-				
r 1				0147
	v		v	evv

The corpus 00000

Experts and common ground 0000000

# Collecting task-oriented dialogues

### David Clausen and Christopher Potts Stanford Linguistics

Workshop on Crowdsourcing Technologies for Language and Cognition Studies Boulder, July 27, 2011



The corpus

Experts and common ground

Conclusions

# Collaborators



David Clausen



Alex Djalali



Sven Lauer

Overview	Scenario and implementation	The corpus	Experts and common ground
	0000000	00000	0000000
	0000000	00000	000000

- 1 The ups and downs of crowdsourcing a time-consuming collaborative task.
- 2 Properties of the resulting corpus.
- Engaging pragmatic theory: expert effects, common ground, and presupposition accommodation

Overview	Scenario and implementation	The corpus	Experts and common ground	Conclusions

Description: Two-player collaborative video game involving

Payment: \$0.40 \$1.00, and up to \$0.20 \$0.50 cents for rich, collaborative problem-solving using meaningful dialogue.
 Restrictions: US IP addresses; at least 95% approval rating

Title: Collaborative Search Game with Chat

dialogue/chat with other Turkers.

ΗΙΤ

•

Experts and common ground 0000000

### Instructions

#### Instructions

- You'll first be presented with an experimental consent form; the transcripts from these games will be used to study collaborative problem solving.
- 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
- 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".
- 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

Overview	-					
Creiview	( )					12
	$\cup$	v		V.		ν.

Experts and common ground

## Gameboard



Experts and common ground 0000000

### Consent form

# 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

Å.

Overview	Scenario and implementation	The corpus	Experts and common
	00000000	00000	0000000

### Scenario

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.

Overview	Scenario and imple	mentation	The corpus	Experts and common ground	Conclusions
Transcr	ipts				
Server Server Server Server	, 0: TASK_COMI , 0: PLAYER_1/ , 0: PLAYER_2/ , 56: MAX_LINI	PLETED2010-06 A1OBNPQ9TFS88 A253Q11TZPQPJ EOFSIGHT3	5-13 01:01:02 BE CZ		
Server	, 118: CREATE	_ENVIRONMENT	NEW_SECTION		
		;	1,2:2D;1,7:K	CH;1,7:9S;1,11:6C;1,13	:QC;1,14:QS;
-		-;	2,18:3H;2,18	3:9H;	
		-;	3,19:4H;4,8:	AC;4,19:3D;	
		-;	4,19:KD;		
		-;	5,14:QH;5,15	5:5S;5,15:2S;5,16:4D;5	,16:10C;5,18:4
-		-;	6,11:KC;6,15	5:9C;	
– b		- :	7,11:2H;7,13	3:7S;	
		-:	8.2:0D:8.4:A	D:8.11:JC:8.20:8S:	
		-:	9.9:10S:9.9:	6H:9.9:8C:9.10:7H:9.14	4:JS:
	b	-:	10.1:2C:10.1	Q:8D:11.14:6D:11.14:10	9H:
		-:	11.18:4C:11.	18:9D:	. ,
		-;	12,10:3S;12,	12:6S;12,16:5H;12,16:	JD;12,20:3C;

13,4:5C;13,4:JH;13,15:KS;

14,2:5D;14,20:10D;15,2:AH;

-; 15,13:7D;15,15:8H;15,17:AS;15,20:7C; Server, 118: MAX\_CARDS3

Server, 118: GOAL\_DESCRIPTION [...]

-:

--- b-----

-----;

-:

# Transcripts

Player 1, 566650: PLAYER\_MOVE7,11 Player 2, 567771: CHAT\_MESSAGE\_PREFIXwhich c's do you have again? Player 1, 576500: CHAT\_MESSAGE\_PREFIXi have a 5c and an 8c Player 2, 577907: CHAT\_MESSAGE\_PREFIXi jsut found a 4 of clubs Player 1, 581474: PLAYER\_PICKUP\_CARD7,11:8C Player 1, 586098: PLAYER\_MOVE7,10

The corpus

Experts and common ground

Conclusions

# The Pragbot platform

Extensible Java program developed by Karl Schultz. Handles high traffic well. Intuitive transcript design and helpful logging. Plays nicely with the outside world.



- Specify the task (or task family).
- Design the map (simple text format).
- Set all high-level contextual parameters (line of sight, max moves, max cards, hidden walls).
- Two humans, or one human and one bot.

# Data collection

- Data collection in June 2010. •
- PHP wrapper to Pragbot written by Victoria Schwanda.
- Server-side configuration by Chriz Czyzewicz.
- Collection time: 5 batches each lasting about 5 hours, spread • out over two work weeks.
- At peak times: 30 transcripts per hour.
- Total cost: about \$1,000

$\sim$				
				5.7
		v		ν.

