Changing perspective: Local alignment of reference frames in dialogue

Simon Dobnik and Christine Howes
Centre for Language Technology
University of Gothenburg, Sweden
{simon.dobnik,christine.howes}@gu.se

John D. Kelleher
School of Computing
Dublin Institute of Technology, Ireland
john.d.kelleher@dit.ie

Abstract

In this paper we examine how people negotiate, interpret and repair the frame of reference (FoR) in free dialogues discussing spatial scenes. We describe a pilot study in which participants are given different perspectives of the same scene and asked to locate several objects that are only shown on one of their pictures. This task requires participants to coordinate on FoR in order to identify the missing objects. Preliminary results indicate that conversational participants align locally on FoR but do not converge on a global frame of reference. Misunderstandings lead to clarification sequences in which participants shift the FoR. These findings have implications for situated dialogue systems.

1 Introduction

Directional spatial descriptions such as "to the left of green cup" or "in front of the blue one" require the specification of a frame of reference (FoR) in which the spatial regions "left" and "front" are projected, for example "from where I stand" or "from Katie's point of view". The spatial reference frame can be modelled as a set of three orthogonal axes fixed at some origin (the location of the landmark object) and oriented in a direction determined by the viewpoint (Maillat, 2003).

A good grasp of spatial language is crucial for interactive embodied situated agents or robots which will engage in conversations involving such descriptions. These agents have to build representations of their perceptual environment and connect their interpretations to shared meanings in the common ground (Clark, 1996) through interaction with their human dialogue partners. There are two main challenges surrounding the computational modelling of FoR. Firstly, there are several ways in which the viewpoint may be assigned. If the FoR is assigned by the reference object of the description itself ("green cup" in the first example above) then we talk about *intrinsic* reference frame (after (Levinson, 2003)). Alternatively, the viewpoint can be any conversational participant or object in the scene that has an identifiable front and back in which case we talk about a *relative* FoR. Finally, one can also to refer to the location of objects where the viewpoint is external to the scene, for example, as a superimposed grid structure on a table top with cells such as A1 and B4. In this case it is an *extrinsic* reference frame. There are a number of factors that affect the choice of FoR, including: 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). The second challenge for computational modelling is that the viewpoint may not be overtly specified and must be recovered from the linguistic or perceptual context. Such underspecification may lead to situations where conversational partners fail to accommodate the same FoR leading to miscommunication.

Psycholinguistic research suggests that interlocutors in a dialogue align their utterances at several levels of representation (Pickering and Garrod, 2004), including their spatial representations (Watson et al., 2004). However, as with syntactic priming (Branigan et al., 2000), the evidence comes from controlled experiments with a confederate and single prime-target pairs of pictures, and this leaves open the question of how well such effects scale up to longer unconfined free dialogues. In the case of syntactic priming, corpus studies suggest that interlocutors actually diverge syntactically in free dialogue (Healey et al., 2014).

Semantic coordination has been studied using the Maze Game (Garrod and Anderson, 1987), a task in which interlocutors must produce location descriptions, which can be figurative or abstract. Evidence suggests that dyads converge on more abstract representations, although this is not explicitly negotiated. Additionally, the introduction of clarification requests decreases convergence, suggesting that mutual understanding, and how misunderstandings are resolved is key to shifts in description types (Mills and Healey, 2006). However, both participants see the maze from the same perspective, in contrast to our egocentric, embodied perceptions of everyday scenes.

We are interested in how participants align their spatial representations in free dialogue when they perceive a scene from different perspectives. If the interactive alignment model is correct, although participants may start using different FoRs (using e.g. an egocentric perspective (Keysar, 2007)), they should converge on a particular FoR over the course of the dialogue. We are also concerned with how they identify if a misalignment has occurred, and the strategies they use to get back on track in dialogues describing spatial scenes.

In contrast to several previous studies, this paper investigates the coordination of FoR between two conversational participants over an ongoing dialogue. Our hypotheses are that (i) there is no baseline preference for a specific FoR; (ii) participants will align on spatial descriptions over the course of the dialogue; (iii) sequences of misunderstanding will prompt the use of different FoRs.

