Towards a computational model of spatial perspective taking

Simon Dobnik
Dept. of Philosophy, Linguistics & Theory of Science
University of Gothenburg, Sweden
simon.dobnik@gu.se

In order to capture meaning and reference of such spatial descriptions such as "to the left of" and "above" one needs to include (i) perceptual knowledge obtained from scene geometry, (ii) world knowledge about the interaction of objects involved, and (iii) shared knowledge that is established as the common ground that includes both the history of dialogue and perceptually attended scene. A good example that demonstrates the interaction of such information is the assignment of perspective which determines the orientation of the frame of reference frame (FoR). For example, the same table may be described to be "to the left of the chair", "to the right of the chair", "behind the chair" or "South of the chair". The FoR, may be described linguistically "from your view" or "from there" but in a free, spontaneous conversation it is frequently omitted. This means that conversational participants must adopt certain strategies to recover it. For example, they may rely on alignment in the dialogue common ground, or the perceptual properties of the attended scene (salient objects) or on some principles of interaction, for example conversational roles such as information giver and information receiver. In this presentation we describe three lines of work which are leading towards building a computational model that would capture the dynamics of the human FoR assignment for situated artificial agents. Such models resolve considerable ambiguity that agents are facing when interpreting and generating spatial language as well as they lead to a more natural, human-like dialogue.

In the first line of work we use a constrained 3-d virtual environment setup as an online experiment through which we sample interaction data with human participants. The system elicits interactions and records human responses. We investigate what is most likely FoR to start the interaction with (in this visual and discourse environment), whether priming with a particular FoR develops into alignment in the next turn, whether alignment is persistent over several turns or it degrades, and the effects of the change in the roles of information giver and information receiver. This allows us to build a statistical model for interpretation and generation of FoR over short stretches of dialogue which we subsequently test in the reversed interaction with humans.

In the second line of work we investigate strategies of FoR assignment in free, open dialogue between two humans in a similar virtual environment. The results of the dialogue corpora show that humans do align FoR locally over several (1-3) turns but there is no global preference for assignment of particular FoR. We isolate several conversational games where the dynamics of the FoR assignment appears to be linked to other properties of interaction between the agents, for example whether they are focusing on a particular part of the scene or whether they are iden-

tifying individual objects scattered over the entire scene. It follows that alignment is consistently used as a strategy but there are other factors that trigger the change in FoR.

In the third line of work we examine whether a selection/change of the FoR could be predicted from the (textual) dialogue data.We hypothesise that this would contain sufficient information about the dialogue games that conversational participants are engaged in and to which the FoR assignment appears to be linked. Through quantitative data analysis we attempt to identify features that are predictive of FoR changes and which would be useful for annotating and extending our corpus described above. The overall goal of this corpus is to provide a training dataset for machine learning that would allow us to build a model of FoR assignment. Finally, we also investigate a suitability of different machine learning models for the task.

Towards a computational model of spatial perspective taking

Simon Dobnik

with John D Kelleher and Chris Howes

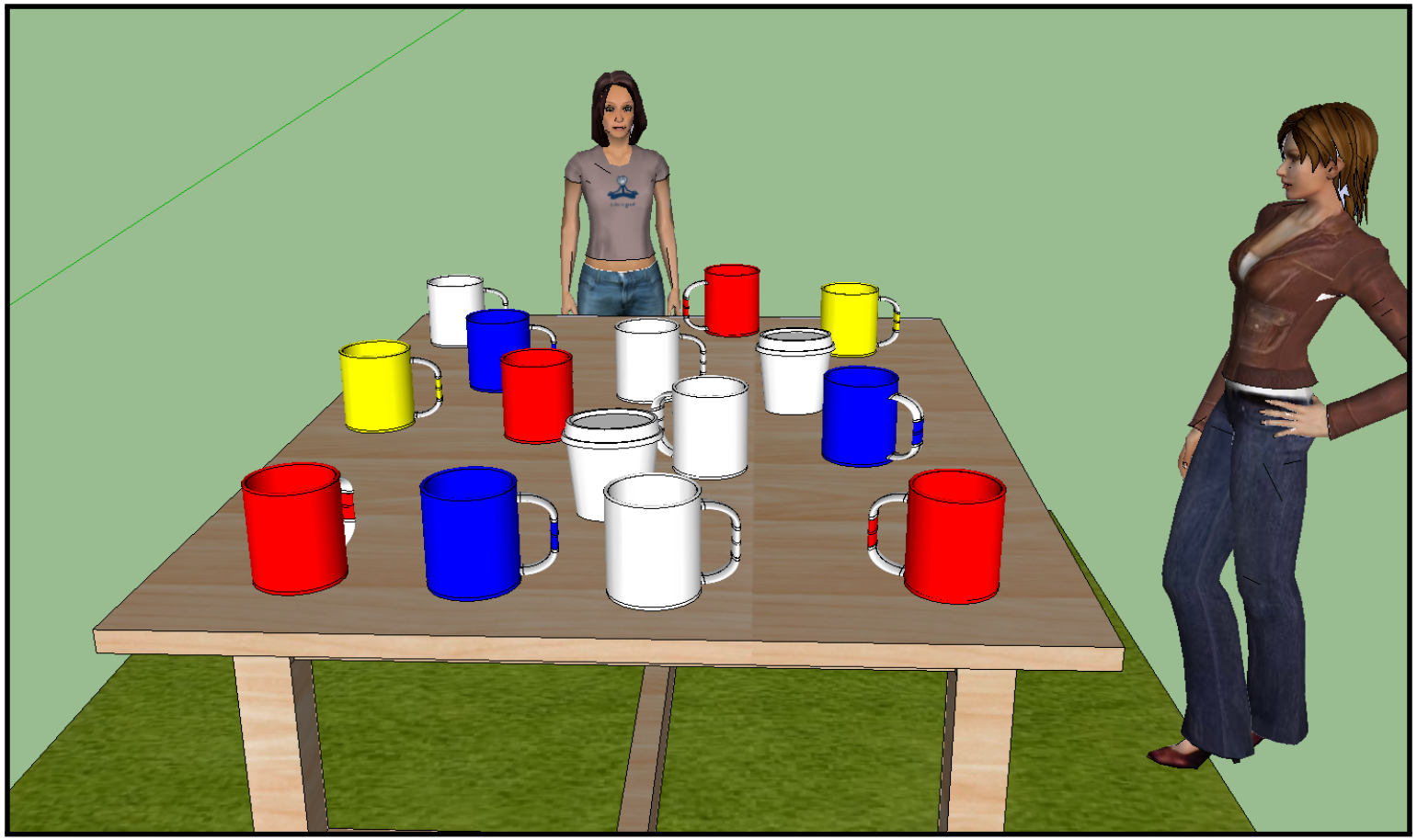
CLASP centre for linguistic theory and studies in probability

University of Gothenburg

Languages as Mechanisms for Interaction: Explorations and Repercussions
Ghent 24th June 2016

Outline

Where is the yellow mug?



