Local Alignment of Frame of Reference Assignment in English and Swedish Dialogue

Simon Dobnik[1][0000−0002−4019−7966], John D. Kelleher[2][0000−0001−6462−3248], and Christine Howes[1][0000−0002−2794−1586]

[1] Centre for Linguistic Theory and Studies in Probability (CLASP), Department of Philosophy, Linguistics and Theory of Science (FLoV), University of Gothenburg, Box 200, SE 405 30 Gothenburg, Sweden
{simon.dobnik,christine.howes}@gu.se
[2] ADAPT Research Centre
Technological University Dublin (TU Dublin)
191 Park House, North Circular Road, Dublin, D07EWV4, Ireland
john.d.kelleher@TUDublin.ie

Abstract. In this paper we examine how people assign, interpret, negotiate and repair the frame of reference (FoR) in online text-based dialogues discussing spatial scenes in English and Swedish. We describe our corpus and data collection which involves a coordination experiment in which dyadic dialogue participants have to identify differences in their picture of a visual scene. As their perspectives of the scene are different, they must coordinate their FoRs in order to complete the task. Results show that participants do not align on a global FoR, but tend to align locally, for sub-portions (or particular conversational games) in the dialogue. This has implications for how dialogue systems should approach problems of FoR assignment – and what strategies for clarification they should implement.

Keywords: spatial descriptions · frame of reference · dialogue · alignment · computational models

1 Introduction

There are two main challenges surrounding the computational modelling of spatial perspective taking which are reflected in the grounding of the origin of the frame of reference (FoR). First, there are several ways in which the viewpoint may be assigned and hence multiple frames of reference may be applicable [19]. Second, the viewpoint may not be overtly specified and must be recovered from linguistic or perceptual context. Such underspecification may lead to situations where conversational partners fail to adopt the same FoR leading to miscommunication [13].[3] In this paper we examine how participants assign, interpret, negotiate and repair their spatial representations in dyadic text dialogues when

[3] This also presents serious challenges for learning spatial language by robots from human descriptions [4,26].

they perceive a scene from different perspectives. In particular we are interested in how they select FoRs, how they identify if an FoR misalignment has occurred, and what strategies they use to realign or clarify FoR misalignment. The paper is structured as follows: in Section 2 we review previous work on FoR in spatial language, and approaches related to FoR underspecification and alignment in dialogue, in Section 3 we present our hypotheses, in Section 4 we describe our data and results and finally in Section 5 we present conclusions and directions for further work on computational modelling of FoR in dialogue.

2 Frames of Reference

Frames of reference have for a long time been recognised as important phenomena of study in both spatial cognition and spatial language research. In the context of spatial language the standard modern interpretation of a FoR is a set of six half-line axes anchored at an origin (defined as a point on the landmark object) [11], though there is a diversity in FoR systems both across and within languages.

Following [18] we distinguish between three FoR as follows:

relative: locates a target relative to a landmark from a particular viewpoint e.g. "the blue cup to my left of the red cup" relative to the speaker;
intrinsic: locates a target relative to a landmark e.g. "the blue cup to the left of the red cup" relative to the orientation of the red cup;
extrinsic: locates a target relative to a landmark "the blue cup to the north of the red cup".

The fact that the intended FoR is often implicit in a spatial description can lead to ambiguity in contexts where different FoRs assign different canonical directions to a directional spatial preposition, which can lead to multiple FoRs being activated [1]. Where multiple FoRs are activated (e.g. between "right" intrinsically aligned and "right" aligned with a relative FoR), they compete, and the resulting spatial template is a weighted merging of the competing spatial templates [2]. Furthermore, for prepositions (canonically) aligned with the vertical axis ("above" and "below") there is a strong weighting towards an extrinsic (gravitationally aligned) FoR [2]. However, for spatial prepositions aligned with the horizontal axis ("behind", "right", "left", *etc.*) where a landmark has an intrinsic FoR there is a weighting towards the intrinsic FoR interpretation [14]. This preference towards intrinsic FoR has been demonstrated in object selection tasks, both when the object array was perceptually available or retrieved from memory [20]. A number of computational models have been developed to accommodate FoR underspecification in locative descriptions [15,25]. However, these studies and models have focused on the interpretation of a locative description in a one-off setting, as opposed to within the context of an ongoing dialogue.

Much of the work on FoR selection in dialogue is based on route description tasks. An early example [17] argues that individuals have a cognitive style which gives them a preference towards using one FoR: some individuals consistently use an intrinsic FoR whereas others are more likely to use an ego-centric (or relative)

FoR. Furthermore, [17] argues that for extended route descriptions people adopt a single FoR and use it consistently, because consistency promotes coherence and comprehension.

More generally, research in dialogue shows that conversational participants align representations at several levels of representations [22], and a number of studies show that this is also the case for FoR. In experiments with confederate-priming [29] show that speakers tend to use the same FoR that the confederate has just used regardless of whether the same or a different spatial description was used. In terms of priming versus preference on FoR selection [12] show effects of priming and a decreased preference for a FoR for a particular pair of objects. However, work on less constrained dialogues finds that speakers frequently switched FoR, indicating that although FoR "is needed locally to define a spatial relation it is not needed throughout to ensure coherence" [27, pg.389].

