIT and return to MTurk** at the top of the screen.
5. You will receive a bonus of up to \$0.50 for good collaboration with meaningful communication.
6. Notes:
 - Once the game begins, you can move around with the arrow keys or the on-screen buttons.
 - The chat window is near the top of the screen.
 - You can't see your partner, and your partner can't see you!
 - The yellow boxes mark cards in your "line of sight".
7. **Caution: Players who do not complete the task properly will not be paid. It is vital that you read and understand the task description and follow the instructions.**
8. Feel free to play multiple times — especially useful if you get good at the game.

Annotated game screen

HIT set-up

The screenshot shows a HIT interface for a card game. At the top, there are navigation tabs: Overview, The Switchboard Dialog Act Corpus, Context, The Cards Corpus (selected), Collaborative reference, and Conclusion. Below the tabs is a header area with the text 'TYPE HERE' and a 'Disable Sound' button. The main interface is divided into several sections:

- Chat Log:** Shows messages: 'Received: hi', 'Sent: I have the JH', and 'Received: I have the 8H'.
- Task Description:** A scrollable text box containing the instruction: 'Gather six consecutive cards of a particular suit (decide which suit together). Each of you can hold only three cards at a time, so you'll have to coordinate your efforts. You can talk all you'.
- Text Input:** A text box with the placeholder 'Type text here:' and a 'Disable Sound' button. The current text is: 'I'm on 2D, which isn't too useful. There are cards to my right and below, though. I'll check them out.'
- Game Board:** A maze-like grid with a player icon 'P1' and three yellow boxes representing cards in the player's line of sight.
- Game State:** Displays 'P1 turns remaining: 546' and 'P2 turns remaining: 599', along with an 'Indicate Task Complete' button.
- Card Selection:** A central area with 'up', 'down', 'left', and 'right' directional buttons. It contains two cards: '2D' (picked up) and 'JH' (to be dropped).

Annotations with red arrows point to various elements:

- 'TYPE HERE' points to the text input field.
- 'Yellow boxes mark cards in your line of sight.' points to the yellow boxes on the game board.
- 'You are on 2D' points to the '2D' card in the selection area.
- 'Task description: Six consecutive cards of the same suit' points to the task description text.
- 'The cards you are holding' points to the 'JH' card in the selection area.
- 'Move with the arrow keys or these buttons.' points to the directional buttons.

HIT set-up

Title of project: Pragmatic enrichment and contextual inference

Principal investigator: Dr. Christopher Potts, Department of Linguistics, Stanford University

Purpose of Study

In this research, we are investigating how people communicate when working together to resolve a joint goal. We collect transcripts of dialogues of players playing a simple game that requires participants to strategize together to solve tasks.

Procedures

In this study, you will play a game with another unseen player over a network. You will see instructions on the screen, and you may communicate with the other player by keyboard. The game consists of a map and several hidden objects. You have a limited number of moves to collect the objects according to the instructions you receive. You must coordinate with your teammate to achieve the goal. The time of a single game is variable depending on the players. A typical game lasts between 10 and 20 minutes.

Risks

There are no known risks involved in this experimental procedure.

Benefits

HIT set-up

*Gather six consecutive cards of a particular suit (decide which suit together), or **determine that this is impossible**. Each of you can hold only three cards at a time, so you'll have to coordinate your efforts. You can talk all you want, but you can make only a limited number of moves.*

Transcripts: game play

Agent	Time	Action type	Contents
Player 1	566650	PLAYER_MOVE	7,11
Player 2	567771	CHAT_MESSAGE_PREFIX	which c's do you have again?
Player 1	576500	CHAT_MESSAGE_PREFIX	i have a 5c and an 8c
Player 2	577907	CHAT_MESSAGE_PREFIX	i jsut found a 4 of clubs
Player 1	581474	PLAYER_PICKUP_CARD	7,11:8C
Player 1	586098	PLAYER_MOVE	7,10


Turker Nation discussion

Initial post announcing the new HITs, our goals for them, and warning about wait-times and potential bugs.

Click here to go to the NEW Turker Nation forum :: Requesters Boardroom :: Everyone Else :: Collaborative Search Game with Chat







Page 1 of 11 » Jump to page Go

[remove bookmark](#) [share](#) [print](#)

Author	Topic: Collaborative Search Game with Chat (Read 1,444 times)
<p>stanfordpraglab Grizzled Turker ☆☆☆☆ member is online</p> <p style="text-align: center;"></p> <p>Joined: Jun 2010 Posts: 126</p>	<p>Collaborative Search Game with Chat « Thread Started on Aug 10, 2011, 10:21am » modify delete</p> <hr/> <p>Turker Nation!</p> <p>I'm writing to let you all know that we've some posted some more HITs called Collaborative Search Game with Chat</p> <p>In the game, you navigate around in a maze-world. You and your partner try to gather six consecutive cards of the same suit before you run out of moves. You and your partner decide which suit and which sequence.</p> <p>You can see where cards are (yellow boxes) when they enter your "line of sight". Navigation is with the keyboard or with the buttons on the right of the screen. When you're on a card, its name displays on the right and you can pick it up by clicking on it.</p> <p>You can hold only three cards at once. so you have to coordinate on your strateav.</p>




Turker Nation discussion

Responding to player queries while the HITs were running.

