Priming and Alignment of Frame of Reference in Situated Conversation

Simon Dobnik[1], **John D. Kelleher**[2] **and Christos Koniaris**[1]
[1]University of Gothenburg, Centre for Language Technology,
Dept. of Philosophy, Linguistics & Theory of Science, Gothenburg, Sweden
[2]Dublin Institute of Technology, Applied Intelligence Research Centre,
School of Computing, Dublin, Ireland
{simon.dobnik,christos.koniaris}@gu.se, john.d.kelleher@dit.ie

Abstract

In this paper, we study how the frame of reference (FoR) or perspective is communicated in dialogue through mechanisms such as linguistic priming and alignment (Pickering and Garrod, 2004). In order to isolate the contribution of these mechanisms we deliberately work with a constrained artificial dialogue scenario. First we collect data that deal with human behaviour in interpreting descriptions that are ambiguous in terms of the FoR. From these interpretations we extract and identify strategies for FoR assignment in conversations which we then apply to generate descriptions and measure human agreement with the system. Our findings confirm that both speakers and hearers rely on such mechanisms in conversation.

1 Introduction

A necessary basis for a successful human-machine interaction in a situated dialogue is the ability of the machine to understand and generate spatial references to objects in the spatio-temporal and discourse contexts. Studies of human-human communication, e.g., (Levelt, 1989), reveal that the speaker often uses projective spatial descriptions, e.g., "to the left of the chair" or "in front of the chair" without explicitly specifying the frame of reference, or perspective, according to which the hearer should interpret a scene. In principle, these spatial descriptions may be interpreted *relative* to either of the conversational participants ("…from my perspective", "…from your perspective") or to any other individual or object in the scene ("…from sofa's/Alex's position"). In order to be able to set the orientation of the coordinate frame such objects must have identifiable front and back. We avoid describing FoR as speaker-relative and hearer-relative as in a conversation their roles may change. Instead we refer to system-relative (S) and human-relative (H) FoR. Finally, the FoR may also be assigned *intrinsically* by the landmark/reference object ("the chair") (Levinson, 2003) which we mark as I.

Our long term research goal is to create artificial conversational agents that can participate in situated dialogue. Such an agent must be able to understand and use locative expressions, including those that are dependent on FoR. The agent must resolve the FoR before a geometric spatial template, representing, for example, a region corresponding to "to the left of", can be applied as the FoR sets the origin and the orientation of the coordinate system in which the spatial template is projected (Maillat, 2003). Possibly the simplest approach to handling the FoR issue that can be adopted when creating an artificial conversational agent is to assume or require that all FoR usage is relative to the artificial agents perspective. Unfortunately, however, our earlier work with a situated robot (Dobnik, 2009) shows that relativising all human spatial descriptions to the perspective of the robot adds considerable noise to the data which affects the performance of classifiers that attempt to capture spatial templates. Trafton et al. (2005) show that robots capable of making perspective shifts are more effective in interpreting human descriptions and Steels and Loetzsch (2009) show that they are more successful in learning and generating situated language. However, both approaches do not equip the robots with a model of perspective of the most likely FoR their conversational partner would expect which is the focus of our current study.

There are a number of factors that affect the choice FoR, including: task (Tversky, 1991), personal style (Levelt, 1982), arrangement of the scene and the position of the agent (Carlson-Radvansky and Logan, 1997; Kelleher and

Costello, 2009; Li et al., 2011), and the presence of a social partner (Duran et al., 2011). In this work, however, we focus on *linguistic priming* and *alignment*. By "linguistic" we mean expression of and exposure to content of linguistic utterances. We use the term linguistic priming to distinguish it from and relate it to other forms of priming, for example visual priming by the visual properties of the scene, and priming by the participant role in conversation (speaker/hearer). By alignment we mean adoption of common patterns of behaviour. Watson et al. (2004) conduct psychological studies that confirm the alignment of FoR between conversational partners following a linguistic priming. Johannsen and de Ruiter (2013) investigate further whether the alignment is due to priming or due to preference for a particular FoR in conversation and conclude that there is an interplay of both factors. In contrast to (Watson et al., 2004) and (Johannsen and de Ruiter, 2013) we designed a more complex structure of dialogue games where, for example, a priming step is followed by two interpretive steps before switching the communicative roles of participants, which allows us to study the attenuation of priming and the development of alignment.