More recent research on spatial dialogue has studied locative descriptions. For example, [28] examined communicative success in a task where one participant had to describe how to arrange and orient a set of objects in a dolls house and their partner had to furnish it based on these descriptions.[4] They found that a number of factors affected communication success, including: (a) the functional features of the spatial arrangement of the furniture (e.g., did the target orientation of a chair relative to a table align with expected/canonical arrangement of chairs and tables), (b) previous task experience, and (c) dialogue features such as description length and the inclusion of orientation information.

There is also evidence that in conversations where speakers have to describe locations of objects in a complex display they adopted the perspective of the person who did not know which object was being selected indicating that communicative role within a conversation and the participant's knowledge of the information both affect FoR selection [24]. Furthermore, FoR selection is not dictated by minimising an individual's processing demands but that participants resolved referential underspecification in terms of their partner's perspective and that this effect was consistent even when the presence of the partner had to be assumed [7]. More recently, [8] describe a human-human spatial dialogue experiment where one participant learnt the layout of an array of objects and then described this from memory to a partner whose task was to reconstruct the layout. The main results were that there was not a dominant strategy in FoR usage, with speakers sensitive to a range of factors (in this instance awareness of the partner's viewpoint, and representational cues, such as viewpoint alignment with the symmetry of the array). FoR selection appears to be flexible and dynamic and sensitive to a range of factors, including social and perceptual factors. In particular, dialogue partners follow the *principle of least collaborative effort* [3] with speakers exhibiting a willingness to adopt their partner's relative FoR in contexts where this would reduce the overall cognitive burden of the dialogue. However, these studies did not focus on the role of FoR priming or whether participants reduced cognitive load by adopting a globally consistent FoR throughout a dialogue.

[4] The 'Dolldialogue' corpus is available online at `www.dolldialogue.space`

In one of the few studies of the effects of priming in terms of alignment over longer stretches of structured dialogue (between a system and a user) and also over conversational role changes [6], a high degree of alignment to linguistic priming was found. In most cases the participant assumed the same FoR that the system used in the first game of the experiment. However, in cases where alignment was not observed, there was a preference for the intrinsic FoR set by the properties of the scene at the expense of the FoR set by the conversational partner, despite the fact that the user is presented with a sequences of descriptions and does not have the ability to interact with their conversational partner with additional utterances to resolve potential ambiguities in other ways.

Results from a pilot study [5] which recorded and annotated in detail two dyadic dialogues in English using the task described below in Section 4 suggest that there was no general preference of FoR in dialogue but the choice is related to the communicative acts of particular dialogue or conversational games (a sequence of dialogue moves centred towards a particular goal [16,23]) at specific points in the dialogue. There was also evidence that participants aligned their FoR locally over a sequence of turns, but not globally; at points of misunderstanding it may be prudent to shift FoR in order to get the conversation back on track. The pilot study [5] isolates 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 identifying individual objects scattered over the entire scene. It follows that alignment is consistently used as a strategy but there are other factors that trigger changes in FoR. For example, assignment of FoR is also driven by strategies for *mutual understanding* and *resolution of misunderstanding*.

3 Hypotheses

The preceding discussion shows that there are several factors which can influence conversational participants choice of FoR, and that FoR choices can compete with each other. Therefore, a natural continuation is to investigate these choices in a free dialogue between human conversational participants. In particular, we are interested in (i) what the possible choices of FoR assignment in a particular discourse (task) and perceptual (arrangement of the scene) are; (ii) whether (different) participants always behave in the same way in such scenes; and (iii) what are the strategies for alignment. The *interactive alignment model* [22] would suggest that interlocutors would converge on a single FoR. However, previous research has shown that interlocutors diverge syntactically [9] and that in semantic coordination clarification requests (taken to be an indicator of miscommunication) decrease convergence [21]. Rather than alignment, description types are driven by mutual understanding and strategies for resolution of misunderstanding: identification of misalignment and strategies for getting back on track.

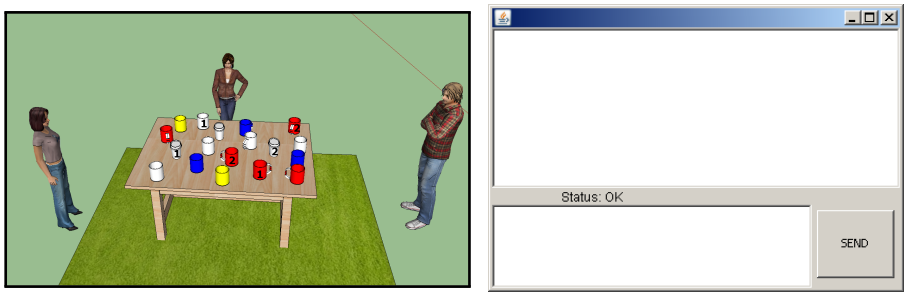
Following these observations we form the following hypotheses:

(1) There is no baseline preference for a specific FoR in free dialogue.
(2) Participants will align on spatial descriptions over the course of the dialogue.
(3) Participants will only use explicit descriptions of the FoR at the beginning, before they have aligned.
(4) Sequences of misunderstanding will prompt the use of different FoRs.