<p>stanfordpraglab Grizzled Turker ★★★★★ member is online</p> <p></p> <p>Joined: Jun 2010 Posts: 126</p>	<p> Re: Collaborative Search Game with Chat « Reply #9 on Aug 10, 2011, 11:04am »</p> <p> modify  delete</p> <hr/> <p>Aug 10, 2011, 10:50am, moniquitta wrote:</p> <div data-bbox="377 422 1163 471" style="border: 1px solid black; padding: 5px;"><p>Would this be weird but can we like state TN here in the beginning so we know we are playing someone from here?</p></div> <p>This is totally fine by us --- even adds a new community dimension to the data. ---Chris</p> <p>Report to Mod - Link to Post - Back to Top  Logged</p>
<p>moniquitta Turkaholic ★★★★★ member is offline</p> <p></p> <p>Joined: Feb 2011 Gender: Female ♀ Posts: 722</p>	<p> Re: Collaborative Search Game with Chat « Reply #10 on Aug 10, 2011, 11:05am »</p> <hr/> <p>Sweet awesome. OH and good to see you on here working with us.</p> <p>Report to Mod - Link to Post - Back to Top  Logged</p>




Turker Nation discussion

Responding to player queries while the HITs were running.

<p>stanfordpraglab Grizzled Turker ★★★★★ member is online</p> <p></p> <p>Joined: Jun 2010 Posts: 126</p>	<p>Re: Collaborative Search Game with Chat « Reply #80 on Aug 11, 2011, 10:55am »</p> <p> modify  delete</p> <hr/> <p>Hi Turkers,</p> <p>You will absolutely get approved if you make a genuine attempt to solve the game and run out of moves. Hope this version isn't too hard.</p> <p>Stanford Praglab</p> <p>Report to Mod - Link to Post - Back to Top  Logged</p>
---	---




Turker Nation discussion

The players helped us spot trouble.

<p>timefactor Turker ☆☆☆ member is offline</p> <p></p> <p>Joined: Jun 2011 Gender: Male ♂ Posts: 13</p>	<p> Re: Collaborative Search Game with Chat « Reply #152 on Aug 11, 2011, 4:28pm »</p> <hr/> <p>Aug 11, 2011, 4:24pm, sleepstar wrote:</p> <div data-bbox="370 419 1166 450" style="border: 1px solid black; padding: 2px;"><p>wow.. high paying, fast HITs.. go figure the stupid scammers would show up. ughhhhhhhh.</p></div> <p>Seriously. This game is not that hard and wouldn't even take much longer than just blindly submitting. This batch was disappointing since three of my games were these losers who didn't play and I was too scared to submit them.</p> <p style="text-align: right;">Report to Mod - Link to Post - Back to Top  Logged</p>
---	---

Turker Nation discussion

And resolve trouble.

<p>bid Turkaholic ★★★★★ member is offline</p> <p></p> <p>Joined: Jan 2011 Gender: Male  Posts: 769</p>	<p>Re: Collaborative Search Game with Chat « Reply #158 on Aug 11, 2011, 4:45pm »</p> <hr/> <p>Aug 11, 2011, 4:35pm, moniquitta wrote:</p> <div style="border: 1px solid black; padding: 5px;"> <p>I wonder if a qual would help.</p> </div> <p>I wondered the same thing. I suppose it depends on how many of these there will be. Of course, with a qual, they could weed out some, shall we say less than dedicated participants, increase the quality of data being obtained and (can't believe I am saying this) make it increasingly challenging to succeed.</p> <p style="text-align: right;"> Report to Mod - Link to Post - Back to Top  Logged </p>
---	--

Turker Nation discussion

That was actually a pretty fun hit.

The game with chat was great and like to see more HITs from you.

These HITs were really enjoyable. Hopefully you will put more on the site. You state that we can keep doing them, but right now if I click on your HIT, it tells me there are no more available for me. Is there something I can do to try again? Thanks.

I waited 1.22 before someone showed up. They never talked to me and didn't finish the job before leaving. Am I still out because they didn't cooperate?

By the numbers

- 1,266 transcripts
- Game length mean: 373.21 actions (median 305, sd 215.20)
- Actions:
 - Card pickup: 19,157
 - Card drop: 12,325
 - Move: 371,811
 - Utterance: 45,805
 - Utterance length mean: 5.69 words (median 5, sd 4.74)
 - Total word count: 260,788
 - Total vocabulary: 5,577 (\approx 4,000 if card references are normalized)

Constants and points of variation

Constants

- Task description ('six consecutive cards of the same suit')
- Max cards in hand: 3

Randomness

- Players' initial positions
- All card positions

Systematic variation (highlights)

- Some game are infeasible because areas of the board are walled off.
- Most games are symmetric: the players each have the same line of sight and number of moves.
- Around 500 games are asymmetric: one player has a very limited number of moves but infinite line-of-sight; the other has a large number of moves but very limited line of sight.
- There are a few different games boards, in a few different sizes.
- The number of moves each player has varies from 100 to 600; these values result in very different play.

Sample runs

P1 start P1 finish P2 finish
 Available cards:
P2 start P2 finish P1 finish

Mouse-over the utterances to see where they were made.