Our study includes two experiments which were performed in a constrained spatial environment and dialogue (i) to control the influence of other non-linguistic priming factors, and (ii) to test how humans assign FoR at those points in dialogue where the FoR assignment is at stake: directly after a priming utterance, dialogue turns following this turn and subsequent dialogue turns where the interlocutors switch their roles (from interpretation to generation and vice versa). By examining the behaviour of dialogue participants at these dialogue points we address the following research questions: (i) do participants align their FoR with the linguistically primed FoR used by their dialogue partner; (ii) does the effect of priming degrade over dialogue games; and (iii) does priming persist over role changes?

Overall, if priming develops into alignment, it shows that agents behave cooperatively to their conversational partner (Clark and Wilkes-Gibbs, 1986). In dialogue each conversational participant has a dialogue game-board which contains their individual representation of the state of the dialogue (Ginzburg and Fernández, 2010). One part of the dialogue game-board is the common ground which contains assumptions that conversational participants believe that they have agreed upon. In the priming game (which contains an unambiguous utterance relative to the visual scene) both the hearer and the speaker push the FoR from the speaker's utterance to their common ground; the speaker when they choose what to describe and the hearer when they confirm that they have understood the utterance. In the subsequent ambiguous games both agents have a choice: should they generate and interpret the utterance relative to the FoR that is in the common ground of their dialogue game-board or should they update the FoR in their common ground with a different one. We hypothesise that if the agents are cooperative, they will tend to minimise the updates to the common ground unless this is not necessary, for example, there is no new priming of the FoR through other priming factors. We interpret the non-variability of the FoR in the common ground as alignment. Note that our notion of alignment is slightly different from (Watson et al., 2004) and (Johannsen and de Ruiter, 2013) who consider alignment to occur if a hearer primed with a particular FoR would use this FoR in their next utterance as a speaker. In our framework, alignment occurs earlier, at the point after the hearer updates their common ground with the primed utterance.

Our experiments study the dynamics of FoR updates to common ground in a restricted scenario. In Experiment I the system has no knowledge of the strategies for FoR assignment, instead we try to capture them through observing the behaviour of a human. The system primes the human with an unambiguous scene description and we capture what a human would do in terms of FoR assignment in the subsequent conversational games over visually ambiguous scenes, first when they have a role of the interpreter and finally when they become a generator. In Experiment II we test whether the human strategies for assigning FoR from Experiment I can be used by the system and whether human observers evaluate such behaviour positively. Here, the human primes the system in the first conversational game and in the subsequent games the system has a role of the generator and finally an interpreter of visually unambiguous scenes.

2 Experiment I: alignment of FoR

The focus of the reported research is to investigate the role of linguistic priming and alignment in FoR-usage in constrained situated dialogues in order to discover an inventory of strategies that an intelligent virtual agent could use to generate and interpret FoR-dependent locative expressions correctly. As a basis for our analysis of strategies, we collected a dataset of situated dialogues. To collect the data, we created a virtual scene embedded in a web-page in which a pre-scripted agent interacts with a human through a series of utterances in particular spatial scene configurations as shown in Figure 1.

Katie: I chose the blue box to the left of the chair.

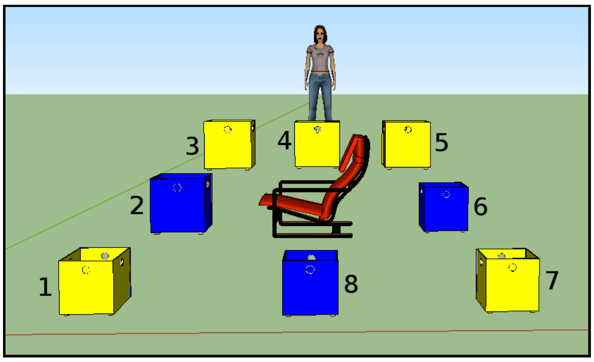

Figure 1: A scene from the virtual environment as seen by a human (not including numbers). The system (the character Katie) generates a description for which the human should decide on the most likely FoR by clicking on a box (2 = H*uman*, 6 = S*ystem* and 8 = I*ntrinsic*).