Factors affecting the choice of FoR

- task (Tversky, 1991)

Factors affecting the choice of FoR

- task (Tversky, 1991)
- personal style (Levelt, 1982)

Factors affecting the choice of FoR

- task (Tversky, 1991)
- personal style (Levelt, 1982)
- arrangement of the scene and the position of the agent (Taylor and Tversky, 1996; Carlson-Radvansky and Logan, 1997; Kelleher and Costello, 2009; Li et al., 2011),

Factors affecting the choice of FoR

- task (Tversky, 1991)
- personal style (Levelt, 1982)
- arrangement of the scene and the position of the agent (Taylor and Tversky, 1996; Carlson-Radvansky and Logan, 1997; Kelleher and Costello, 2009; Li et al., 2011),
- the presence of a social partner (Duran et al., 2011)

Factors affecting the choice of FoR

- task (Tversky, 1991)
- personal style (Levelt, 1982)
- arrangement of the scene and the position of the agent (Taylor and Tversky, 1996; Carlson-Radvansky and Logan, 1997; Kelleher and Costello, 2009; Li et al., 2011),
- the presence of a social partner (Duran et al., 2011)
- the communicative role and knowledge of information (Schober, 1995)

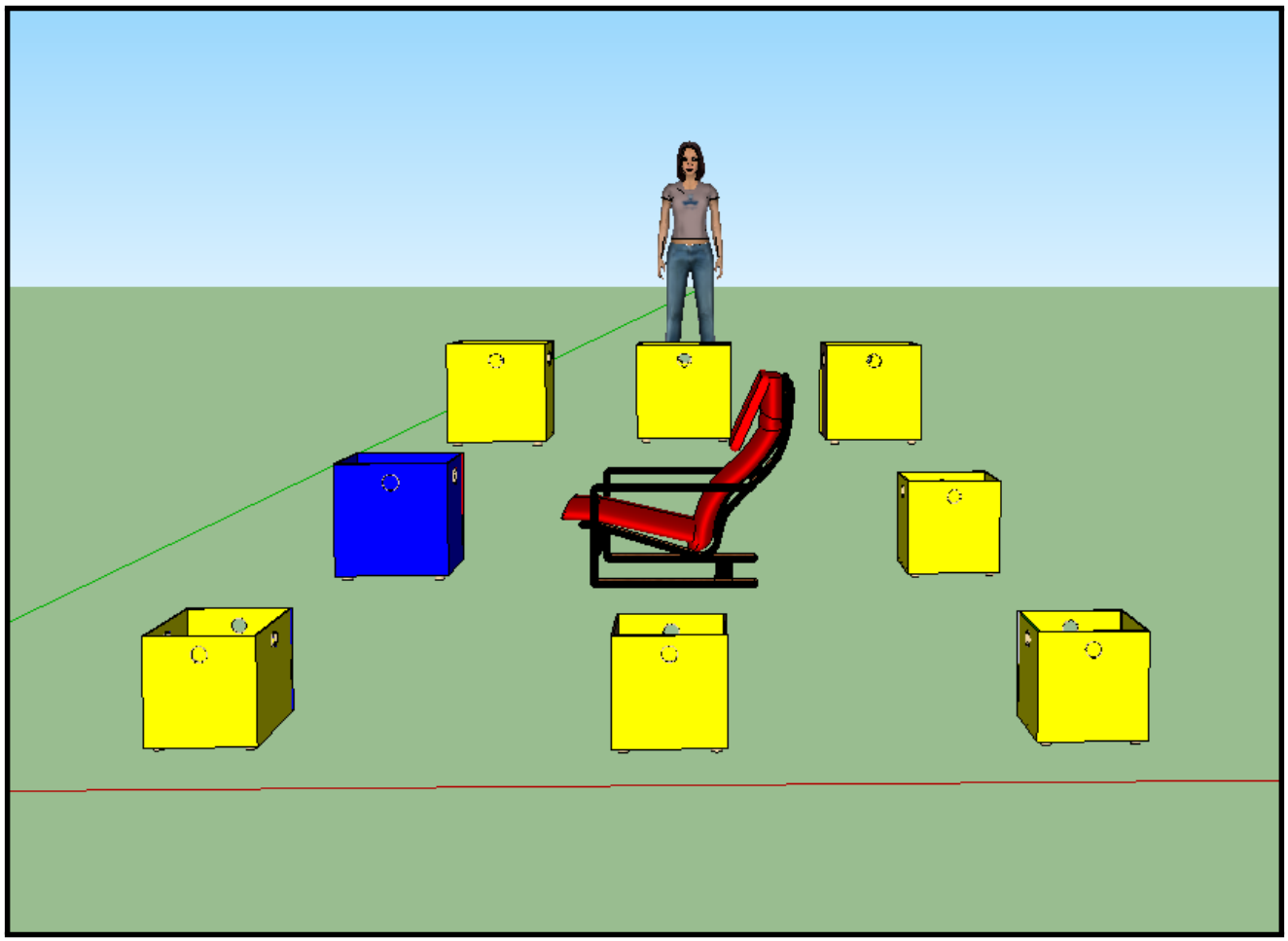
Alignment and FoR

UNIVERSITY OF GOTHENBURG

- Conversational participants align representations (Pickering and Garrod, 2004)
- (Watson et al., 2004; Johannsen and de Ruiter, 2013)
- (Dobnik, Kelleher, and Koniaris, 2014)
  - ... alignment over several turns
  - ... alignment over conversational role changes