4 Situated Dialogue and FoR: the Cups Dataset

4.1 Method

Task Using 3D modelling software we designed a virtual scene depicting a table with several mugs of different colours and shapes placed on it as shown in Figure 1a. The scene also includes three people standing at different sides of the table. The people at the opposite sides of the table are the avatars of the participants (the man = P1 and the woman = P2). There is also a third person at the side of the table who was described to the participants as an observer "Katie" who is not taking part in the conversation. As shown in [6] participants prefer to assign FoR to a neutral landmark that is not one of the conversational participants. In this experiment Katie fulfils this role.

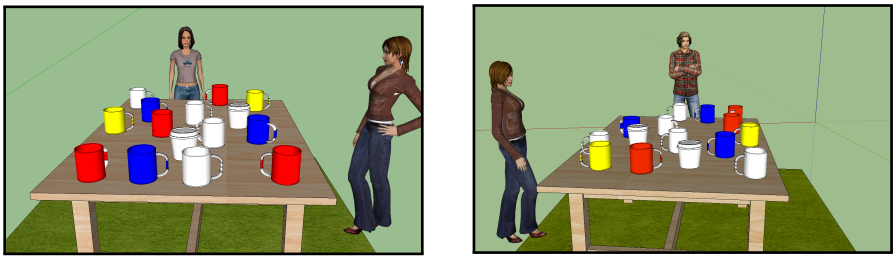


(a) (b)

Fig. 1: (a) 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. (b) The DiET chat tool.

In order to ensure that participants engaged in longer dialogue involving spatial descriptions and in order to create the ambiguity involving spatial reference we designed the task as a "spot-the-difference" game. Each participant was shown the scene from one of the avatar's points of view (see Figure 2), and was informed that some of the objects on the table were missing from their picture, but these objects were visible to their partner. Equally they are able to see some objects that are missing from their partner's view. Their joint task was to discover the missing objects by interacting through conversation. The objects that were hidden from each participants are marked with their ID in Figure 1.

The task proved challenging as it requires fine dialogic negotiation and reasoning since there are identical (*distractor*) objects in close proximity to the missing objects and therefore the dialogues exhibit rich linguistic data going beyond modelling of FoR but also generation of referring expressions (mugs of different colours and at different locations), anaphora resolution and conversational dynamics such as reasoning in dialogue, incrementality, clarification and repair.



(a) (b)

Fig. 2: The scene from Figure 1a as seen by (a) Participant 1 & (b) Participant 2.

The task ensures all possible FoR assignments: (i) most of the mugs on the table have handles which means that they have orientation and can assign intrinsic FoR – participants interpret the handle as the back of the mug; (ii) the surface of the table grounds the extrinsic FoR; and (iii) the conversation participants and the observer Katie can assign relative FoR.

Participants The participants were 8 native Swedish speakers, taken from the student population at Gothenburg University, paired into dyads.

Procedure Each participant was seated at their own computer and separated from their dialogue partner so that they could not see each other or each other's screens. Communication was through an online text based chat tool (Dialogue Experimental Toolkit, DiET, [10]), which records each key press and associated timing data. Participants saw the chat tool as shown in Figure 1 and were instructed that they should chat to each other until they found the missing objects or for at least 30 minutes. They were also asked to mark the missing objects and any notes on the printed image of their view of the scene. Following completion of the task participants were debriefed about the nature of the experiment.

Data Annotation Following [5], we annotated the dialogue data at the level of individual turns with the annotation categories shown in Table 1. A turn may contain several spatial descriptions in which case all of the categories are

annotated. The same is true if a spatial description is ambiguous. However this is very uncommon as in nearly all cases annotators were able to resolve the intended referents of the objects from the visual scene and the surrounding dialogue.

Table 1: Annotation scheme for manual annotation of the corpus

Tag	Value		Explanation
`is-spatial`	**sp**	y/n	For all turns: does this turn contain a spatial description
`viewpoint`	**P1/P2/katie/ obj/intr/extr**		Where `is-spatial=y`: what viewpoint does the FoR use: P1, P2, Katie, object, intrinsic, extrinsic?
`topological`	**top**	y/n	Where `is-spatial=y`: does the turn contains a topological spatial description such as "near" or "at"?
`explicitness`	**exp**	y/n	Where `is-spatial=y`: is the FoR explicitly referred to, e.g. "on my left"?
`repair`	**rep**	y/n	The utterance is a repair.
`acknowledgement`	**ack**	y/n	The utterance is an acknowledgement.

The annotation requires resolution of the reference of the expressions and is considerably cognitively demanding. It also requires expert annotators who understand the model of the FoR we use (see Section 2). Two of the authors annotated the first 100 turns of D1 and first 105 turns of D2 (for explanation see below) for which we obtained high inter-annotator agreement (Cohen's kappa $\kappa = 0.8121$).

Figure 3 shows example annotations of dialogues between participant pair D1 (English) and participant pair D7 (Swedish).

4.2 Results

In this paper we describe an extension of the corpus captured using the same task in Swedish with 4 pairs of Swedish native speakers. We investigate if the findings from the English study [5] hold cross-linguistically, when resources for resolving misunderstandings may not be the same across languages. Table 2 lists a general comparison of the corpora in terms of participant's native language, duration of the conversation and the number of turns produced. Overall, the English corpus contains 598 turns, the Swedish corpus is 794 turns and the combined corpus contains 1392 turns.