Conversational Games I

1. The system primes a human for FoR unambiguously: the scene contains only one blue box.

   S: "I chose the blue box to the left of the chair."
   H: Clicks on the intended target object.

2. The system generates an ambiguous description: the scene contains 3 blue boxes, one for each FoR interpretation (cf. Figure 1).

   S: "I chose the blue box to the left of the chair."
   H: Clicks on the intended target object.

3. Identical to Game 2 but with a different spatial description ("to the right of") and a different arrangement of blue boxes.

4. The system asks a human to describe the object that it chose (and marked by an arrow).

   S: "Tell me: which box did you choose?"
   H: Types in their description.

We deliberately opt for such a constrained artificial scenario for two reasons arising from our previous work where we examined assignment of FoR in unrestricted conversation between humans (Dobnik, 2012). Firstly, even if a dialogue task is designed to maximise the usage of spatial descriptions, for example as a variant of the map task (Anderson et al., 1991), longer sequences of potentially ambiguous utterances in respect of the FoR assignment are in minority and therefore one would need to collect a several times larger corpus to obtain a representative number of examples. Secondly, previous studies have shown the FoR assignment is influenced by several factors (task, arrangement of the scene, position of the agent and presence of the social partner) and hence a constrained scenario may be to our advantage as these factors can be controlled. In this study it is not our intention to model human dialogue as a whole but to extract the strategies of FoR assignment through linguistic priming at particular points of dialogue where its assignment is at stake in such a way that the strategies can be used for assignment or disambiguation of FoR in a dialogue manager.

We represent these points in dialogue as a sequence of four dialogue games (each consisting of two turns) which we summarise under the heading Conversational Games I. The conversation was initiated by the system in what we call the priming step (Game 1). This was followed by three games which were intended to show the development of linguistic priming into an alignment of the other agent, the human. Game 2 tested the effectiveness of FoR priming, Game 3 tested the persistence of priming under the same speaker-hearer roles and Game 4 tested the persistence of priming if the speaker-hearer roles change. The system had no knowledge about the FoR assignment (human (H), system (S) or intrinsic (I), i.e., relative to the chair). Rather, the study was intended to capture what FoR an interpreter and finally a generator of an utterance would assume after being linguistically primed for a particular FoR.

Data were collected from both supervised lab sessions and anonymous online contributions. In both cases the same web-interface was used. In total there were 75 trials from which 51 were completed and used in the study. Each participant made judgements for 12 games in total, i.e., 4 games for each of the 3 primed FoRs. All subjects were primed for FoR in the same order which was H $>$ I $>$ S. Table 1 shows conditional probabilities of a human selecting a particular FoR in each subsequent dialogue game following linguistic priming in Game 1. They reveal that priming in

Game 1 does have a strong effect on the human's choice of FoR in the subsequent games (the highest probabilities for each game given each priming are emphasised). Generally, humans align to all 3 FoR primed by the system in Game 2 and to H and I in Games 3 and 4. In Games 3 and 4 the alignment to S loses to the preference for I. This indicates that priming to H and I is persistent in conversation over several games but not priming to S the use of which persistently drops across subsequent games. The priming to H and I also carries over to the fourth conversational game where the speaker-hearer roles change. In more detail, the transition from Game 2 to 3 shows that the alignment to the primed FoR weakens for H and S but it grows stronger for I as shown by the spread of probabilities. This means that as the conversation proceeds there is more variation in the choice of S and H and less in the choice of I. This is because in each game following Game 1 the chosen FoR also adds secondary priming for the following game. If this FoR is the same as in Game 1, it will further strengthen the alignment to the primed FoR, otherwise it will weaken it. In Game 4 where roles change, i.e., human becomes a speaker and system becomes a hearer, an increase in the preference for H and a decrease in the preference for S relative to the previous game is found. This may be because at this stage priming by the speaker role for H is introduced (speakers being egocentric) which competes with the linguistic priming. Overall, at the end of the conversation (Game 4) the perspective that decreases the most is S and the one that remains the most dominant of all three is I.