2 Method

We describe below our pilot experimental set-up in which participants were required to discuss a visual scene in order to identify objects that were missing from one another's views of the scene.

2.1 Task

Using 3D modelling software (Google SketchUp) we designed a virtual scene depicting a table with several mugs of different colours and shapes placed on it. As shown in Figure 1, the scene includes three people on different sides of the table. The people standing at the opposite side of the table were the avatars of the participants (the man = P1 and the woman = P2), and a third person at the side of the table was described to the participants as an observer "Katie".

Each participant was shown the scene from their avatar's point of view (see Figures 2 and 3), and informed that some of the objects on the table were missing from their picture, but visible to their partner. Their joint task was to discover the missing objects from each person's point of view and mark them on the printed sheet of the scene provided. The objects that were hidden from each participants are marked with their ID in Figure 1.

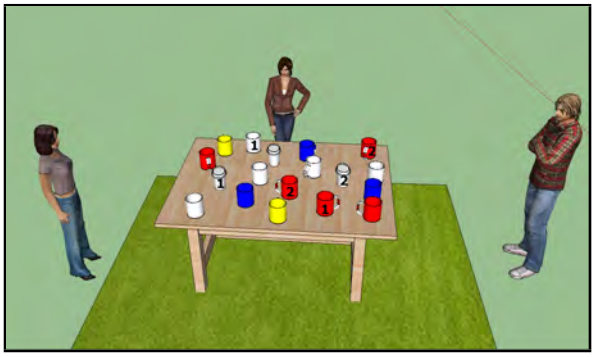

Figure 1: A virtual scene with two dialogue partners and an observer Katie. Objects labelled with a participant ID were removed in that person's view of the scene.

2.2 Procedure

Each participant was seated at their own computer and the participants were separated by a screen so that they could not see each other or each other's computer screens. They could only communicate using an online text based chat tool (Dialogue Experimental Toolkit, DiET, (Healey et al., 2003)).[1] The DiET chat tool resembles common online messaging applications, with a chat window in which participants view the unfolding dialogue and a typing window in which participants can type and correct their contributions before sending them to their interlocutor. The server records each key press and associated timing data.

In addition to the chat interface each participant saw a static image of the scene from their view, as shown in Figure 2, which shows the scene from P1's view and Figure 3, which shows the same scene from P2's view.

2.3 Participants

In the pilot study reported here, we have recorded two dialogues. Both dialogues were conducted in English but the native language of the first pair was Swedish while the second pair were native British English speakers. Participants were instructed that

[1] http://cogsci.eecs.qmul.ac.uk/diet/

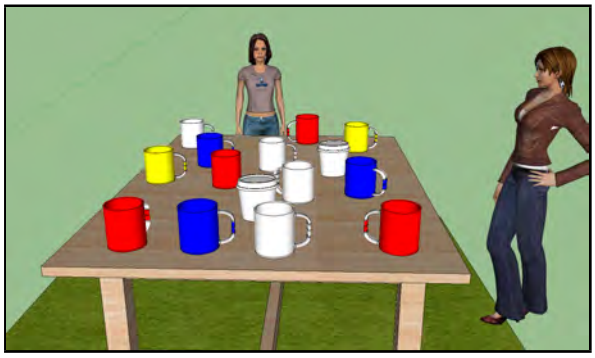

Figure 2: The table scene as seen by Participant 1.

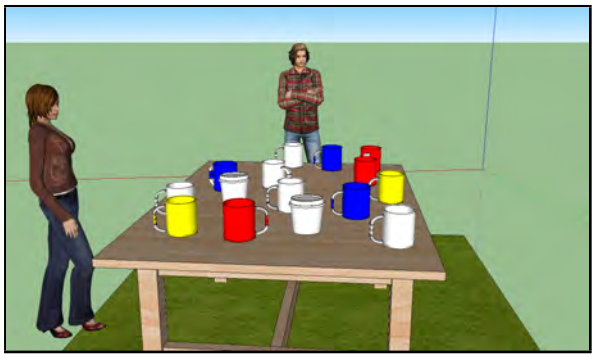

Figure 3: The table scene as seen by Participant 2.

they should chat to each other until they found the missing objects or for 30 minutes. The first dyad took approximately 30 minutes to find the objects and produced 157 turns in total. The second dyad (native English speakers) discussed the task for a little over an hour, during which they produced 441 turns. Following completion of the task participants were debriefed about the nature of the experiment.