FoR Alignment and Change In order to assess whether there is a baseline preference for a specific FoR in dialogue (hypothesis 1) we examine the overall spread of FoR assignments used in our data. Table 3 summarises the assignment of FoR. In support of hypothesis 1, there does not seem to be any baseline preference for a specific FoR, with speaker/addressee/external(Katie+extrinsic)

English D1

20 *P1:* from her right I see yell, white, blue red **sp, katie, exp**
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 **sp, katie, exp, top**
24 *P2:* because i see a normal mug too, right next to the yellow one, on the left **sp, katie, top**
25 *P1:* ok, is your white one closer to katie than the yellow and blue? **sp, katie, top**
26 *P2:* yes **ack**
27 closest to me, from right to left: **sp, p2, top**
28 *P1:* ok, got it **ack**
29 *P2:* white mug, white thing with funny top, red mug, yellow mug (the same as katies)

Swedish D7

72 *P2:* från katies sida av bordet **sp, p2, katie, exp**
73 *P2:*en blå mugg
74 *P1:* en blå mugg nästan mitt framför henne? **sp, p2, katie, rep**
75 *P2:* en vit mugg, som nog står något närmare dig **sp, top, p1, p2, katie**
76 *P2:* ja, jag skulle vilja säga att den är något närmare dig än mitten **sp, top, p1, extr, ack**
77 *P2:* men oavsett
78 *P2:* den är blå
79 *P1:* den är blå, ja **ack**
80 *P2:* sen står som sagt en vit mugg **sp, p2, katie**
81 *P1:* mhm **ack**
82 *P2:* sedan något större mellanrum åt höger till den sista som är röd **sp, p2, katie**
83 *P1:* har den röda ett hantag? **rep**
84 *P2:* jag kan inte riktigt se det handtaget ordentligt, men det verkar stå mot dig **sp, p1**
85 *P1:* jag ser inget handtag **rep**
86 *P2:* nähä

Fig. 3: Example annotation

descriptions occurring approximately equally in the dialogues. In the Swedish dialogues there are fewer spatial descriptions but FoR is indicated more often as all percentages are higher. Although there is too little data to draw any firm conclusions, it appears that in the Swedish data speakers may be more precise in overtly specifying the FoR. However, the rankings between the assignments are similar: P1 > P2 and Speaker > Addressee. The first observation may be explained by the fact that P1 has a better view of the scene and a more focused line of objects (Figure 2). The second observation shows that speakers are egocentric which is contrary to the observation of [24] who shows that information givers adapt to information receivers. The FoR relative to Katie and the extrinsic FoR are neutral perspectives (neither of the conversational partners) and the figures suggest that their preference may be independent of language and instead dependent on other factors such as personal preference. For example, in the two English dialogues we observe that one pair prefers the FoR relative to Katie and the other the extrinsic FoR. The percentages in Table 3 do not add up to 100 because in some turns there are several spatial descriptions using different

Table 2: Overview of data

Dialogue	*Language*	*Native*	*Duration (min)*	*Length (turns)*
#1	English	Swedish	≈30	157
#2	English	British	≈60	441
#4	Swedish	Swedish	≈30	75
#5	Swedish	Swedish	≈60	163
#6	Swedish	Swedish	≈60	248
#7	Swedish	Swedish	≈60	308

FoRs and therefore there is over-specification. This is related to ensuring greater precision of reference.

Table 3: FoR assignment in English and Swedish dataset

Category	*English*		*Swedish*	
	Turns	*%*	*Turns*	*%*
Contains a spatial desc.	245	40.97	273	34.38
FoR=P1	88	35.92	122	44.69
FoR=P2	66	26.94	83	30.40
FoR=speaker	81	33.06	107	39.19
FoR=addressee	72	29.39	98	35.90
FoR=Katie	15	6.12	52	19.05
FoR=extrinsic	61	24.90	38	13.92
Topological description	44	17.96	52	19.05
Total turns	598		794	

Alignment Given that there is no baseline FoR preference, we now turn to the issue of whether dialogue participants align on one type of FoR (which may be different for each dialogue pair) as the dialogue progresses (hypothesis 2). To assess this hypothesis, we examine the distribution of the FoR assignment of all the dialogues (illustrated by two example dialogues in Figure 4). In terms of alignment the Swedish dialogues show a very similar trend to that found in English pilot study. Participants tend to align on FoR over several turns but the alignment is local, not global (contra hypothesis 2). This is shown in the graphs by clusters of one type of FoR for a period of turns before switching to a different FoR and a new cluster. Of course, we do not expect the same FoRs to occur at the same points in each dialogue as there is no fixed order in which participants must complete the task, but the general clustering pattern strongly

suggests that once an FoR is used, it continues to be used for a portion of the dialogue, before there is a general switch to another FoR.