- Player 1: hello
- Player 2: hi
- Player 2: top right
- Player 1: i am...right in the very middle in the large room
- Player 1: sitting on a dD
- Player 2: ac
- Player 2: I go top
- Player 1: 10C nearby
- Player 2: want to try c
- Player 1: OK
- Player 1: I'll try to work around to the bottom and right
- Player 2: k
- Player 1: grabbed a KC
- Player 2: k
- Player 1: got the Queen
- Player 1: so we have 10-QKA
- Player 1: need the 9-J
- Player 2: k
- Player 1: looks like you're using up your moves
- Player 1: where are you going
- Player 2: bottom left
- Player 2: look so top and side
- Player 1: ah
- Player 1: my hand is full so you probably want to stop moving
- Player 1: until I find what were looking for
- Player 2: k
- Player 1: ok found the J
- Player 2: where
- Player 1: where are you standing
- Player 1: all the way on the bottom right corner
- Player 1: one up from the bottom
- Player 2: you see the opening on the bottom
- Player 1: yeah
- Player 2: im three over one u to the right
- Player 2: I mean left
- Player 1: ok i dont think 16 moves will make it here
- Player 2: right under the u
- Player 1: the U?
- Player 2: you see the u on the bottom right
- Player 2: near the opening on the bottom
- Player 1: i'm not sure what the U is
- Player 1: are you at the break in the bottom wall
- Player 2: 2 left one u from it
- Player 1: ok
- Player 1: yeah i dont think you can make it here from there
- Player 1: if the 9 is near me maybe i can bring it halfway
- Player 2: can you make it to me and back
- Player 1: not sure yet
- Player 1: yeah.
- Player 1: depends on where the 9 is
- Player 1: make sure to leave a turn left to pick up the card
- Player 2: ok found the 9
- Player 2: im at the break in the bottom
- Player 1: oh man
- Player 1: yeah too far away
- Player 2: where are you
- Player 1: what do you have in yr hand
- Player 2: ac
- Player 1: ok i dropped the King Q to the right of the break in sub bottom wall
- Player 1: I'll see if i can make it back to the other cards
- Player 2: k
- Player 1: yeah, couldn't make it... dam
- Player 1: can you pick up the cards i dropped
- Player 1: we can still finish
- Player 2: y
- Player 1: when you run out of moves

Sample runs

Novices

Player 1: Hello. Are you here?
Player 2: yes
Player 2: do you see any cards
Player 1: Yes. I see a yellow spot. Those are our cards. We'll only be able to see the ones that are in our view
Player 1: until we move with our arrows.
Player 2: i see 3 of them
Player 1: We only have a certain number of moves, so we should decide how we're going to do this before we use them, do you think?
Player 2: sure
Player 1: Ok. So, we have to pick up six cards of the same suit, in a row...
Player 1: each of us can hold three, so...
Player 1: I think I should get my three, then you should get your three or vice versa
Player 2: ok
Player 2: you go ahead
Player 1: What suit should we do?
Player 1: And which six cards do you want to try for?
Player 2: whatever you want
Player 1: I'm __REDACTED_NAME__, by the way- nice to meet you.
Player 2: i'm __REDACTED_NAME__...nice to meet you too
Player 1: Hi __REDACTED_NAME__. How about we go for hearts? And take 234567
[...]

Sample runs

Experienced players

These players have explored and are now forming a strategy:

Player 1 I have 9 clubs and K clubs

Player 1 want to look for clubs?

Player 2 ok

[...]

The players then find various clubs, checking with each other frequently, until they gain an implicit understanding of which specific sequences to try for (either 8C-KC or 9C-AC):

Player 1 so you are holding Jc and Kc now?

Player 2 i now have 10d JC and KC

Player 2 yes

Player 1 drop 10d and look for either 8c or Ace of clubs

Sample runs

Experts

Player 2: hi

Player 1: hi--which side r u on?

Player 2: right side

Player 2: u?

Player 1: left/middle

Player 1: ok i gathered everything in my area

Player 2: i think i have all of them also

Player 1: how bout 5C - 10C?

Player 2: ok

Player 1: i have 5C, 8C, 9C, and you should have 6C, 7C, 10C

Player 2: got them

Sample runs

Asymmetric play: one player has a very limited number of moves but infinite line-of-sight; the other has a large number of moves but very limited line of sight.

```

Player 1: Hi
Player 2: hi where are you
Player 1: near the upper right
Player 2: ok any cards that way
Player 1: lots of cards near me to the upper right corner
Player 2: did you get that
Player 1: get wjat ?
Player 2: the drop in the top right
Player 1: I have not gone there yet
Player 2: ok I'll wait
Player 2: we have the 4 8 j h
Player 2: 3 k c
Player 1: ok
Player 1: the cards are pretty scattered
Player 1: did you check the entire right column?
Player 2: doing it now
Player 2: what hearts do we have
Player 1: 4 8 J
Player 1: uh.. need any help ?
Player 2: are there any on the bottom
Player 2: I have the 6 7 hearts
Player 1: yes the bottom row has 6 cards
Player 1: 4 8 J K Q A H
:
:

```

Player annotations

- 879 of the transcripts include player Ids.
- 1183 distinct players

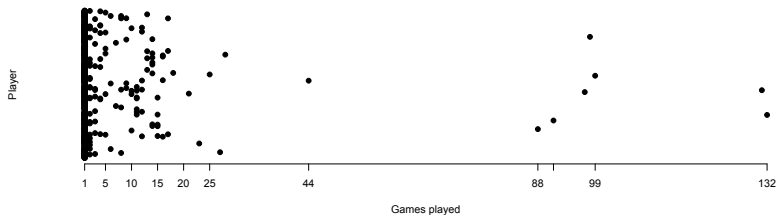


Figure: Number of games played by each player. The data points have been jittered randomly along the y-axis to make their clustering evident.