2.4 Data annotation

The turns were annotated manually for the following features: (i) does a turn (T) contain a spatial description; (ii) the viewpoint of the FoR that the spatial description uses (P1, P2, Katie, object, extrinsic); a turn may contain several spatial descriptions with different FoR in which case all were marked; (iii) whether a turn contains a topological spatial description such as "near" or "at" which do not require a specification of FoR; and (iv) whether the FoR is explicitly referred to by the description, for example "on my left".

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
29 *P2:* white mug, white thing with funny top, red mug, yellow mug (the same as katies)

The example also shows that topological spatial descriptions can be used in two ways. They can feature in explicit definitions of FoR as "away" in T23, be independent as "right next to" in T24 and "closest to me" in T27 or sometimes they may be ambiguous between the two as " closer to Katie" in T25. In addition to referring to proximity, topological spatial descriptions also draw attention to a particular part of the scene that dialogue participants should focus on to locate the objects and to a particular FoR that has already been accommodated, in this case relative to Katie. Strictly speaking, this is not an explicit expression of a FoR but is used to add additional salience to it.

2.5 Dialogue Acts and entropy

We tagged both conversations with a dialogue act (DA) tagger trained on the NPS Chat Corpus (Forsyth and Martell, 2007) using utterance words as features as described in Chapter 6 of (Bird et al., 2009) but using Support Vector Machines rather than Naive Bayes classifier (F-score 0.83 tested on 10% held-out data). Out of 15 dialogue acts used, the most frequent classifications of turns in our corpus are (in decreasing frequency) Statement, Accept, yAnswer, ynQuestion and whQuestion and others. In parallel to DA tagging we also marked turns that introduced a change in the FoR assignment. Turns with no projective spatial description and hence no FoR annotation are marked as no-change. We process the dialogues by introducing a moving window of 5 turns and for each window we calculate the entropy of DA assignments and the entropy of FoR changes.

3 Results and Discussion

3.1 Overall usage of FoR

Table 1 summarises the number of turns that use each FoR in the dialogues. The data shows that the majority of FoR is assigned relative to dialogue participants (P1: 36%, P2: 27% and Speaker:

33%, Addressee: 29%, all values relative to the turns containing a spatial description). Extrinsic FoR is also quite common (25%) followed by the FoR relative to Katie (6%). In 10% of turns containing a spatial description the FoR could not be determined, most likely because a turn contained only a topological spatial description. Topological spatial descriptions are used in 18% of spatial turns. Note that since one turn may contain more than one spatial description, the number of turns of these does not add up to the total number of turns containing a spatial description.

Category	Turns	%
Turns in total	598	1.0000
Contains a spatial description	245	0.4097
FoR=P1	88	0.3592
FoR=P2	66	0.2694
FoR=speaker	81	0.3306
FoR=addressee	72	0.2939
FoR=Katie	15	0.0612
FoR=extrinsic	61	0.2490
FoR=unknown	26	0.1061
Topological description	44	0.1796

Table 1: Overall usage of FoR

In our data there are no uses of the intrinsic reference frame relative to the landmark object. This may be because the objects in this study were mugs and they are used as both target and landmark objects in descriptions. Although they may have identifiable fronts and backs and are hence able to set the orientation of the FoR, they are not salient enough to attract the assignment of FoR relative to the presence of the participants. This observation is orthogonal to the observation made in earlier work where the visual salience properties of the dialogue partners and the landmark object were reversed compared to this scene (Dobnik et al., 2014). Note, however, that we annotate descriptions such as "one directly in from of you" (D(ialogue) 1, T146) as relative FoR to P1, although this could also be analysed as an intrinsic FoR. We opt for the relative interpretation on the grounds that otherwise important information about which contextual features attract the assignment of FoR would be lost. In our system there is therefore no objectively intrinsic FoR but FoR assigned to different contextually present entities.