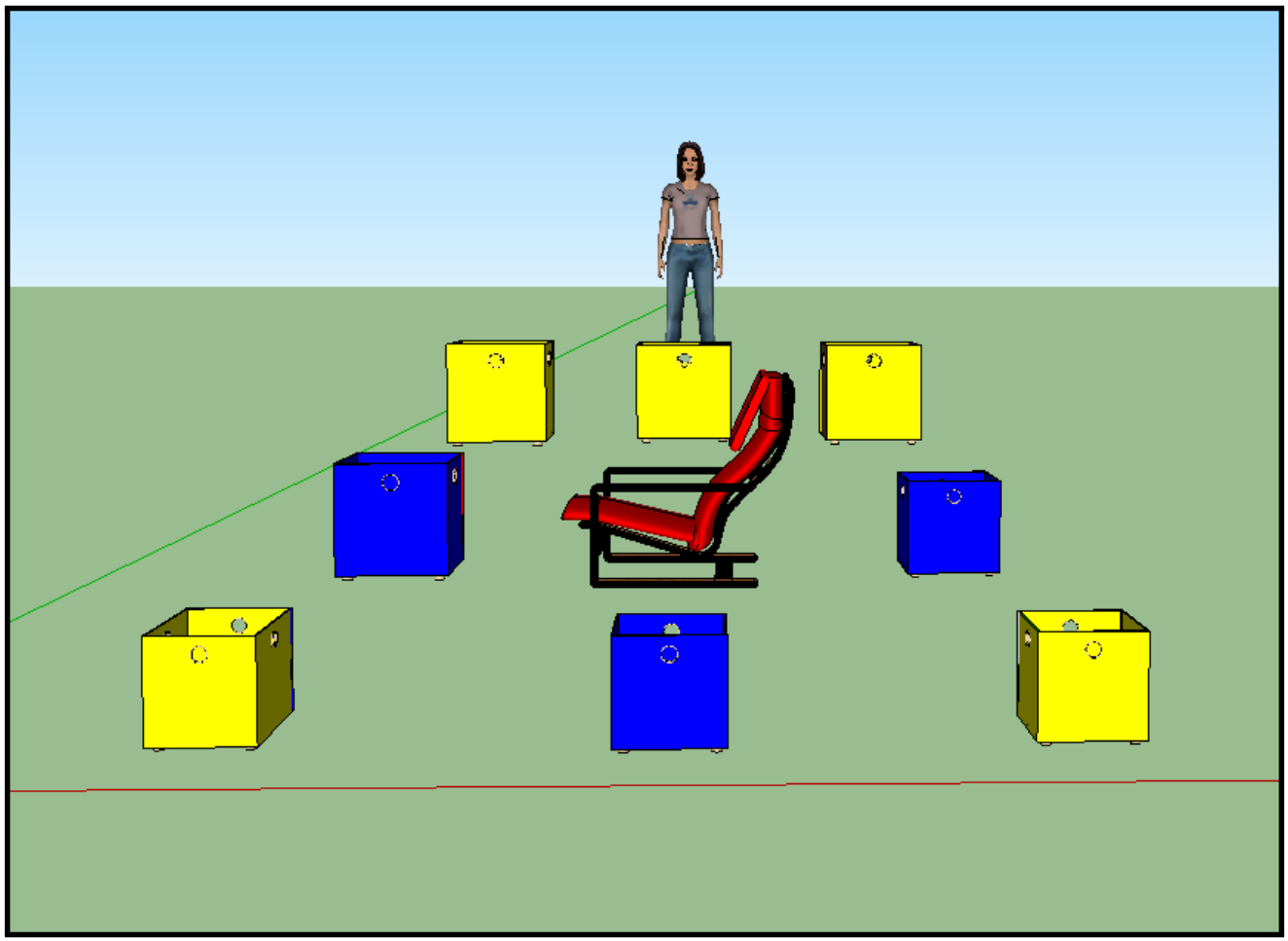
Game 1: Priming

S: “I chose the box in front of the chair.”



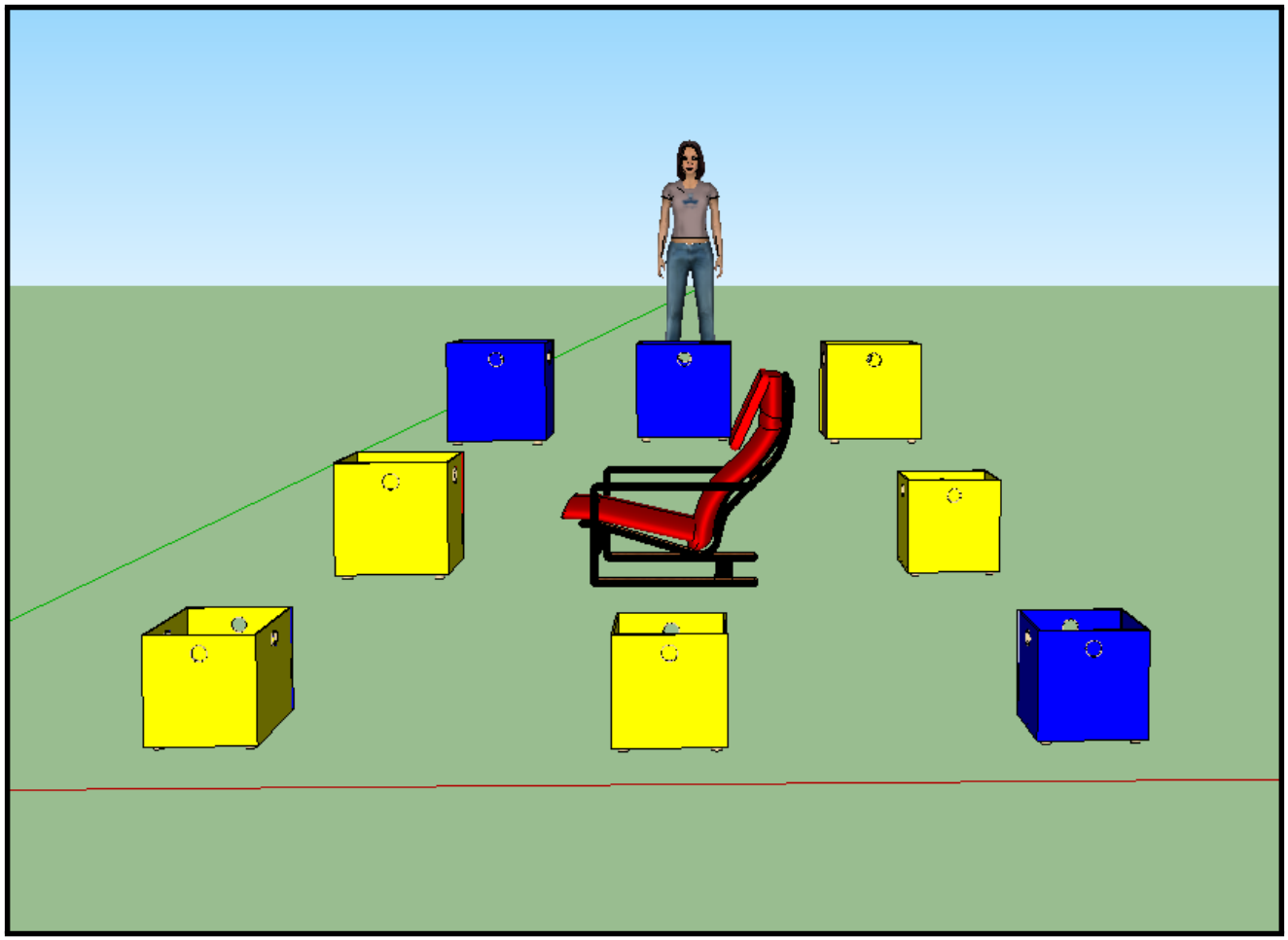
Game 2: Alignment?

S: “I chose the box to the left of the chair.”