(Djalali et al. 2011: expertise affects strategy and language)

Language in context

Each transcript is a data structure that is intuitively a list of temporally-ordered states

$$(\text{context}, \text{event})$$

The context includes

- local information (the state of play at that point)
- historical information (the events up to that point)
- global information (limitations of the game, the task, etc.)

When the event is an utterance, we can interpret it *in context*.

This is what pragmatics is all about, but it is very rare to have a dataset that truly lets you do it.

Other task-oriented dialogue corpora

Corpus	Task type	Domain	Task-orientation	Dialogues	Format
Switchboard	discussion	wide-open	very loose	2,400	aud/txt
SCARE	search	3d world	tight	15	aud/vid/txt
Cards	search	2d grid	tight	745	txt
TRAINS	routes	map	tight	120	aud/txt
Map Task	routes	map	tight	128	aud/vid/txt

(See also Blaylock and Allen 2005)

Other task-oriented dialogue corpora

Corpus	Task type	Domain	Task-orientation	Dialogues	Format
Switchboard	discussion	wide-open	very loose	2,400	aud/txt
SCARE	search	3d world	tight	15	aud/vid/txt
Cards	search	2d grid	tight	745	txt
TRAINS	routes	map	tight	120	aud/txt
Map Task	routes	map	tight	128	aud/vid/txt

(See also Blaylock and Allen 2005)

Chief selling points for Cards:

- Pretty large.
- Controlled enough that similar things happen often.
- Very highly structured — the only corpus whose release version allows the user to replay all games with perfect fidelity.

Collaborative reference

- Clark (1996) argues that language use is collaborative: speakers and hearers always work together to create meaning and guide the discourse.
- This section first reports on a seminal set of experiments by Clark and Wilkes-Gibbs (1986) highlighting the collaborative nature of referring.
- I then review the COREF system of DeVault (2008), DeVault et al. (2005), and DeVault and Stone (2007, 2009), which the creators have evaluated using human-subjects methods and machine learning methods.
- I then sketch two ways in which these insights and experiments can be adapted to work with the Cards Corpus.

Collaboration (Clark and Wilkes-Gibbs 1986)

The 'literary' model of reference in discourse

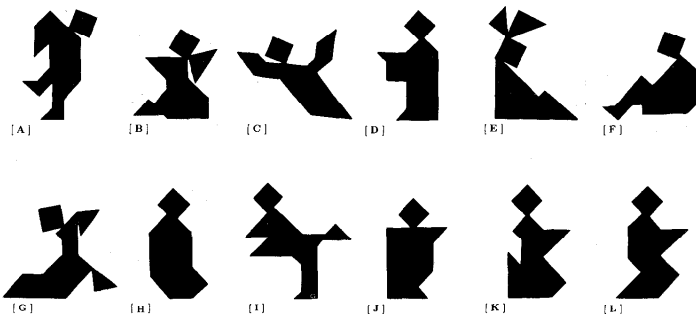
- The speaker is in complete control.
- All noun phrases are well-formed and contextually unambiguous.
- Uttering makes it so — the addressee perceives the speaker's intentions perfectly simply in virtue of the speaker's linguistic choices.

The collaborative model of reference in discourse

- Speakers' referring expressions often involve self-correction and impromptu addition of extra descriptive content.
- The addressee often steps in to make corrections.
- The speaker often invites addressee-participation via intonational cues, dummy phrases like *what's his name*, and physical gestures inviting the addressee to fill in descriptive content.

The tangrams experiment

Figure 1. *The 12 Tangram figures arranged by directors and matchers.*



- 1 Paired participants draw lots for the director *D* and matcher *M* roles.
- 2 *D* gets an arrangement of the above twelve figures in a particular order.
- 3 *M* gets the figures in a random order.
- 4 *D* is tasked with getting *M* to rearrange her figures so that they are in the same order as *D*'s.
- 5 Each pair does six trials.

Examples

Matcher acknowledgement

A: Number 4's the guy leaning against the tree.

B: Okay.

Installment noun phrase with Matcher acknowledgement

A: And the next one is the one with the triangle to the right . . .

B: Okay.

A: With the square connected to it.

Examples

The Matcher fills in the Director's dummy descriptor

A: And number 12 is, uh, ...

B: Chair.

A: With the chair, right.

B: Got it.

The Matcher signals lack of certainty

A: Okay, the next one is the rabbit.

B: Uh—

A: That's asleep, you know, it looks like it's got ears and a head pointing down.

B: Okay.

* * * * *

A: Uh, person putting a shoe on.

B: Putting a shoe on?

A: Uh huh. Facing left. Looks like he's sitting down.

B: Okay.

Findings

For a broad picture of what occurred, consider this very simple series of utterances by one director for figure I on trials 1 through 6:

1. All right, the next one looks like a person who's ice skating, except they're sticking two arms out in front.
2. Um, the next one's the person ice skating that has two arms?
3. The fourth one is the person ice skating, with two arms.
4. The next one's the ice skater.
5. The fourth one's the ice skater.
6. The ice skater.

- ① For the group as a whole, directors' referential phrases were longer in earlier trials than in later ones (mean of 41 words in trial 1; mean of 8 in trial 6).
- ② For the group as a whole, directors took more turns in earlier trials than later ones (mean of 3.7 turns/figure on trial 1; mean of 1 turn per figure on trial 6).
- ③ For the group as a whole, the pairs became more efficient during each trial as well, as the number of unplaced pictures declined. In trial 1, the efficiency gains were greatest, in trial 6 they were the smallest. By that time, they had settled on preferred referential phrases.
- ④ For the group as a whole, referential phrases were more standard in later trials than earlier ones.
- ⑤ Basic exchanges, consisting of a non-collaborative presentation and a simple asserted or presupposed acceptance, were far more frequent in early trials than in later ones.