Correlation tests support this impression, with significant partial auto-correlations on each binary FoR variable: Each variable P1, P2, Katie and Extrinsic correlates positively with itself ($p < 0.05$) at 1–3 (English) and 1–2 (Swedish) turns lag. The use of a particular FoR makes a reuse of that FoR more likely in the immediately following turns. However, this is less so in the Swedish dialogues where the change may come more often. This supports the observation that Swedish speakers seem to use overt specification of the FoR more often. There are no significant cross-correlations between the variables in the English data (the use of one FoR does not predict the subsequent use of another one) but there are significant cross-correlations between P2 and Katie in the Swedish data. Examining the graphs in Figure 4 it can be seen that there are parts of the conversation where there is alignment of the FoR but also that there are parts where FoR frequently changes. Qualitative assessment of these sequences suggests that FoR assignment is linked to particular dialogue games or communicative strategies that participants are using in that part of the dialogue. We return to the discussion of this question in Section 4.2.

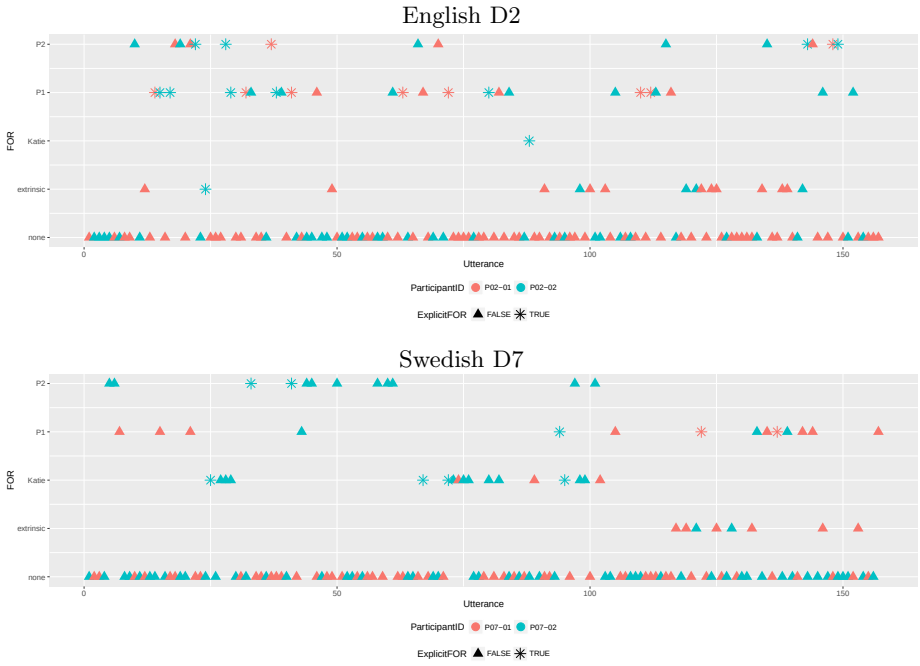
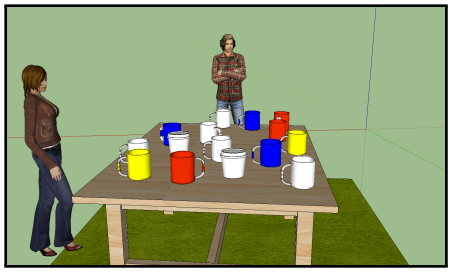


Fig. 4: The FoR assignment over the first 157 turns of the D2 (English) and D7 (Swedish) dialogues. The number of turns is chosen to ensure a comparison with D1 (the shortest dialogue in the data) reported in [5].

Explicitness As we did not find support for global alignment of FoR, only local alignment, we now turn to an amended version of hypothesis 3, that participants will only use explicit descriptions of the FoR at the beginning *of a local alignment cluster*: there should be no need to describe FoR overtly when local alignment is established and therefore explicit definitions of FoR should only be at the beginning of the aligned sequences. However, examination of the data (see Figure 4) shows that even a local version of hypothesis 3 does not hold for either English or Swedish dialogues since FoR is referred to explicitly every couple of turns. This may be related to the task in which there is high potential for referential ambiguity and precision is critical for success, or because switching FoR is common (as shown in Figure 4), so even when it is not changed this may need explicit marking to avoid potential misunderstandings.

Qualitative Analysis of FoR Assignment Informed by the graphs shown in Figure 4 we qualitatively examined the dialogues in detail at points which show changes in FoR, which we now discuss in this section, under a number of subheadings that seem to trigger such changes.

Visual properties of the scene The visual properties of the scene from each person's perspective may trigger a change of FoR. For example, Figure 5 repeats a section of D1 dialogue given in Figure 3. Participants start and continue using Katie's perspective. P1 cannot see the white mug farthest to the left from P2's perspective. In the preceding conversation they are evaluating expressions in the lateral dimension from Katie's view and hence the two white mugs are linguistically less distinguishable using this strategy since they are arranged front-back and the only available description is non-directional "close". Changing to the FoR of P2 allows the use of a further description of the visual scene that they are attending to with a more precise reference to several objects. Changing the FoR therefore aids the resolution of referential ambiguity.



P2

English D1

25 *P1:* ok, is your white one closer to katie than the yellow and blue? **sp, katie, top**
26 *P2:* yes
27 closest to me, from right to left: **sp, p2, top**
28 *P1:* ok, got it **ack**
29 *P2:* white mug, white thing with funny top, red mug, yellow mug (the same as katies)

Fig. 5: Visual properties of the scene and FoR change

Dialogue game Another factor may be the (sub-)task or dialogue game that conversational participants are engaged in at each point in the dialogue. In this case, participants may use different strategies to perceptually attend to and discuss the scene. Figure 6 shows two different strategies that are used. In dialogue D1 between turns 42–48 the conversational participants are focusing on a particular part of the scene where (i) there is a spatial continuum between the objects and hence previously located objects can serve as later landmarks or be referred back to with definite descriptions; (ii) there are fewer distractors and therefore higher FoR alignment and less explicit FoR marking (although this is not shown in this short example). Consequently, the same perspective is used over a stretch of the conversation.