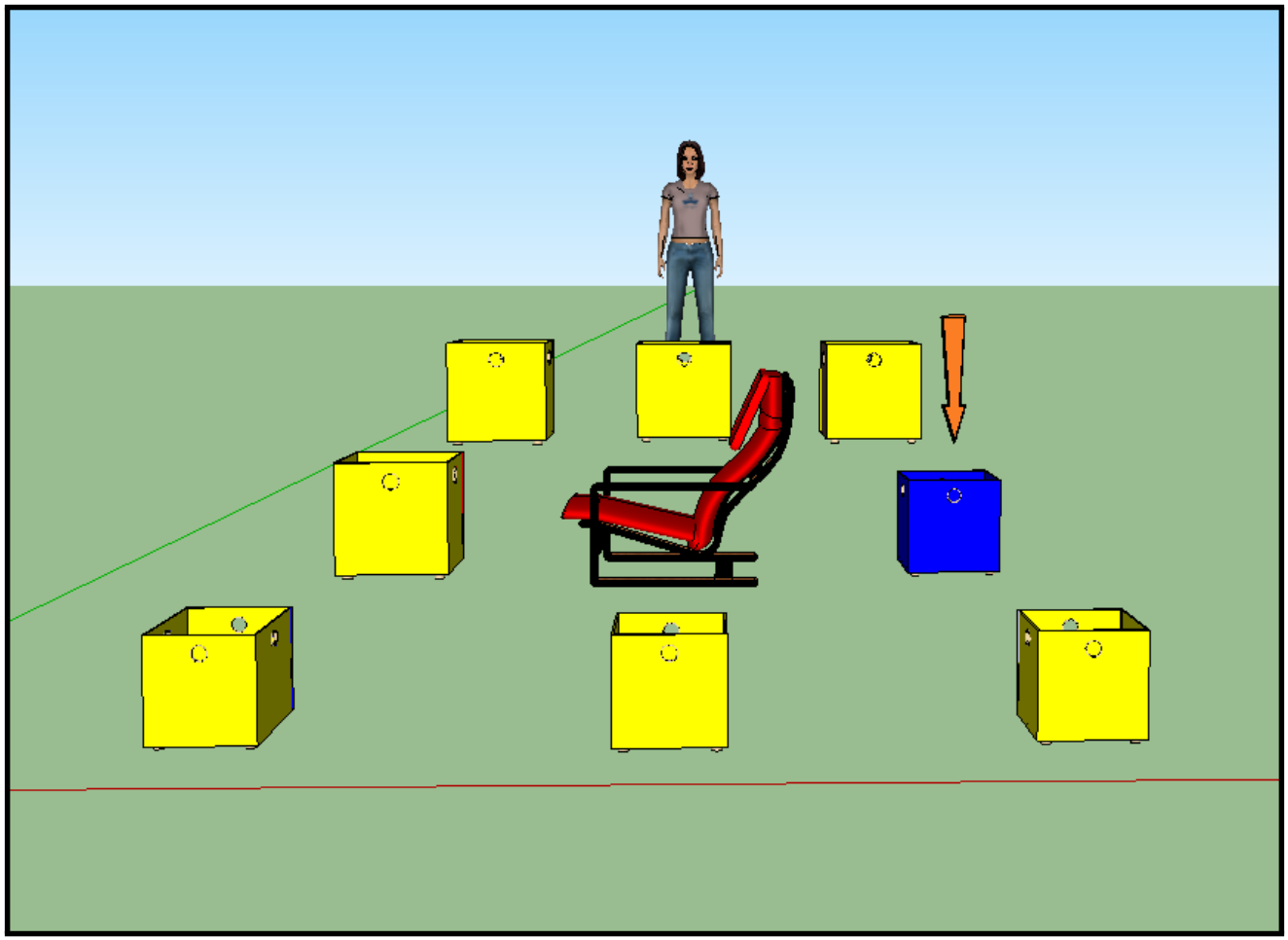
Game 3: Persistent alignment?

S: “I chose the box to the right of the chair.”



Game 4: Change of speaker-hearer roles

S: “Tell me: which box did you choose?”



Observations

UNIVERSITY OF GOTHENBURG

- High degree of alignment to linguistic priming
- Also visual priming and conversational-role priming

Alignment and free dialogue

UNIVERSITY OF GOTHENBURG

- Interactive alignment hypothesis: interlocutors would converge on a FoR
- However...
  - ... they diverge syntactically (Healey et al., 2014)
  - ... In semantic coordination clarification requests decrease convergence (Mills and Healey, 2006)
- Description types driven by mutual understanding and strategies for resolution of misunderstanding.