The corpus ○●○○○ Experts and common ground

# Sample run

000	) Pragb	iot game	board		+	_				_		_		_	Pragbot	gamebo	ard				
START	← BA	CK F	ORWAR	aD →	FINISH	(keyt	ooard na	vigatior	n: ← bac	k I → fo	rward I J	start I (	finish)								
																					Player 1 holding: 85, 75, 55 Player 2 holding: 65, 95, 105
					QH																Current event Player 1, 902946: CLOSE_SOCKETS
		-												-						-	<ol> <li>Player 1 (36836): OK, we need to choose a suit - spades, etc. Any preference?</li> </ol>
		QD					8C							5C	AH					2H	<ol> <li>Player 2 (63662): 1 just got here. Doesn't matter to me.</li> <li>Player 1 (73356): Ok, how about spades?</li> <li>Player 2 (79177): Sounds and</li> </ol>
		-						-												4D	<ol> <li>Player 1 (98577): Use arrow keys to move around maze, look for yellow boxes that have S</li> </ol>
				9C												QS		10D			6. Player 1 (122058); 7. Player 1 (136792); Picked up 8 S 8. Player 2 (143520); Great
						_				35		_				6D, KS			5D	л	<ol> <li>Player 1 (170895): We need consecutives: 5,6,7, or 9,10,11</li> <li>Player 1 (286170): picked up 10 Spades; we need the 9S</li> </ol>
						70								к							<ol> <li>Player 2 (302728): I haven't found any yet. I'm still looking.</li> <li>Player 2 (350053): I keep getting red squares. Do you know what they mean?</li> </ol>
											9D 2D	00			108						13. Player 1 (362998): No, I was wondering. 14. Player 2 (538526): I have a 6 and 9S
															AS, 6H						<ol> <li>Player 1 (552754): picked up 7 spades. I can only hold 3 cards at a time.</li> <li>Figure 2 (552065): So what do no hous non?</li> </ol>
		811					45		AU.				m						л		17. Player 1 (500774): Wowl 18. Player 1 (500774): Wowl 18. Player 1 (588214): We either need 9S or 5S
										25	KC			7D					8D		<ol> <li>Player 2 (624182): OK. I just saw the 2 and 3.</li> <li>Player 2 (630221): Fill keep going.</li> <li>Baren 1 (68924): OS 55 in theme shift commerciate commerciation.</li> </ol>
	3H																	AC			<ol> <li>Player 1 (711580): Should drop 10?</li> <li>Player 2 (711608): Outside the maze?</li> </ol>
												4H	2C					кн			<ol> <li>Player 1 (750054): Our mazes might be different. Maybe only MY maze has it at lower right. Til drop 10S</li> <li>Flayer 1 (782180): Ob. was used a mr. 5, 2, and 8 of smokes.</li> </ol>
		3C	10C			9H															<ol> <li>Hayer 2 (808180): I have 6 &amp; 9.</li> <li>Player 2 (808180): I can grab 10</li> </ol>
	3D																			P1, P2	<ol> <li>Player 2 (825741): Will that be it?</li> <li>Player 2 (830685): So now I have 6, 9, 10. What do you have?</li> <li>Rear 1 (255161): Terriford think are printed by 5 and 7 and 8</li> </ol>
	SH																		4C	KD	<ol> <li>Player 2 (865446): Yeah! That's it! Thanks.</li> </ol>
Julie																					u 2

# Basic corpus stats

- 439 transcripts
- 111 unique players
- Game length mean: 465 actions (median 392, sd 263)
- Actions:
  - Card pickup: 8,330
  - Card drop: 6,105
  - Move: 175,503
  - Utterance: 12,280
    - Utterance length mean: 5.28 words (median 4, sd 4.78)
    - Total word count: 64,900
    - Total vocabulary: 3,149 (stemmed and with card-reference regularization: 2,255)

# Email feedback from our Turkers

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?

The corpus ○○○○● Experts and common ground 0000000

### A variation: Some games are impossible

a +	- BA	ск в	ORWAI	RD →	FINISH	(keyl	board na	vigation	n: ← bac	k I → fo	rward I J	start I	† finish)								
																					Player 1 holding: Player 2 holding:
				4C			45	JS						кн					3D		Current event Player 2, 1230: PLAYER_INITIAL_LOCATION3,1
	75																4H				
1	чн	5D					ASP1		JC	10H			P2						QC		
0	2D					6D			QS												
										55				7C					KS		
				-		-								-	6C				-		
	-									5C		8C	7H								
	85									9C, 98		105					68				
	зн						7D					-	2H	5H					кс		
	-						9H, 2C		2D	-				-					-		
,	AC												JD, 10D	-							
	л															10C	38				
	_							25				8H						AD	3C		
	_	KD																		QH	
	_			6H									4D			9D		8D			