General principles

Minimizing collaborative effort

“Our proposal is that speakers and addressees try to minimize *collaborative effort*, the work both speakers and addressees do from the initiation of the reference process to its completion” (p. 26).

Principle of mutual responsibility

“The participants in a conversation try to establish, roughly by the initiation of each new contribution, the mutual belief that the listeners have understood what the speaker meant in the last utterance to a criterion sufficient for current purposes” (p. 33)

The COREF system of David DeVault and Matthew Stone

A task-oriented dialogue system

- COREF and its human interlocutor collaborate on a simple referential task, improving forms and resolving ambiguities using contextual and linguistic information.
- The experimental task is the same as Clark and Wilkes-Gibbs's (though with simpler and fewer shapes), but now one of the agents can be artificial.

possible contexts	agent	actor
c_1	COREF:	is the target pink?
c_2	A18:	no
c_3	A18:	it is brown
c_{4l}, c_{4d}	COREF:	do you mean dark brown?
c_{5l}, c_{5d}	A18:	yes
c_6	COREF:	(privately adds the brown diamond)
c_6	COREF:	done

Figure 2: COREF asks a clarification question.

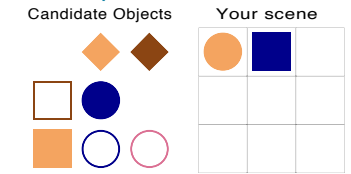
(DeVault and Stone 2007)

Figure 1: A human user plays an object identification game with COREF. The figure shows the perspective of the user (denoted c_4). The user is playing the role of director, and trying to identify the diamond at upper right (indicated to the user by the blue arrow) to COREF.

Figure 2: The conversation of Figure 1 from COREF's perspective. COREF is playing the role of matcher, and trying to determine which object the user wants COREF to identify.

(DeVault and Stone 2009)

COREF's interpretations



- ② COREF: is the target round?
- ③ c4: brown diamond
- ④ COREF: do you mean dark brown?
- ⑤ c4: yes

COREF's possible interpretation for ③

```

i2,1 = { c4:tacitAbandonTasks [2],
          c4:addcr [t7, rhombus (t7) ],
          c4:setPrag [inFocus (t7) ],
          c4:addcr [t7, saddlebrown (t7) ]}

i2,2 = { c4:tacitAbandonTasks [2],
          c4:addcr [t7, rhombus (t7) ],
          c4:setPrag [inFocus (t7) ],
          c4:addcr [t7, sandybrown (t7) ]}
  
```

Managing the ambiguities via probability

- COREF maintains a probability distribution over up to three potential dialogue states.
- The nature of this distribution should be given by numerous contextual and decision-making factors.
- In the implemented system, a simple measure is used, which assigns higher probability to interpretations that require fewer tacit actions.

COREF's utterances

For ambiguity resolution, COREF specializes in three areas

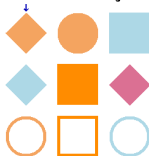
- 1 word-sense ambiguities (e.g., *brown* in the previous example)
- 2 referential ambiguities
- 3 speech-acts (e.g., distinguishing statement from request)

COREF's strategies for eliminating possible interpretations

- Querying the user for additional information (as in the previous example).
- Pooling information from multiple utterances:

- 1 c1: color is orange
- 2 c1: shape is diamond
- 3 COREF: ok

Candidate Objects



Experimental results (DeVault and Stone 2007)

- 20 human subjects participated over the Net, collaborating with COREF.
- Each subject saw 29 objects in three groups:
 - Round 1: 4 objects, 2 × 2 grid
 - Round 2: 9 objects, 3 × 3 grid
 - Round 3: 16 objects, 3 × 3 grid
- COREF or the subject randomly chosen to be Director for Round 1; subsequently, they alternated.

correct	no object	skipped	wrong
75.0%	14.3%	7.4%	3.3%

Table 1: Overall distribution of object outcomes.

1 context	2 contexts	3 contexts
83.4%	6.8%	9.8%

Table 2: Number of possible contexts perceived when utterances or actions occur.

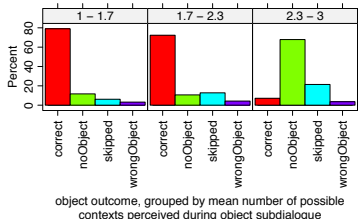


Figure 3: Object outcome vs. context uncertainty.

Learning interpretive preferences

The previous experimental data provides labels that can be used to fit probabilistic models that capture dialogue preferences (DeVault 2008:§4.4).

Learning interpretive preferences

The previous experimental data provides labels that can be used to fit probabilistic models that capture dialogue preferences (DeVault 2008:§4.4).

```

i1,1 = { COREF : tacitNop[[c4 does clickContinue[]]],
          c4 : pushCollabRef[c4, COREF, t7],
          COREF : pushYNQ[COREF, c4, addcr[t7, P], negcr[t7, P]],
          COREF : askYNQ[addcr[t7, circleFigureObject(t7)]],
          COREF : setPrag[inFocus(Y), inFocus(t7)]}

```

- ② COREF: is the target round?
-

Learning interpretive preferences

The previous experimental data provides labels that can be used to fit probabilistic models that capture dialogue preferences (DeVault 2008:§4.4).