The focused region is chosen by P2 who notices that in their view there is an empty space close to P1. P2 requests that P1 take the role of information-giver/describer of the scene while P2 follows the descriptions until they discover inconsistencies. We observe a tendency at this point and other points in the data that the perspective of the person who is providing information is taken as the region is visually more accessible to them. They are also expected to give a complete and consistent account of what they see, while the other participant validates the description in their own view. Under the principle of least collaborative effort, each participant takes on one cognitively costly task (P1, the description task; P2, the non ego-centric perspective taking), thus splitting the load between them. This avoids the situation where the speaker takes on all the cognitive load by using the FoR of the participant who checks the scene. This observation is consistent with the observation by [24] that communication role affects the assignment of the FoR. However, contrary to his findings that information givers adapt to information receivers who have less information, in our task the FoR aligns with information givers who have more information.

English D1

42 *P2:* there is an empty space on the table on the second row away from you **sp, p1, exp, top**
43 between the red and white mug (from left to right) **sp, p1**
...
48 to my left from that red mug there is a yellow mug **sp, p1, exp, top**

English D2

131 *P1:* and the blue ones are one on the second row from you, to the right from you **sp, p2, exp**
132 one slightly to my left **sp, p1, exp**
133 and one in front of katie in the first row **sp, katie, exp**

Fig. 6: Task and FoR change

On the other hand, in dialogue D2 between turns 131–133 the conversational participants are scanning the scene to locate individual objects that are blue (see Figure 6). In this task (i) there is no spatial continuum between the objects they are referring to; (ii) there are several potential referents and distractors for

each description. As a result descriptions have to be more precise and there is less alignment of FoR and more explicit FoR marking. Each referring expression is made with a different FoR. The FoR is chosen from which a good referring expression can be generated.

Resolving (potential) misunderstanding Finally, let us consider the assignment of FoR in **clarification and repair**, which we hypothesised would lead to FoR change (hypothesis 4). Although our annotations are not (yet) rich enough to quantify whether and how sequences of misunderstanding prompt the use of different FoRs (hypothesis 4), examination of the data offers some pointers in this regard.

Figure 7 from dialogue D1 between turns 14–32 shows that the distinction between information givers and information receivers [24] is difficult to maintain in dialogues that transition to clarification as it is not clear when their roles change or which participant in the clarification dialogue has more or less information at any of these points. In turn 14 P1 is the information giver but in turn 17 they are the information receiver and there is no change of the FoR. In turn 18 P1 asks a clarification question and is therefore both information giver and receiver. Furthermore, if the participant that starts a clarification dialogue is information receiver then according to Schober the FoR should align to them but in this case P1 changes the FoR to the FoR relative to P2.

The FoR change can be successfully explained if one adopts the task-based assignment of FoR that we described above. In this scenario participants take one of the two roles: (i) the describer who has visual focus on the scene, clarifies it and attracts the FoR and (ii) the follower who checks the descriptions until an inconsistency is detected. A clarification request triggers a change of roles and a different perceptual focus on the scene. During turns 14–17 P1 is the describer and P2 is the follower. In turn 18 P2 is the describer and P1 is the follower. In turn 28 P2 transitions the visual focus and the assignment of the FoR back to P1 and therefore initiates a change of roles back to those before the first clarification request. Hence, the changes in FoR are task dependent and clarification requests initiate a new dialogue game and therefore a change of the strategy for FoR assignment. The clarification game is embedded within the original game and after it is completed the FoR also transitions back with it, as in turn 32.

In the Swedish dialogue D5 between turns 36–43 in Figure 7 a clarification about the FoR is explicitly raised because the roles of the describer and the follower are not clear. In turn 36 P1 is the describer and P2 the follower. However, in turn 39 P2 makes a description, followed by another description in turn 40 but this time without an overt specification of the FoR. As P2 is now taking on the role of the describer P1 appears to be confused about the FoR P1 is using and they raise a clarification request about the perspective that they are assuming. The example is a clear demonstration that FoR may be associated with participant roles and that deviation from these conversational strategies requires a resolution of misunderstanding.

Misunderstanding also occurs because descriptions are underspecified and describers make errors. Clarification and repair dialogue strategies – including

English D2

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

...

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) **sp, p1, exp**

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

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

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

...

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

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 **sp, p1, p2, exp**

29 *P2:* the second row of cups from your end **sp, p1, exp**

30 *P1:* Yes, I understand now!