The corpus ○○○○● Experts and common ground 0000000

# A variation: Some games are impossible

	vagbot ga	neboard FORWA	RD→ F	INISH (	keyboard	navigatior	i: ← bac	k I → for	rward I 🕽	start I	finish)	file:	// - Prag	bot garr	eboard	_			् इ
	S H 51		4C		9D, / 4S	D,		IC 9H				P2				4H		3D OC	Player 1 holding: AS, QS Player 2 holding: KH Current even Hyper 1, 142025; CLOSE, SOCKETS 1, Hyper 2, 12373; hepyr 2, Hyper 1 (14532); showing? 3, Hyper 1 (14532); showing? 4, Player 2 (12299); sure
	D							10H, 7D, 2C, 6DP1 5S		8C	7H	70	60					KS	S. Flager 1 (431); have QS and AS     Flager 1 (431); have QS and AS     Flager 2 (1967b; S)     Flager 2 (197b; S)
	s H .C						2D	90,98		108	2H JD, 10D	SH				68		KC	<ol> <li>Piger 2 (19255): heres"</li> <li>Piger 2 (19255): heres"</li> <li>Piger 2 (19757): just typical typical typical typical typical (19757): just typical (19757): here the like 920 (19757): here the like 920 (19757): here typical (197577): here typical (197577): here typical (1975777): here typical (197577777): here typical (197577777777777777777777777777777777777</li></ol>
	H KI		6H			25				8H	4D			100		38	6,8D	3C	22. Figure 1 (0329), then is two of every stati 32. Figure 1 (03290), then is two of every stati 30. Figure 2 (03299), then means its impact 30. Figure 2 (03199), then, (opensate) 31. Figure 1 (03379, state), (opensate) 33. Figure 1 (033794), years     33. Figure 1 (033794), years
Done																			

The corpus

Experts and common ground

Conclusions

### **Experience** levels



- The more a person played, the fewer utterances they used. This is true regardless of whether their partner was also experienced.
- If both players were experienced, the effect was even more dramatic.
- Expert transcripts were not necessarily shorter, though; some experts exhaustively searched independently, gathered subsets of the cards, and then assessed what they had found.

Player

Overview	Scenario and implementation	The corpus	Experts and common ground
	0000000	00000	000000

### Novice strategy

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 Courtney, by the way- nice to meet you.
- Player 2: i'm becky....nice to meet you too
- Player 1: Hi Becky. How about we go for hearts? And take 234567

[...]

Overview	Scenario and implementation	The corpus	Experts and common ground	Conclu
	0000000	00000	000000	

### Journeyman strategy

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

Experts and common ground

# Expert strategy

Player	2:	hi
Player	1:	hiwhich 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

### Depth annotations Depth 1

How do I interact with the game world?

What are the meanings of the various technical terms?

#### Depth 2

What is the goal of the game generally?

#### Depth 3

What is the configuration of the game board?

What is the expertise of my fellow player?

#### Depth 4

What is the goal of this game specifically?

### Depth 5

How do we achieve this goal generally?

### Depth 6

What cards do we need to achieve this goal specifically?

### Depth 7

Have we completed the game?

P2:	what suit do we want?	(4)
P1:	I hit a KD.	(3)
	I think we should see what we get	,
	and keep the most promising suit	(3)
P2:	i have a JD	(3)
P1:	That works	
P2:	so we are looking for Ds?	(4)
P1:	I vote Ds.	(4)
P2:	okay i have 10D , 9D and JD	(6)
P1:	7D	(6)
P1:	okay do you think my cards work?	(5)
P1:	So we're looking for 8D,	
	and 6D or QD	(6)
P2:	You should be good (I'm slow at	
	this)	(6)

# Initial depth and rich common ground

- We expect experts to be more likely than novices to assume that their partners will accommodate rich contextual knowledge.
- This predicts that experts will initiate discourse at a deeper level in the question graph than novices.
- Furthermore, these effects should be amplified if it is mutual knowledge between the two players that they are both experts.

Overview	Scenario and implementation	The corpus	Experts and common ground	Conclusions

Results

- Greater number of games played (greater expertise) reliably correlates with lower initial utterance depth, though the effect is weak, presumably because one game suffices to become an expert.
- Let **Expert** be true of a player–game pair (*P*, *T*) iff *T* is not *P*'s first game. **Expert** is a very highly correlated with lower initial utterance depth in *T*.
- At the game level, expert–expert pairings have the lowest mean utterance depth, following by mixed pairs, and then novice–novice pairs.

Overview	Scenario and implementation	The corpus	Expe
	0000000	00000	000

# Conclusions

- Crowdsourcing is an effective method for observing collaborative behavior.
- Long wait-times can be problematic.
- But they don't always get in the way of the fun; more than half of our players returned to play again.
- These repeat players permitted us to study the effects of rich common ground on strategic interaction.