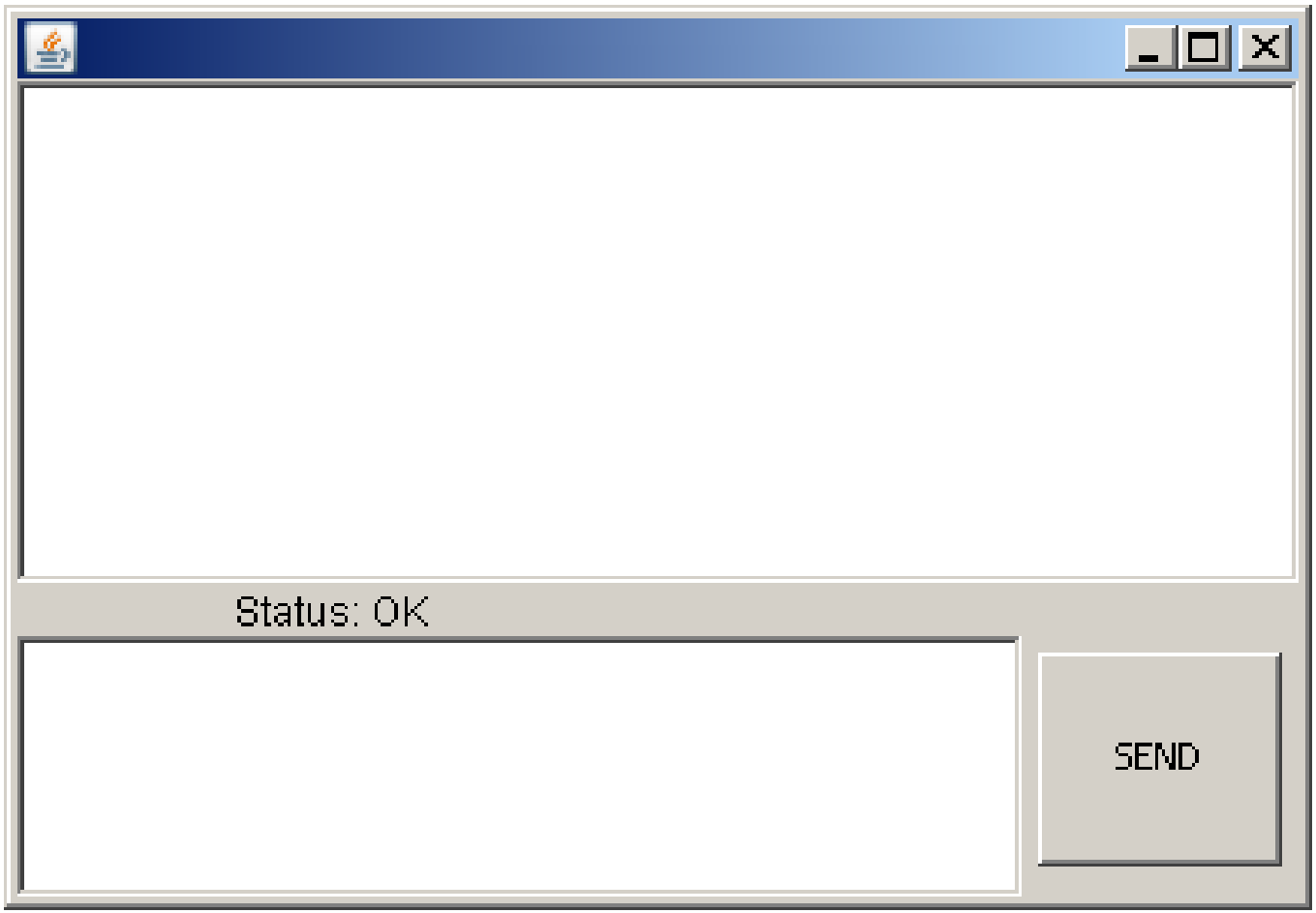
Hypotheses

(1) There is no baseline preference for a specific FoR

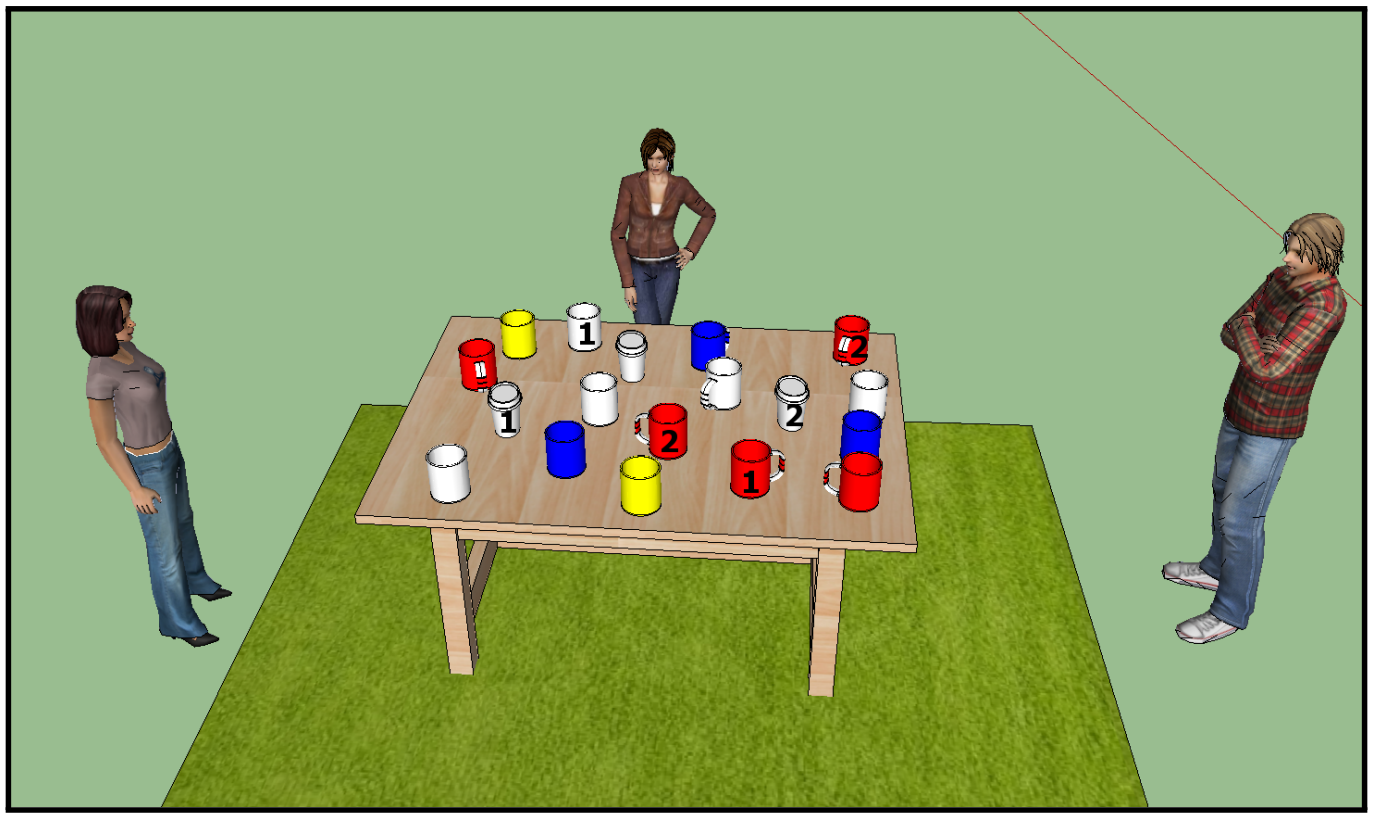
(2) Participants will align on spatial descriptions over the course of the dialogue

(3) Sequences of misunderstanding will prompt the use of different FoRs

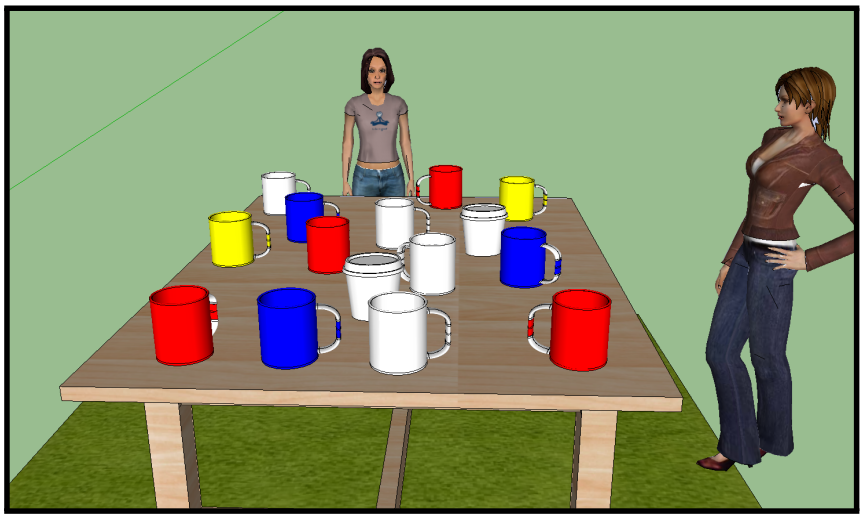
The DiET chat tool



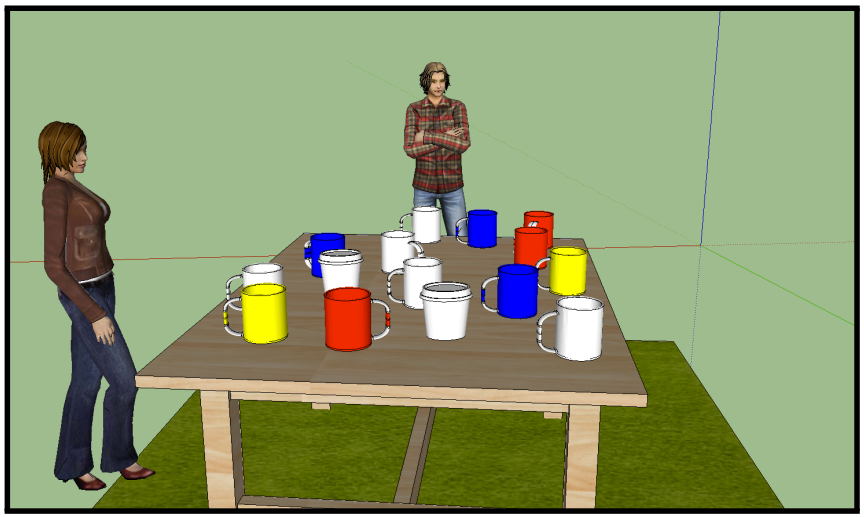
The task



The views



View for participant 1



View for participant 2

The data

- Two pairs of participants (so far!!)
  - Dialogues in English
- Dialogue 1:
  - Swedish first language
  - 30 minutes
  - 157 turns
- Dialogue 2:
  - British English first language
  - 60 minutes
  - 441 turns