2 COREF: is the target round?

3 c4: brown diamond

```

i1,1 = { COREF : tacitNop[[c4 does clickContinue[]]],
          c4 : pushCollabRef[c4, COREF, t7],
          COREF : pushYNQ[COREF, c4, addcr[t7, P], negcr[t7, P]],
          COREF : askYNQ[addcr[t7, circleFigureObject(t7)]],
          COREF : setPrag[inFocus(Y), inFocus(t7)]}

```

```

i2,1 = { c4 : tacitAbandonTasks[2],
          c4 : addcr[t7, rhombusFigureObject(t7)],
          c4 : setPrag[inFocus(Y), inFocus(t7)],
          c4 : addcr[t7, saddlebrownFigureObject(t7)] }

```

```

,
i2,2 = { c4 : tacitAbandonTasks[2],
          c4 : addcr[t7, rhombusFigureObject(t7)],
          c4 : setPrag[inFocus(Y), inFocus(t7)],
          c4 : addcr[t7, sandybrownFigureObject(t7)] }

```

Learning interpretive preferences

The previous experimental data provides labels that can be used to fit probabilistic models that capture dialogue preferences (DeVault 2008:§4.4).

- 2 COREF: is the target round?
- 3 c4: brown diamond
- 4 COREF: do you mean dark brown?

```

i1,1 = { COREF : tacitNop[[c4 does clickContinue[]]],
         c4 : pushCollabRef[c4, COREF, t7],
         COREF : pushYNQ[COREF, c4, addcr[t7, P], negcr[t7, P]],
         COREF : askYNQ[addcr[t7, circleFigureObject(t7)]],
         COREF : setPrag[inFocus(Y), inFocus(t7)]}
  
```

```

i2,1 = { c4 : tacitAbandonTasks[2],
         c4 : addcr[t7, rhombusFigureObject(t7)],
         c4 : setPrag[inFocus(Y), inFocus(t7)],
         c4 : addcr[t7, saddlebrownFigureObject(t7)]
        ,
        i2,2 = { c4 : tacitAbandonTasks[2],
               c4 : addcr[t7, rhombusFigureObject(t7)],
               c4 : setPrag[inFocus(Y), inFocus(t7)],
               c4 : addcr[t7, sandybrownFigureObject(t7)]}
        }
  
```

Learning interpretive preferences

The previous experimental data provides labels that can be used to fit probabilistic models that capture dialogue preferences (DeVault 2008:§4.4).

- 2 COREF: is the target round?
- 3 c4: brown diamond
- 4 COREF: do you mean dark brown?
- 5 c4: **yes**

```

i1,1 = { COREF : tacitNop[[c4 does clickContinue[]]],
          c4 : pushCollabRef[c4, COREF, t7],
          COREF : pushYNQ[COREF, c4, addcr[t7, P], negcr[t7, P]],
          COREF : askYNQ[addcr[t7, circleFigureObject(t7)]],
          COREF : setPrag[inFocus(Y), inFocus(t7)]}

```

```

i2,1 = { c4 : tacitAbandonTasks[2],
          c4 : addcr[t7, rhombusFigureObject(t7)],
          c4 : setPrag[inFocus(Y), inFocus(t7)],
          c4 : addcr[t7, saddlebrownFigureObject(t7)] }

```

```

,
i2,2 = { c4 : tacitAbandonTasks[2],
          c4 : addcr[t7, rhombusFigureObject(t7)],
          c4 : setPrag[inFocus(Y), inFocus(t7)],
          c4 : addcr[t7, sandybrownFigureObject(t7)] }

```


Learning interpretive preferences

The previous experimental data provides labels that can be used to fit probabilistic models that capture dialogue preferences (DeVault 2008:§4.4).

2 COREF: is the target round?

3 c4: brown diamond

4 COREF: do you mean dark brown?

5 c4: **yes**

```
i1,1 = { COREF : tacitNop[[c4 does clickContinue[]]],
          c4 : pushCollabRef[c4, COREF, t7],
          COREF : pushYNQ[COREF, c4, addcr[t7, P], negcr[t7, P]],
          COREF : askYNQ[addcr[t7, circleFigureObject(t7)]],
          COREF : setPrag[inFocus(Y), inFocus(t7)]}
```

```
i2,1 = { c4 : tacitAbandonTasks [2],
          c4 : addcr[t7, rhombusFigureObject(t7)],
          c4 : setPrag[inFocus(Y), inFocus(t7)],
          c4 : addcr[t7, saddlebrownFigureObject(t7)] } CORRECT
```

```
,
i2,2 = { c4 : tacitAbandonTasks [2],
          c4 : addcr[t7, rhombusFigureObject(t7)],
          c4 : setPrag[inFocus(Y), inFocus(t7)],
          c4 : addcr[t7, sandybrownFigureObject(t7)] } INCORRECT
```

Cards Corpus connections: underspecified card references

Player 2: Look for 2.

Player 1: and the 3?

Cards Corpus connections: underspecified card references

The players are holding {4H,KH}

Player 2: Look for 2.

Player 1: and the 3?

Cards Corpus connections: underspecified card references

Underspecified referential expressions are extremely common:

The players are holding {4H,KH}

Player 2: Look for 2.

Player 1: and the 3?