We explain the increased preference for I at the expense of S if priming was followed on the grounds of the visual priming introduced by the chair. This is more visually salient than the system avatar. It is placed in the middle of the room, appears closer and larger to the human and is red. On the other hand the system avatar is a static character and therefore may lack the salience of an animate person speaking. Given this salience imbalance, humans performing the task may simply forget that they are talking to an agent and consequently focus on the chair. We hypothesise that this is the main reason why the usage of S is in decline in Games 3 and 4, although note that at the beginning of the conversation in Game 2 the likelihood of S following a primed S is higher than H following a primed H. Furthermore, the chair is also a convenient compromise to ground the FoR in for both the system and a human as it is not one of the agents speaking. Visual priming of the chair is constant throughout the conversation whereas speaker-related priming changes from one agent to another.

	Followed by		
Primed by	H	S	I
	Game 1		
H	1.000	0.000	0.000
S	0.000	1.000	0.000
I	0.000	0.000	1.000
	$\chi^2(4) = 388, p < 2.2 \times 10^{-16}$		
	Game 2		
H	**0.513**	0.145	0.342
S	0.073	**0.564**	0.364
I	0.098	0.131	**0.771**
	$\chi^2(4) = 75.250, p = 1.764 \times 10^{-15}$		
	Game 3		
H	**0.460**	0.108	0.432
S	0.111	0.426	**0.463**
I	0.083	0.117	**0.800**
	$\chi^2(4) = 52.828, p = 9.256 \times 10^{-11}$		
	Game 4		
H	**0.508**	0.127	0.365
S	0.308	0.250	**0.442**
I	0.175	0.018	**0.807**
	$\chi^2(4) = 33.613, p = 8.945 \times 10^{-7}$		

Table 1: The probabilities of selecting a particular FoR for each subsequent game given some priming (Game 1). The system primes all FoRs equally and the figures show that all participants correctly identified the unambiguous target object. The χ^2-test confirms the statistical significance of the differences in observed assignments/probabilities. We calculate the χ^2 statistic for each game separately which ensures independence of observations in respect to individuals.

Table 1 shows us whether linguistic priming of FoR initiated by the system in equal proportions develops into alignment of a human. Unfortunately, for this reason we are not able to extract the preference of humans for FoR in the priming Game 1. This would tell us the overall preference for FoR in this spatial and dialogue contexts in the absence of linguistic priming. We estimate this preference in Experiment II in Section 4.

3 Strategies of FoR assignment

How can the strategies for FoR assignment discussed in the previous section be integrated within a dialogue manager of a conversational agent? One way of representing them is using a simple 4-state graphical model as shown in Figure 2, where each state represents a dialogue game and contains a conditional probability table representing the likelihood of the chosen FoR (H, I or S) in that game, given that a particular FoR was chosen in the previous game. The graphical model can be applied as a classifier within dialogue rules that update the dialogue game-board.

Table 2 shows the conditional probabilities tables for states G_2, G_3 and G_4 of the network. If we choose maximum a posteriori hypothesis, the most likely choice of a FoR for a dialogue manager is always the same FoR as in the preceding step, except at the switch of the conversational roles in Game 4 where S chosen in Game 3 is followed by H in Game 4. Hence, due to the strong alignment of subjects in our experimental scenario the FoR assignment could be implemented in a dialogue manager with only two rules: **If** you are changing your role from interpreter to generator **and** the last FoR was grounded in the location of your conversational partner, **then** ground the FoR in your location; **else** do nothing.

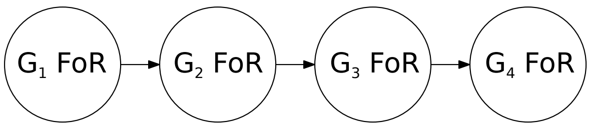

Figure 2: Block diagram of the Bayesian network. Each state of the network represents the model of FoR assignment for a particular dialogue game ($G_1 \dots G_4$), a sequence of generative and interpretive turns.