Annotation scheme

Tag	Value	Explanation
`is-FoR`	y/n	For all turns: does this turn contain an FoR
`viewpoint`	category	Where `is-FoR=y`: what viewpoint does the FoR use? P1, P2, Katie, object, extrinsic
`topological`	y/n	Where `is-FoR=y`: does the turn contains a topological spatial description such as "near" or "at"?
`explicitness`	y/n	Where `is-FoR=y`: whether the FoR is explicitly referred to, e.g. "on my left"

$\kappa = 0.8121$

Example

20 *P1:* from her right I see yell, white, blue red
spatial, relative-katie, explicit

21 and the white has a funny thing around the top

22 *P2:* then you probably miss the white i see

23 *P1:* and is between yel and bl but furhter away from katie
spatial, relative-katie, explicit, topological

24 *P2:* because i see a normal mug too, right next to the yellow one, on the left
spatial, relative-katie, topological

25 *P1:* ok, is your white one closer to katie than the yellow and blue?
spatial, relative-katie, topological

26 *P2:* yes

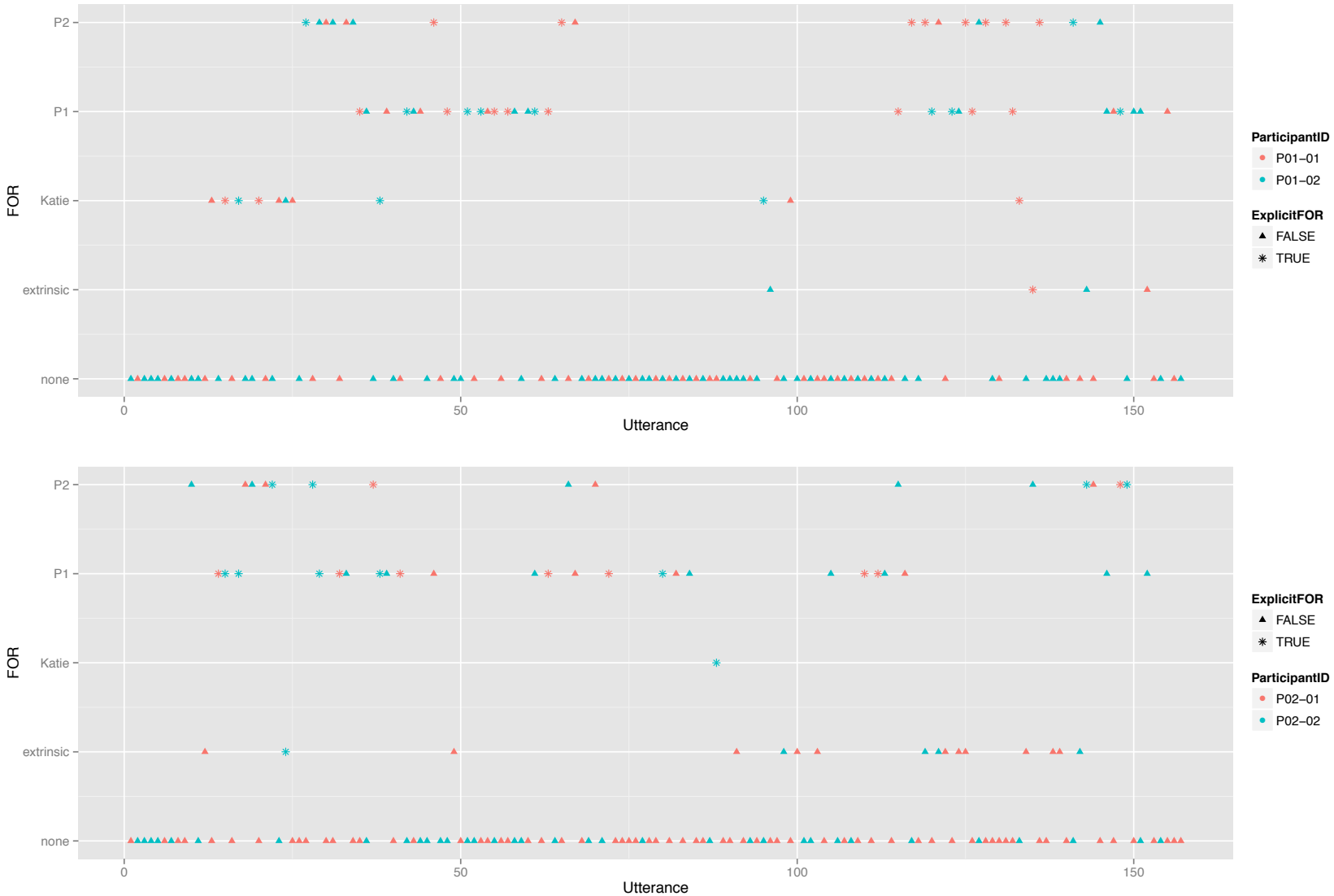
27 closest to me, from right to left:
spatial, relative-p2, topological

28 *P1:* ok, got it

Results: Overview

Category	Turns	%
Contains a spatial description	245	40.97
FoR=P1	88	35.92
FoR=P2	66	26.94
FoR=speaker	81	33.06
FoR=addressee	72	29.39
FoR=Katie	15	6.12
FoR=extrinsic	61	24.90
FoR=unknown	26	10.61
Topological description	44	17.96
Total turns	598	