3.2 Local alignment of FoR

Figure 4 show the uses of FoR over the length of the entire D1 and the same length of utterances of D2. The plots show that although there is no global preference for a particular entity to assign the FoR one can observe local alignments of FoR that stretch over several turns which can be observed as lines made of red (P1) and green (P2) shapes. This supports the findings in earlier work (Watson et al., 2004; Dobnik et al., 2014) that participants tend to align to FoR over several turns.

Partial auto-correlations on each binary FoR variable in Figure 4 (P1, P2, Katie and Extrinsic) confirm this. Each correlates positively with itself ($p < 0.05$) at 1–3 turns lag, confirming that the use of a particular FoR makes reuse of that FoR more likely. Cross-correlations between the variables show no such pattern.

The graph also shows that the alignment is persistent to a different degree at different parts of both dialogues. For example, in D1 the participants align considerably in the first part of the dialogue up to turn 75, first relative to Katie, then to P2 and finally to P1. After approximately T115 both FoR relative to P1 and P2 appear to be used interchangeably in a threaded manner as well as the use of the extrinsic perspective. In D2 the situation is reversed. The participants thread the usage of the FoR in the first part of the dialogue but converge to segments with a single FoR shortly before T100 where they both prefer the extrinsic FoR and also FoR relative to P1. We will discuss these segments further in Section 3.4

Overall, the data show that the use of FoR is not random and that different patterns of FoR assignment and coordination are present at different segments of the dialogue. In order to understand how FoR is assigned we therefore have to examine these segments separately.

3.3 Explicitness of FoR

With an increase in (local) alignment, as discussed above, we might expect that there is less necessity for dialogue participants to describe the FoR overtly after local alignment has been established. Explicitness of FoR is therefore indicated in Figure 4: stars indicate that the FoR is described explicitly wheres triangles indicate that it is not. However, contrary to our expectation that the FoR would only be described explicitly at the beginning of a cluster of aligned FoR turns, it appears