Phrase type	Count
Fully specified	103 (37%)
Underspecified	172 (63%)
Total	275

Table: Annotations for 10 transcripts.

Cards Corpus connections: underspecified card references

Underspecified referential expressions are extremely common:

The players are holding {4H,KH}

Player 2: Look for 2.

Player 1: and the 3?

Phrase type	Count
Fully specified	103 (37%)
Underspecified	172 (63%)
Total	275

Table: Annotations for 10 transcripts.

The corpus provides extensive information that can be used to disambiguate such utterances:

- The cards the players are currently holding.
- The cards picked up in the past.
- The cards mentioned in the past.
- The cards manipulated immediately before/after underspecified references.
- The players' utterances surrounding the underspecified reference.

(Djalali et al. 2012)

Cards Corpus connections: quantifier domains

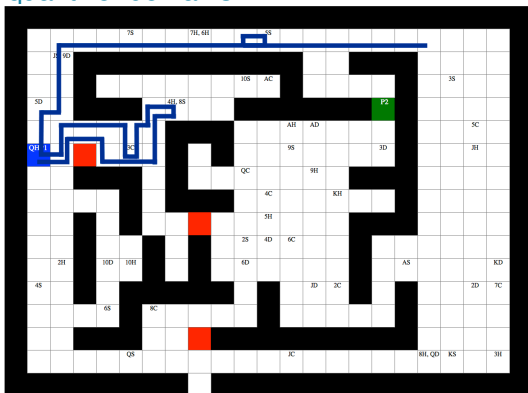
P2: lets do spades

P2: I have the as, qs,
and ks

[...]

P2: ok, i found js

P1: Ok. I haven't found
anything...lol



Cards Corpus connections: quantifier domains

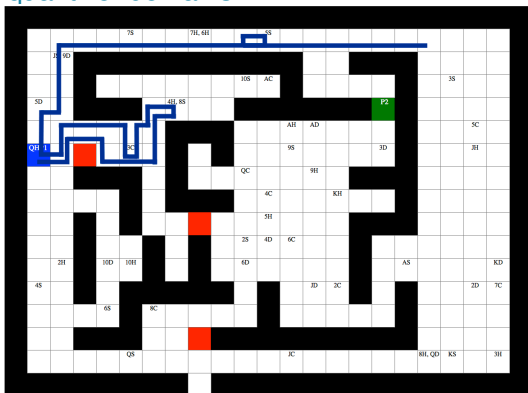
P2: lets do spades

P2: I have the as, qs,
and ks

[...]

P2: ok, i found js

P1: Ok. I haven't found
anything...lol



Quantifier	Literally true	Literally false
<i>anything</i>	2	6
<i>nothing</i>	0	6
Total	2 (14%)	12 (86%)

Table: In context (find|found|see|saw) (any|no)thing. Effectively no tokens are interpreted literally.

Summary of corpus resources

- **SwDA:**
<http://www.stanford.edu/~jurafsky/ws97/>
- **SwDA with Treebank3 alignment:**
<http://comp Prag.christopherpotts.net/swda.html>
- **Edinburgh Map Corpus:**
<http://groups.inf.ed.ac.uk/maptask/>
- **TRIPS:**
<http://www.cs.rochester.edu/research/cisd/projects/trips/>
- **TRAINS:**
<http://www.cs.rochester.edu/research/cisd/projects/trains/>
- **Cards:**
<http://CardsCorpus.christopherpotts.net/>
- **SCARE:**
<http://slate.cse.ohio-state.edu/quake-corpora/scare/>
- **The Carnegie Mellon Communicator Corpus (human-computer transcripts):**
<http://www.speech.cs.cmu.edu/Communicator/>

Summary and prospects

Summary

- Dialogue demands knowledge of the world/domain and knowledge of conversational norms.
- Dialogue is situated; implemented systems must be grounded in the world/action.
- Dialogue is joint action (Clark 1996); our discourse moves are collaborative.

Prospects (exciting ideas)

- Human-subjects testing for evaluation (an old idea reborn!).
- Implemented systems should provide their own training data (DeVault and Stone 2009; Artzi and Zettlemoyer 2011).
- Flexible notions of rich context: webpages, mobile device states, semi-structured databases.
- Language as action: interpretation and production are driven by the evolving goals of the conversation.

References I

- Allen, James F.; Nathanael Chambers; George Ferguson; Lucian Galescu; Hyuckchul Jung; Mary Swift; and William Taysom. 2007. PLOW: A collaborative task learning agent. In *Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence*, 1514–1519. Vancouver, British Columbia, Canada: AAAI Press.
- Artzi, Yoav and Luke Zettlemoyer. 2011. Bootstrapping semantic parsers from conversations. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, 421–432. Edinburgh, Scotland, UK.: Association for Computational Linguistics. URL <http://www.aclweb.org/anthology/D11-1039>.
- Bangalore, Srinivas; Giuseppe Di Fabbrizio; and Amanda Stent. 2006. Learning the structure of task-driven human-human dialogs. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, 201–208. Sydney, Australia: Association for Computational Linguistics. URL <http://www.aclweb.org/anthology/P/P06/P06-1026>.
- Blaylock, Nate and James F. Allen. 2005. Generating artificial corpora for plan recognition. In Liliana Ardissono; Paul Brna; and Antonija Mitrovic, eds., *User Modeling 2005*, Lecture Notes in Artificial Intelligence, 179–188. Berlin: Springer.
- Büring, Daniel. 1999. Topic. In Peter Bosch and Rob van der Sandt, eds., *Focus — Linguistic, Cognitive, and Computational Perspectives*, 142–165. Cambridge: Cambridge University Press.
- Calhoun, Sasha; Jean Carletta; Jason Brenier; Neil Mayo; Daniel Jurafsky; Mark Steedman; and David Beaver. 2010. The NXT-format Switchboard corpus: A rich resource for investigating the syntax, semantics, pragmatics and prosody of dialogue. *Language Resources and Evaluation* 44(4):387–419.
- Clark, Herbert H. 1996. *Using Language*. Cambridge: Cambridge University Press.
- Clark, Herbert H. and Deanna Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition* 22(1):1–39.
- Clark, Robin and Prashant Parikh. 2007. Game theory and discourse anaphora. *Journal of Logic, Language, and Information* 16(3):265–282.