31 *P1:* Gotcha

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

Swedish D5

36 *P1:* okej, nästa rad mot mitten **sp, p1, exp**

37 *P1:* från mitt håll står det en take-away bakom den vita muggen **sp, p1, exp**

38 *P1:* snett vänster om **sp, p1**

39 *P2:* Ok. Här det en vanlig vit mugg strax till höger om den vita närmast dig. **sp, p1, exp**

40 *P2:* Till höger och innåt bordet då. **sp, p1?**

41 *P1:* höger för dig eller mig? **rep**

42 *P2:* För dig. **sp, p1, exp**

43 *P1:* okej, den ser jag

Fig. 7: Clarification and repair: role and information

(in support of hypothesis 4) switching FoR – allow conversational partners to resolve them. Figure 8 shows Swedish dialogue D6 between turns 55–72.

In the preceding dialogue the participants were discussing the scene using the FoR assigned to Katie. However, in turn 55 P2 generates a description using their own FoR in response to which P1 raises a clarification request about the perspective they intend to use: P2's or Katie's. P2 answers "Katie's" but what they mean is close to or starting from Katie using P2's FoR (to the right of Katie using Katie's FoR) which is a cause of a misunderstanding that is resolved several turns later when P1 explicitly states that they should take P2's perspective. However, explicit definition appears to be a last resort for negotiating a FoR: participants start negotiating a FoR by simply generating descriptions using that FoR, using explicit reference if necessary, and expect that their conversational partners accommodate that FoR. Using P2 FoR in this dialogue also confirms both preferences discussed earlier. P2 has the best visual focus of the part of the scene that they are focusing on (good referring expressions can be generated) and P2 also takes the role of the describer (and P1 the role of the follower).

Swedish D6

55 *P2:* okej, fortsätter längs kanten på vänster sida? **sp, p2**
56 *P1:* vems perspektiv? **rep**
57 *P2:* Katies **sp, katie, exp**
58 *P1:* okej på kates vänstra sida innåt framför dig finns det en röd mugg **sp, p2, katie?, exp?**
59 *P1:* ditt höger **sp, p2?, katie?, exp**
60 *P1:* nej vänster **sp, p2, exp**
61 *P2:* va?? **rep**
62 *P1:* hahaha
63 *P2:* okej närmast mig då **sp, top, p2, exp**
64 *P2:* längst från dig, och Katies högra sida **sp, p1, katie, exp**
65 *P1:* japp snätt åt vänster framför dig **sp, p2**
66 *P1:* ditt vänster dvs **sp, p2, exp**
67 *P2:* röd, sen vit med lock, sen vit med öra i mitt nedre högra hörn **p2, exp**
68 *P1:* vi tar ditt perspektiv nu tycker jag, OKEJ! **p2**
69 *P2:* OKEJ
70 *P1:* ;)
71 *P1:* jag har bra perspektiv
72 *P2:* klart du har

English D2

146 *P2:* so you see that yellow cup to be right on teh corner? **p1**
147 *P1:* Yes
148 A yellow cup, on my right your left, with the handle facing east to me, west to you. **p1, p2, exp**

Fig. 8: Clarification, (explicit) repair and precision

During clarification there is also stronger demand for precision and hence over-specification. This is also clearly demonstrated in turns 146–148 of the English D2 dialogue in Figure 8.

5 Discussion and Future Work

We presented a study of how FoR is negotiated in free dialogue in English and Swedish. The observed strategies for choosing an FoR are similar between the two languages, differences appear mostly to be due to personal style and preferences of participants. Returning to the hypotheses from Section 3, p.4 we have provided evidence that

- there is no baseline preference for a specific FoR;
- there is no general alignment of FoR over dialogue but local alignment;
- participants do not use explicit descriptions at the beginning of alignment sequences;
- misunderstandings are associated with FoR change but there are also other factors related to the particular dialogue game in play.

In order to produce or interpret a spatial description, conversational partners need to take into account several sources of knowledge: (i) perceptual properties of the scene from which objects and geometrical arrangement of the scene can be conceptualised; (ii) knowledge about objects, properties and affordances and their interaction; (iii) interaction strategies with conversational partners

(including the language used). As discussed in this article all three modalities affect the assignment of the FoR and interact with each other. Here we mostly focused on (iii) because linguistic interaction through dialogue provides an overarching modality: it involves interaction with conversational partners and also with the environment in which they are located. We argued that FoR appears to be dependent on the dialogue games participants are engaged in, that is the communicative strategies adopted to achieve a task-oriented (sub-)goal in a particular scene. Overall, the assignment of FoR is driven by mutual understanding of each other and the world around us and resolution of misunderstanding. Our future work will focus on extending the corpus of dialogues in such a way that more reliable quantitative analyses can be performed, in particular with respect to identifying the features that are indicative of FoR change. Our ultimate goal is to model (human-like) spatial perspective taking in spoken dialogue systems.

Acknowledgements

The research of Dobnik and Howes reported in this paper was supported by a grant from the Swedish Research Council (VR project 2014-39) for the establishment of the Centre for Linguistic Theory and Studies in Probability (CLASP) at the University of Gothenburg. The research of Kelleher was supported by the ADAPT Centre for Digital Content Technology which is funded under the SFI Research Centres Programme (Grant 13/RC/2106) and is co-funded under the European Regional Development Fund.