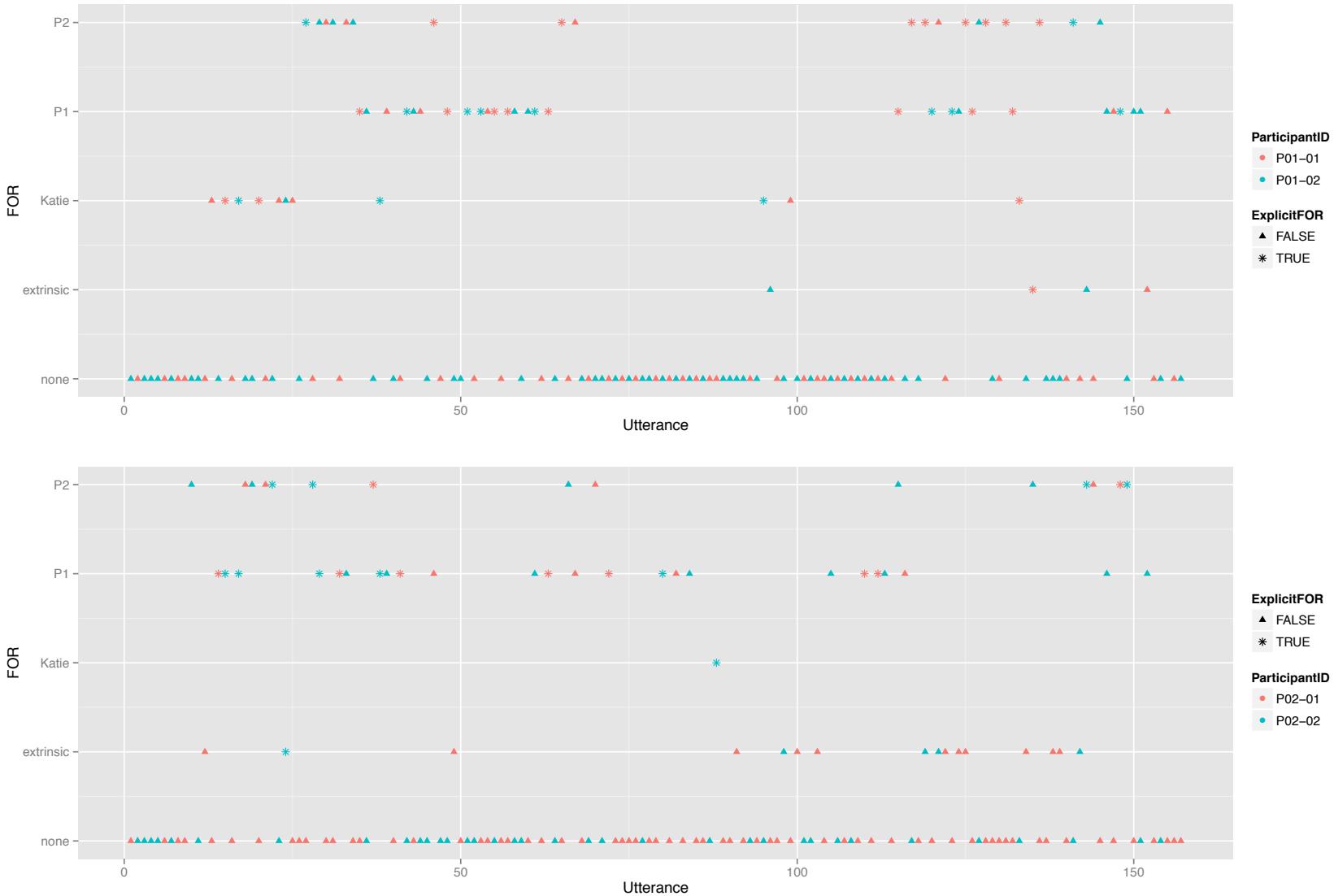
Results: Local alignment



Results: Local alignment

- Participants tend to align to FoR over several turns
- Partial auto-correlations on each binary FoR variable: P1, P2, Katie and Extrinsic
  - Each correlates positively with itself ($p < 0.05$) at 1-3 turns lag
  - use of a particular FoR makes reuse of that FoR more likely
- No significant cross-correlations between the variables

Results: Explicitness



Results: Explicitness

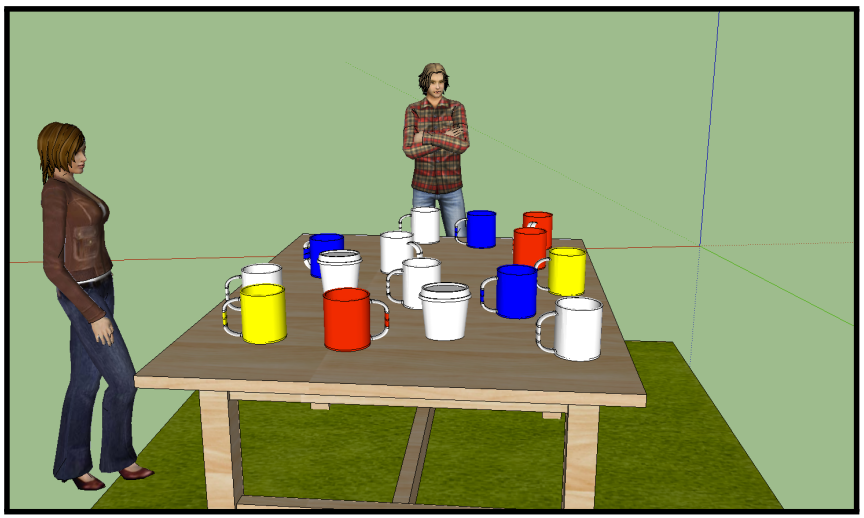
UNIVERSITY OF GOTHENBURG

- Should be no need to describe FoR overtly when local alignment is established
- But FoR referred to explicitly every couple of turns
- Possibly due to task?
  - high potential for referential ambiguity
  - precision critical for success

Changing FoR

Visual properties of the scene

Make a description less ambiguous



P2

25 *P1:* ok, is your white one closer to katie than the yellow and blue? **spatial, relative-katie, topological**

26 *P2:* yes

27 closest to me, from right to left: **spatial, relative-p2, topological**

Task/dialogue game #1

UNIVERSITY OF GOTHENBURG

Focusing on a part of the scene:

- Spatial continuum between the objects
- Fewer distractors
- Higher FoR alignment
- Less explicit FoR marking

Task/dialogue game #1

42 *P2:* there is an empty space on the table on the second row away from you
relative-p1, explicit, topological

43 between the red and white mug (from left to right)
relative-p1

...

48 to my left from that red mug there is a yellow mug
relative-p1, explicit, topological

Task/dialogue game #2

UNIVERSITY OF GOTHENBURG

Scanning the scene for individual objects:

- No spatial continuum between objects
- Several distractors
- Less alignment
- More explicit FoR marking

Task/dialogue game #2

131 *P1:* and the blue ones are one on the second row from you, to the right from you
relative-p2, explicit

132 one slightly to my left
relative-p1, explicit

133 and one in front of katie in the first row
relative-katie, explicit

Clarification and repair

UNIVERSITY OF GOTHENBURG

A new dialogue game and a change of strategy:

- Different perceptual focus on the scene
- A person with the visual access to the scene and clarifying it attracts the FoR (contrary to (Schober, 1995))
- Stronger demand for precision, hence over-specification.

Clarification and repair

Information

14 *P1:* On my first row. I have from the left (your right): . . . Then a red with handle turned to my left.
relative-p1, explicit

. . .

17 *P2:* ok then i think we found a cup of yours that i can't see: the red with the handle to your left (the last one you mention)
relative-p1, explicit

18 *P1:* Okay, that would make sense. Maybe it is blocked by the other cups in front or something?
relative-p2

19 *P2:* yeh, i have a blue one and a white one, either of which could be blocking it
relative-p2

20 *P1:* Yes, I think I see those.