References II

- DeVault, David. 2008. *Contribution Tracking: Participating in Task-Oriented Dialogue under Uncertainty*. Ph.D. thesis, Rutgers University, New Brunswick, NJ.
- DeVault, David; Natalia Kariaeva; Anubha Kothari; Iris Oved; and Matthew Stone. 2005. An information-state approach to collaborative reference. In *Proceedings of the ACL Interactive Poster and Demonstration Sessions*, 1–4. Ann Arbor, MI: Association for Computational Linguistics.
- DeVault, David and Matthew Stone. 2007. Managing ambiguities across utterances in dialogue. In Ron Artstein and Laure Vieu, eds., *Proceedings of DECALOG 2007: Workshop on the Semantics and Pragmatics of Dialogue*.
- DeVault, David and Matthew Stone. 2009. Learning to interpret utterances using dialogue history. In *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)*, 184–192. Athens, Greece: Association for Computational Linguistics. URL <http://www.aclweb.org/anthology/E09-1022>.
- Djalali, Alex; David Clausen; Sven Lauer; Karl Schultz; and Christopher Potts. 2011. Modeling expert effects and common ground using Questions Under Discussion. In *Proceedings of the AAAI Workshop on Building Representations of Common Ground with Intelligent Agents*. Washington, DC: Association for the Advancement of Artificial Intelligence.
- Djalali, Alex; Sven Lauer; and Christopher Potts. 2012. Corpus evidence for preference-driven interpretation. In Maria Aloni; Vadim Kimmelman; Floris Roelofsen; Galit Weidman Sassoon; Katrin Schulz; and Matthijs Westera, eds., *Proceedings of the 18th Amsterdam Colloquium: Revised Selected Papers*. Amsterdam: Springer.
- Ginzburg, Jonathan. 1996. Dynamics and the semantics of dialogue. In Jerry Seligman, ed., *Language, Logic, and Computation*, volume 1. Stanford, CA: CSLI.
- Groenendijk, Jeroen and Martin Stokhof. 1982. Semantic analysis of wh-complements. *Linguistics and Philosophy* 5(2):175–233.
- Jurafsky, Daniel and James H. Martin. 2009. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Englewood Cliffs, NJ: Prentice-Hall, 2nd edition.

References III

- Kumar Rangarajan Sridhar, Vivek; Srinivas Bangalore; and Shrikanth Narayanan. 2009. Combining lexical, syntactic and prosodic cues for improved online dialog act tagging. *Computer Speech and Language* 23(4):407–422. doi:\bibinfo{doi}{10.1016/j.csl.2008.12.001}. URL <http://www.sciencedirect.com/science/article/pii/S0885230808000569>.
- Partee, Barbara H. 1973. Some structural analogies between tenses and pronouns in English. *The Journal of Philosophy* 70(18):601–609. Reprinted in Partee (2004), 50–58. Page reference to the reprinting.
- Partee, Barbara H. 2004. *Compositionality in Formal Semantics: Selected Papers of Barbara H. Partee*, volume 1 of *Explorations in Semantics*. Oxford: Blackwell Publishing.
- Perrault, C. Raymond and James F. Allen. 1980. A plan-based analysis of indirect speech acts. *American Journal of Computational Linguistics* 6(3-4):167–182.
- Roberts, Craige. 1996. Information structure: Towards an integrated formal theory of pragmatics. In Jae Hak Yoon and Andreas Kathol, eds., *OSU Working Papers in Linguistics*, volume 49: Papers in Semantics, 91–136. Columbus, OH: The Ohio State University Department of Linguistics. Revised 1998.
- Shriberg, Elizabeth; Rebecca Bates; Paul Taylor; Andreas Stolcke; Daniel Jurafsky; Klaus Ries; Noah Coccaro; Rachel Martin; Marie Meteer; and Carol Van Ess-Dykema. 1998. Can prosody aid the automatic classification of dialog acts in conversational speech? *Language and Speech* 41(3–4):439–487.
- Stolcke, Andreas; Klaus Ries; Noah Coccaro; Elizabeth Shriberg; Rebecca Bates; Daniel Jurafsky; Paul Taylor; Rachel Martin; Marie Meteer; and Carol Van Ess-Dykema. 2000. Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational Linguistics* 26(3):339–371.
- Thompson, Henry S.; Anne Anderson; Ellen Gurman Bard; Gwyneth Doherty-Sneddon; Alison Newlands; and Cathy Sotillo. 1993. The HCRC map task corpus: Natural dialogue for speech recognition. In *HLT '93: Proceedings of the workshop on Human Language Technology*, 25–30. Morristown, NJ: Association for Computational Linguistics.