References

1. Carlson-Radvansky, L., Irwin, D.: Frames of reference in vision and language: Where is above? Cognition **46**, 223–224 (1993)
2. Carlson-Radvansky, L., Logan, G.: The influence of reference frame selection on spatial template construction. Journal of Memory and Language **37**, 411–437 (1997)
3. Clark, H.H., Wilkes-Gibbs, D.: Referring as a collaborative process. Cognition **22**(1), 1 – 39 (1986)
4. Dobnik, S.: Teaching mobile robots to use spatial words. Ph.D. thesis, University of Oxford: Faculty of Linguistics, Philology and Phonetics and The Queen's College, Oxford, United Kingdom (September 4 2009)
5. Dobnik, S., Howes, C., Kelleher, J.D.: Changing perspective: Local alignment of reference frames in dialogue. In: Howes, C., Larsson, S. (eds.) Proceedings of SemDial 2015 (goDIAL): The 19th Workshop on the Semantics and Pragmatics of Dialogue. pp. 24–32 (2015)
6. Dobnik, S., Kelleher, J.D., Koniaris, C.: Priming and alignment of frame of reference in situated conversation. In: Rieser, V., Muller, P. (eds.) Proceedings of DialWatt - Semdial 2014: The 18th Workshop on the Semantics and Pragmatics of Dialogue. pp. 43–52. Edinburgh (1–3 September 2014)
7. Duran, N.D., Dale, R., Kreuz, R.J.: Listeners invest in an assumed other's perspective despite cognitive cost. Cognition **121**(1), 22–40 (2011)

8. Galati, A., Avraamides, M.N.: Social and representational cues jointly influence spatial perspective-taking. Cognitive Science **39**(4), 739–765 (2015)
9. Healey, P.G.T., Purver, M., Howes, C.: Divergence in dialogue. PLoS ONE **9**(6), e98598 (Jun 2014)
10. Healey, P.G.T., Purver, M., King, J., Ginzburg, J., Mills, G.J.: Experimenting with clarification in dialogue. In: Proceedings of the 25th Annual Meeting of the Cognitive Science Society. Boston, MA (Aug 2003)
11. Herskovits, A.: Language and spatial cognition: An interdisciplinary study of prepositions in English. Studies in Natural Language Processing, Cambridge University Press (1986)
12. Johannsen, K., de Ruiter, J.: Reference frame selection in dialogue: priming or preference? Frontiers in Human Neuroscience **7**(667), 1–10 (2013)
13. Kelleher, J.: A Perceptually based computational framework for the interpretation of spatial language. Ph.D. thesis, Dublin City University (2003)
14. Kelleher, J., Costello, F.: Cognitive representations of projective prepositions. In: Proceedings of the Second ACL-Sigsem Workshop of The Linguistic Dimensions of Prepositions and their Use in Computational Linguistic Formalisms and Applications. (2005)
15. Kelleher, J., van Genabith, J.: A computational model of the referential semantics of projective prepositions. In: Saint-Dizier, P. (ed.) Syntax and Semantics of Prepositions. Speech and Language Processing, Kluwer Academic Publishers, Dordrecht, The Netherlands (2006)
16. Kowtko, J.C., Isard, S.D., Doherty, G.M.: Conversational games within dialogue. HCRC research paper RP-31, University of Edinburgh (1992)
17. Levelt, W.J.M.: Cognitive styles in the use of spatial direction terms. In: Jarvella, R.J., Klein, W. (eds.) Speech, place, and action, pp. 251–268. John Wiley and Sons Ltd., Chichester, United Kingdom (1982)
18. Levinson, S.: Frame of reference and Molyneux's question: Crosslinguistic evidence. In: Bloom, P.and Peterson, M., Nadell, L., Garrett, M. (eds.) Language and Space, pp. 109–170. MIT Press (1996)
19. Levinson, S.C.: Space in language and cognition: explorations in cognitive diversity. Cambridge University Press, Cambridge (2003)
20. Li, X., Carlson, L.A., Mou, W., Williams, M.R., Miller, J.E.: 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 (2011)
21. Mills, G., Healey, P.G.T.: 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 (Sep 2006)
22. Pickering, M., Garrod, S.: Toward a mechanistic psychology of dialogue. Behavioral and Brain Sciences **27**, 169–226 (2004)
23. Pulman, S.G.: Conversational games, belief revision and Bayesian networks. In: et al., J.L. (ed.) CLIN VII: Proceedings of 7th Computational Linguistics in the Netherlands meeting. pp. 1–25 (1997)
24. Schober, M.F.: Speakers, addressees, and frames of reference: Whose effort is minimized in conversations about locations? Discourse Processes **20**(2), 219–247 (1995)
25. Schultheis, H., Carlson, L.A.: Mechanisms of reference frame selection in spatial term use: Computational and empirical studies. Cognitive Science **41**(2), 276–325 (2017)
26. Steels, L., Loetzsch, M.: Perspective alignment in spatial language. In: Coventry, K.R., Tenbrink, T., Bateman, J.A. (eds.) Spatial Language and Dialogue. Oxford University Press (2009)

27. Taylor, H.A., Tversky, B.: Perspective in spatial descriptions. Journal of Memory and Language **35**(3), 371 – 391 (1996)
28. Tenbrink, T., Andonova, E., Schole, G., Coventry, K.R.: Communicative success in spatial dialogue: The impact of functional features and dialogue strategies. Language and Speech **0**(0), 0023830916651097 (2016)
29. Watson, M.E., Pickering, M.J., Branigan, H.P.: 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 (2004)