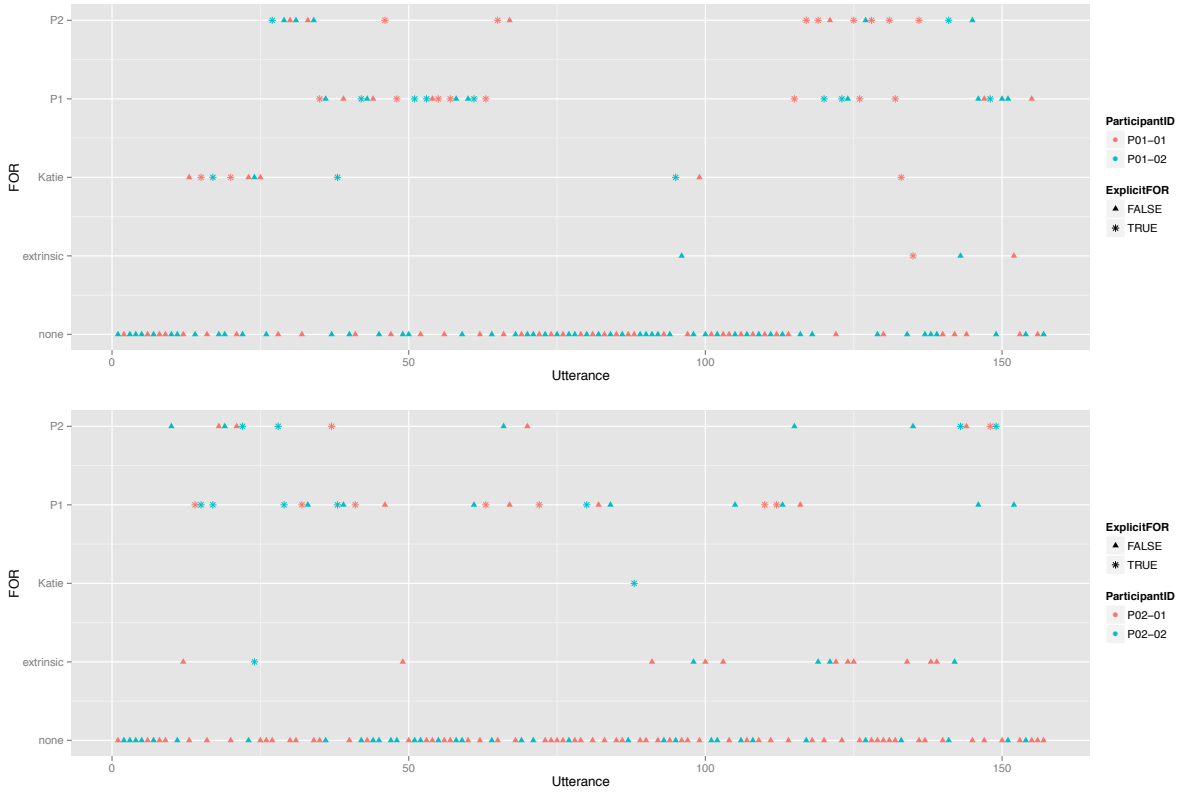

Figure 4: The assignment of FoR over the length of Dialogue 1 (top) and Dialogue 2 (bottom)

that the FoR is explicitly described every couple of utterances even if the participants align as in the first half of D1. This may be because participants are engaged in a task where the potential for referential ambiguity is high and precision is critical for successful completion of the task.

Note also that in D2 at around turn 100 there are clusters of turns where extrinsic FoR was used but this was not referred to explicitly. This is because participants in this dialogue previously agreed on a 2-dimensional coordinate system involving letters and numbers that they superimposed over the surface of the table. Referring to a region "A2" does not require stating "of the table" and hence a lack of explicitness in their FoRs.

3.4 Changing FoR

One of the main consequences of the local, and not global alignment of FoR, as shown in Figure 4 is that there are several shifts in FoR as the dialogue progresses. Below we outline some possible reasons for this, with illustrative examples taken from the dialogues. Due to the sparsity of data in our pilot study, these observations are necessarily qualitative, but they point the way towards some interesting future work.

(i) The scene is better describable from another perspective. Due to the nature of the task and the scene, it is not possible to generate a unique and successfully identifiable referring expression without leading to miscommunication. In D1 we can observe that the dialogue partners take neutral Katie's viewpoint over several turns. In fact, they explicitly negotiate that they should take this FoR: T13 "shall we take it from katies point of view?". However, in T25 P1 says "ok, is your white one closer to katie than the yellow and blue?" which prompts P2 to switch FoR to themselves "closest to me, from right to left:". The change appears to be initiated by the fact that the participants have just discovered a missing white mug but a precise reference is made ambiguous because of another white distractor mug nearby. P2 explicitly changes the FoR because a description can be made more precise from their perspective: from Katie's perspective both white mugs are arranged in a line at her front. Interestingly, in T35 P1 uses the same game strategy and switches the FoR to theirs saying "closest to me, from left to right red, blue, white, red" and the conversation continues using that FoR for a while, until turn 63. The example also shows that participants align in

terms of conversational games for the purposes of identifying the current object and that the nature of dialogue game also affects the assignment of FoR.

(ii) Current dialogue game. The nature of the task seems to naturally lead to a series of different dialogue games, from describing the whole scene to zooming in on a particular area when a potential mismatch is identified. In this case, since the scene in focus is only a part of the overall picture it is less likely that a an identifiable reference to a particular object will fail as there will be fewer distractors. As a result a single FoR can be used over a stretch of the conversation and participants are likely to align. There is less need for explicit perspective marking. See for example D1,T20-29 in the previous dialogue listing which corresponds to a cluster in Figure 4. Another cluster in Figure 4 starts at D1,T42 and is shown below. P2 identifies an empty space in their view which they assume is not empty for P1 and this becomes a region of focus. Since this region is more visually accessible to P1 and since they are information giver they opt for P1's FoR ("away from you" in T42 and T43). As shown in Figure 4 this is a dominant FoR for this stretch of dialogue.