Clarification and repair

UNIVERSITY OF GOTHENBURG

Precision

146 *P2:* so you see that yellow cup to be right on teh corner?
relative-p1

147 *P1:* Yes

148 A yellow cup, on my right your left, with the handle facing east to me, west to you.
relative-p1, relative-p2, explicit

Making it more precise

UNIVERSITY OF GOTHENBURG

- Counting the number of objects of each color
- Agreeing on a specific FoR (D1, Katie)
- Devising a coordinate system with 16 sub-areas

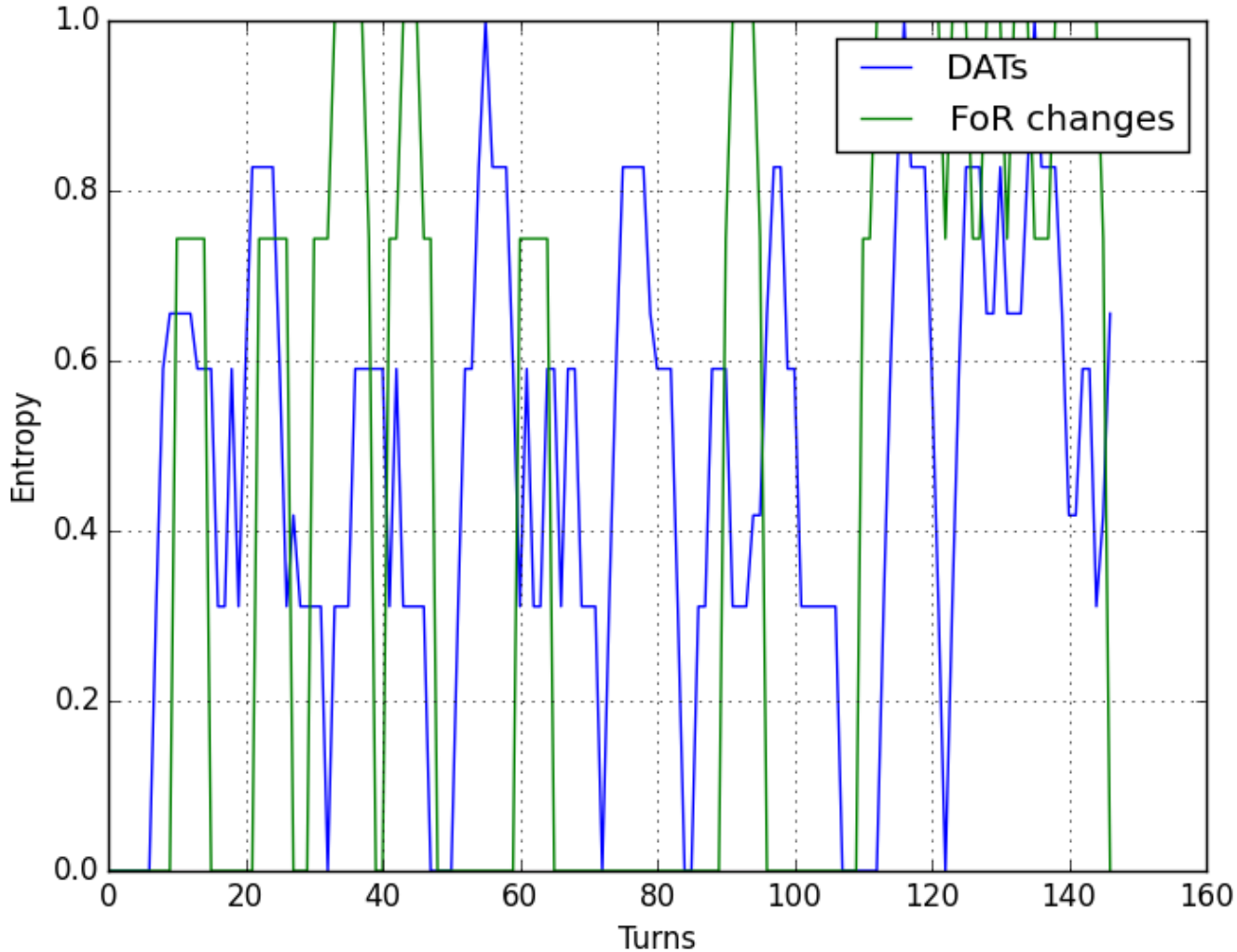
FoR change and dialogue structure

UNIVERSITY OF GOTHENBURG

Does the structure of dialogue games predict FoR change?

- Changes in DA sequences will lead to changes in FoR.
- Tag turns with a DA tagger trained on the NPS Chat Corpus
- Tag turns with changes in FoR
- Entropies of DAs and FoR-changes in a sliding window of 5 turns
  - Spearman's correlation rho $= -0.36$, $p = 0.383$
  - No cross-correlation at different time lags

DA and FoR-change entropies



Conclusions

- A pilot study how FoR is negotiated over several turns of free dialogue
- There is no preference for a specific FoR (H1: supported)
- There is no general alignment of FoR over the entire dialogue but local alignment (H2: unsupported)
- Misunderstanding may be associated with FoR change but there may be other factors (H3: unsupported)
- Strategies/dialogue games used
- Driven by mutual understanding and resolution of misunderstanding

Future work

- Extending the corpus
- Add additional semantic and discourse features that would allow computational modeling of FoR assignment

References I

Carlson-Radvansky, L. A. and G. D. Logan (1997). The influence of reference frame selection on spatial template construction. *Journal of Memory and Language 37*(3), 411–437.

Dobnik, S., J. D. Kelleher, and C. Koniaris (2014, 1–3 September). Priming and alignment of frame of reference in situated conversation. In V. Rieser and P. Muller (Eds.), *Proceedings of DialWatt - Semdial 2014: The 18th Workshop on the Semantics and Pragmatics of Dialogue*, Edinburgh, pp. 43–52.

Duran, N. D., R. Dale, and R. J. Kreuz (2011). Listeners invest in an assumed other's perspective despite cognitive cost. *Cognition 121*(1), 22–40.

Healey, P. G. T., M. Purver, and C. Howes (2014, June). Divergence in dialogue. *PLoS ONE 9*(6), e98598.

Johannsen, K. and J. de Ruiter (2013). Reference frame selection in dialogue: priming or preference? *Frontiers in Human Neuroscience 7*(667), 1–10.

References II

Kelleher, J. D. and F. J. Costello (2009). Applying computational models of spatial prepositions to visually situated dialog. *Computational Linguistics 35*(2), 271–306.

Levelt, W. J. M. (1982). Cognitive styles in the use of spatial direction terms. In R. J. Jarvella and W. Klein (Eds.), *Speech, place, and action*, pp. 251–268. Chichester, United Kingdom: John Wiley and Sons Ltd.

Li, X., L. A. Carlson, W. Mou, M. R. Williams, and J. E. Miller (2011). Describing spatial locations from perception and memory: The influence of intrinsic axes on reference object selection. *Journal of Memory and Language 65*(2), 222–236.

Mills, G. and P. G. T. Healey (2006, September). Clarifying spatial descriptions: Local and global effects on semantic co-ordination. In *Proceedings of the 10th Workshop on the Semantics and Pragmatics of Dialogue (SEMDIAL)*, Potsdam, Germany.

Pickering, M. and S. Garrod (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences 27*, 169–226.

References III

Schober, M. F. (1995). Speakers, addressees, and frames of reference: Whose effort is minimized in conversations about locations? *Discourse Processes 20*(2), 219–247.

Taylor, H. A. and B. Tversky (1996). Perspective in spatial descriptions. *Journal of Memory and Language 35*(3), 371–391.

Tversky, B. (1991). Spatial mental models. *The psychology of learning and motivation: Advances in research and theory 27*, 109–145.

Watson, M. E., M. J. Pickering, and H. P. Branigan (2004). Alignment of reference frames in dialogue. In *Proceedings of the 26th annual conference of the Cognitive Science Society*, pp. 2353–2358. Lawrence Erlbaum Mahwah, NJ.