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
44 *P1:* I have one thing there, a white funny top
relative-p1
45 *P2:* ok, i'll mark it.
46 *P1:* and the red one is slightly close to you
relative-p2, explicit, topological
47 is that right?
48 to my left from that red mug there is a yellow mug
relative-p1, explicit, topological
49 *P2:* hm...

Conversely, when looking for single objects that may be located anywhere on the entire table, for example, the speaker focuses on one object only that may be in a different part of the table than the one referred to in the previous utterance. There is no spatial continuum in the way the scene is processed and there may be several distracting objects that may lead to misunderstanding. Therefore, each description must be made more precise, both in the explicit definition of the FoR and through taking the perspective from which the reference is most identifiable. An example of this can be found towards the end of D1, before turn 115 (cf. Figure 4) where the participants decide to enumerate the mugs of each colour that they can see, P1 leads the enumeration and and describes the location of each object. However, the example also shows effects of continuity that is created by perceptual and discourse salience of objects, i.e. the way the scene is processed visually and the way it is described. In T117 "your left hand" is good landmark which attracts the FoR to P2 in the following spatial utterance in T119 but in T120 the FoR switches to P1 and in T121 back to P2. Turns T131-T136 show a similar object enumerating situation where FoR changes in every term and is also explicitly marked.

115 *P1:* my red ones are two in my first row (one of them close to katie)
relative-p1, explicit
116 *P2:*i mean there is a chance we both see a white that the other one is missing..
117 *P1:* one just next to your left hand
relative-p2, explicit
118 *P2:* yes
119 *P1:* and one on the third row from you slightly to your right
relative-p2, explicit
120 *P2:* is it directly behind the red mug on your left?
relative-p1, explicit
121 *P1:* no, much closer to you
relative-p2, topological
...
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
134 *P2:* yes, that's the same
135 *P1:* and the yellow are on between us to your far right
extrinsic
136 and one quite close to the corner on your left and katies right?
relative-p2, relative-katie, explicit

A switch between dialogue games tends to come with a switch of FoR. For example, in the following segment of D2, P1's FoR is selected initially to describe a row of cups closest to P1 and starting from their left to right (T14-T17). However, at T18 P1 initiates clarification. As P2 is information giver in this case the FoR is switched to theirs. Interestingly, the participants also switch the axis along which they enumerate objects (T21): starting at P2 and proceeding to P1, thus consistent from P2's perspective. At T26 a new clarification game is started and FoR changes to both P1 and P2, and at T32, after the participants exit both clarification games, P1 resumes the original game enumerating objects row-by-row and hence FoR is adjusted back to P1 accordingly.

14 *P1:* On my first row. I have from the left (your right): one red, handle turned to you but I can see it. A blue

cup next. Handle turned to my right. A white with handle turned to right. Then a red with handle turned to my left.
relative-p1, explicit

15 *P2:* first row = row nearest you?
relative-p1, explicit

16 *P1:* Yes.

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.

21 It looks almost like a diagonal line to me. From a red cup really close to you on your left, then a white, then the blue, then this missing red.
relative-p2

22 *P2:* blue with the handle to my left and white with the handle to my rigth/towards me a bit
relative-p2, explicit

...

26 *P1:* You know this white one you just mentioned. Is it a takeaway cup?

27 Because I think I know which cup that is but I don't see the handle.

28 *P2:* no, i was referring to the white handled cup to the right of the blue cup in the second row from you. its handle faces... south east from my perspective
relative-p1, relative-p2, explicit

29 the second row of cups from your end
relative-p1, explicit

...

32 *P1:* Shall we take my next row? Which is actually just a styrofoam cup. It's kinda marooned between the two rows.
relative-p1, explicit

(iii) Miscommunication and repair. We have already shown in the previous section that in line with (Mills and Healey, 2006), clarification triggers a change in FoR, with the explanation that clarification triggered a change of roles between the information giver and information receiver as well as introducing a different perceptual focus on the scene. However, during repair one would also expect that participants describe FoR explicitly more often. In the following example from D1, P1 is not sure about the location P2 is referring to. In T148 P2 explicitly describes the cup that can be found at that location using double specification of FoR. Information giver is thus providing more information that necessary to ensure precision of reference.

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

149 *P2:* ok, from my perspective, there is at least a cup-sized gap between the edge of the table and the yellow cup
relative-p2, explicit

150 *P1:* Yes, I can say that too

As we have already seen, participants also use other strategies to reduce miscommunication, for example by enumerating objects that can be seen at any time of the conversation. From D1:

69 *P1:* so now I have 17 including the ones I've marked, how many do you have?

...

100 *P2:* so then again, it looks like we see everything we can

101 *P1:* yes, you still just got 17?

102 *P2:* yes

(iv) Explicit strategies Participants also devise strategies for processing the scene to find the missing objects. In (D1, T13) participants agree to use Katie's perspective as a reference. In (D2, T51 and following) they negotiate to split the table into a grid of 16 sub-areas where they label the columns with letters and rows with numbers. They negotiate the coordinates so that column labels A-D go from left to right and row numbers go from top to bottom relative to P2's view of the table. Hence, although they devise an extrinsic FoR with areas that they can refer to with coordinates they are forced to combine it with a FoR relative to P2 and therefore they create a more complicated system that involves two viewpoints. Interestingly, P1 clearly marked the axis labels on their printed sheet of the scene, which P2 did not, probably because the coordinate system was more difficult from P1's viewpoint. The negotiation of the coordinate system requires a lot of effort and involves referring to objects in the scene when negotiating where to start the lettering and numbering and how to place the lines for the grid. The participants finish the negotiation in T165, 114 turns later. However, although participants of D1 and D2 both negotiate on some reference perspective they do not use it exclusively as shown in Figure 4. One hypothesis that follows from these observations is that participants would use the reference (combining relative-katie and extrinsic) FoR in turns that involve greater information precision, that is those under repair as demonstrated in T119 of D2. Here the participants are negotiating where to draw the lines that would delimit different areas of the grid.

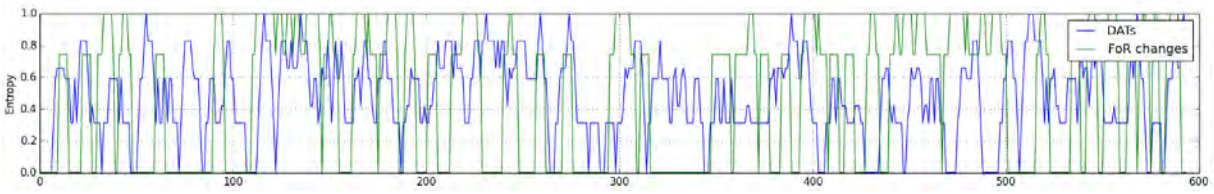

Figure 5: The entropy of DATs and FoR assignment calculated per each moving window of 5 turns. Both dialogues are combined into a single sequence and D2 starts in T158. Entropies were normalised by maximum observable entropy in the dataset.

105 *P2:* so, 2 could be in line with a can you see a blue cup, that is behind the A1 red cup?
relative-p1
...
110 *P1:* Yes. For me the blue cup is in front of the red cup. But yes.
relative-?, explicit
111 It has a handle that perhaps you can't see.
112 Since it is pointing south east for me.
relative-p1, explicit
113 *P2:* what do you mean by "in front of"
114 *P1:* Hmm
115 *P2:* closer to me or closer to you?
116 *P1:* Closer to you
relative-p1, explicit
117 *P2:* ok yep
118 *P1:* Okay
119 *P2:* i cna just see the handle almost pointing to A1
extrinsic

The excerpt shows that FoR itself may be open for repair. In T110 P1 corrects P2 in T105. P2's description contains FoR relative to P1, but P1 mistakenly takes a FoR relative to the landmark "the red cup" (i.e. intrinsic FoR). It is likely that this is because the red cup is very salient for P1 and allows P1 to project their orientation to the cup (the orientation of the FoR is not set by its handle). This is the only example where intrinsic FoR is used in the corpus and since it is repaired we do not count it as such. In T116 P1 comes to an agreement with P2.

3.5 FoR assignment over conversation

The preceding analysis of dialogue shows that FoR assignment is dependent on the type of communicative act or conversational game that participants are engaged in. The changes in perspective are dependent on factors that are involved in that particular game, for example the structure and other perceptual properties of the scene, the participants' focusing on the scene, their conversational role and availability of knowledge, the accommodated information so far, etc. To test whether the FoR assignment could be predicted only from the general dialogue structure we compared the entropy of the Dialogue Act tags with the entropy of the changes in FoR. As shown in Figure 5 there are subsections of the dialogue where the variability of DAs coincides with the variability of the FoR (i.e. where the entropy is high) but this is not a global pattern (Spearman's correlation rho = -0.36, $p = 0.383$). There are also no significant cross-correlations between the variables at different time lags. In conclusion, at least from our pilot data, we cannot predict the FoR from the general structure of conversational games at the level of DAs. This also means that there is no global alignment of FoR assignment and that this is shaped by individual perceptual and discourse factors that are part of the game.

4 Conclusions and future work

We have described data from a pilot study which shows how dialogue participants negotiate FoR over several turns and what strategies they use. The data support hypothesis (i) that there is no general preference of FoR in dialogue but rather this is related to the communicative acts of a particular dialogue game. Examining more dialogues would allow us to design an ontology of such games with their associated strategies which could be modelled computationally. Hypothesis (ii) that participants align over the entire dialogue, is not supported. Rather, we see evidence for local alignment. Hypothesis (iii) is also not supported: while misunderstanding may be associated with the use of different FoRs, there are also other dialogue games where this is the case, for example locating unconnected objects over the entire scene.

We are currently extending our corpus to more dialogues which will allow us more reliable quantitative analyses. In particular we are interested in considering additional perceptual and discourse features (rather than just DAs) to allow us to automatically identify dialogue games with particular assignments of FoR and therefore apply the model computationally.

References

Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural language processing with Python*. O'Reilly.

Holly Branigan, Martin Pickering, and Alexandra Cleland. 2000. Syntactic co-ordination in dialogue. *Cognition*, 75:13–25.

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

Herbert H. Clark. 1996. *Using language*. Cambridge University Press, Cambridge.

Simon Dobnik, John D. Kelleher, and Christos Koniaris. 2014. Priming and alignment of frame of reference in situated conversation. In Verena Rieser and Philippe Muller, editors, *Proceedings of DialWatt - Semdial 2014: The 18th Workshop on the Semantics and Pragmatics of Dialogue*, pages 43–52, Edinburgh, 1–3 September.

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

Eric N. Forsyth and Craig H. Martell. 2007. Lexical and discourse analysis of online chat dialog. In *Proceedings of the First IEEE International Conference on Semantic Computing (ICSC 2007)*, pages 19–26. IEEE.

Simon Garrod and Anne Anderson. 1987. Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, 27:181–218.

Patrick G. T. Healey, Matthew Purver, James King, Jonathan Ginzburg, and Greg J. Mills. 2003. Experimenting with clarification in dialogue. In *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*, Boston, MA, Aug.

Patrick G. T. Healey, Matthew Purver, and Christine Howes. 2014. Divergence in dialogue. *PLoS ONE*, 9(6):e98598, June.

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

Boaz Keysar. 2007. Communication and miscommunication: The role of egocentric processes. *Intercultural Pragmatics*, 4(1):71–84.

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

Stephen C. Levinson. 2003. *Space in language and cognition: explorations in cognitive diversity*. Cambridge University Press, Cambridge.

Xiaoou Li, Laura A. Carlson, Weimin Mou, Mark R. Williams, and Jared 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.

Didier Maillat. 2003. *The semantics and pragmatics of directionals: a case study in English and French*. Ph.D. thesis, University of Oxford: Committee for Comparative Philology and General Linguistics, Oxford, United Kingdom, May.

Gregory Mills and Patrick G. T. Healey. 2006. 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, September.

Martin Pickering and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27:169–226.

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

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

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

Matthew E Watson, Martin J Pickering, and Holly P Branigan. 2004. Alignment of reference frames in dialogue. In *Proceedings of the 26th annual conference of the Cognitive Science Society*, pages 2353–2358. Lawrence Erlbaum Mahwah, NJ.