	Current game		
Previous game	H	S	I
		Game 1 Priming	
		Game 2	
H	**0.513**	0.145	0.342
S	0.073	**0.564**	0.364
I	0.098	0.131	**0.771**
		Game 3	
H	**0.792**	0.021	0.188
S	0.128	**0.766**	0.106
I	0.011	0.011	**0.979**
		Game 4	
H	**0.833**	0.119	0.048
S	**0.515**	0.364	0.121
I	0.064	0.021	**0.915**

Table 2: The conditional probabilities of selecting a particular FoR in the current game given a particular FoR in the previous game.

4 Experiment II: Application of the FoR alignment

In Experiment I we have shown how humans align their interpretation and generation of utterances involving FoRs to the linguistic priming by the system. We can now use the strategies of human alignment in the system to predict the most likely FoR for the utterance in a dialogue after the system has been primed by the human. In Experiment II we examine whether humans agree with the system using these strategies. In particular, would a human choose the same FoR as the system when it is generating unambiguous descriptions in Games 2–3 after being primed by a human in Game 1? Moreover, would a human taking on a speaker role in Game 4 also choose the FoR that the system would predict given the alignment strategies? To answer these questions we tested whether human strategies for interpretation of FoR could be used by the system for generation and vice versa as summarised in Table 3. We hypothesise that in this new scenario our conversational agent is maximally cooperative with its human partner as it is able to predict and foresee their beliefs and thus minimise the differences in their individual common grounds which would lead to misunderstandings. Hence, we expect that humans interacting with the system will evaluate its performance favourably.

Scenario	Games 1–3	Game 4
Experiment I	interpretation	generation
Experiment II	generation	interpretation

Table 3: The application of the FoR strategies in each experiment.

The listing Conversational Games II summarises the dialogues from Experiment II. In Game 1 the human is invited to prime the system. In Games 2 and Games 3 the human is first offered to choose an object whose location should be described, i.e. a box, then the system generates an unambiguous description of the box using the alignment model and asks the human for agreement. The human can acknowledge their agreement or provide a corrective description. We let humans choose the target box themselves as this gives them the opportunity to build their own representation of the scene before they hear the system's description. This way we attempt to counter the secondary priming introduced by the system's description which may lead human evaluators to overly agree with the system. Game 4 is similar to Games 2 and 3 except that in Game 4 both the human and the system generate a description and the system does so in the background. They agree if they both independently choose the same FoR.

We adapted the web-based environment used in Experiment I to the new scenario. The participants were instructed that they were engaged in a conversation with an artificial agent represented by the character facing them at the opposite side of the room (cf. Figure 1). In order to avoid complex descriptions such as “the box at the front and to the left of the chair” that are ambiguous between H and I, the corner boxes 1, 3, 5 and 7 were removed from the scene. The scene thus contained only 4 boxes which were all yellow. Humans communicated with the system by choosing a sentence from a list. This was considered appropriate in this context as we are only interested in the alignment of FoR and not in the spontaneous human generation. The sentences differed in respect to the choice of the spatial description and therefore FoR as shown in Game 2 in the listing. The evaluation was performed entirely through online crowdsourcing. Before starting, each participant had to supply a valid email address which attempted to prevent random participation. In total, judgements from 58 complete trials were collected (whereby one participant completed Games 1–3 twice which gave us 59 judgements for these games).

Conversational Games II

1. Human primes the system by describing a focused box.
   - S: “Where is the blue box?”
   - H: “The blue box is {to the left of | in front of | to the right of} the chair.”
2. Human chooses a box, the system uses the model for FoR, generates a description and asks the human for agreement.
   - S: “Please choose any box.”
   - H: Clicks on one box.
   - S: Using the model and the chosen box: “Aha, you chose the box in front of the chair. Would you agree?”
   - U: “Yes, the box is in front of the chair.” | “No, the box is {to the left of | behind} the chair.”
3. Identical to Game 2.
4. Human chooses a box which becomes the object in focus. The system asks the human to describe it and makes the assumption about the FoR the human would choose. The exchange succeeds if both are the same.
   - S: “Please choose any box.”
   - H: Selects one box by clicking.
   - S: “OK. Now, please tell me: where is the box that you chose?”
   - H: “The box is {to the left of | . . . } the chair.”
   - S: “Thank you.”

4.1 FoR to initialise conversation

In Experiment I the priming of the FoR was a task of the system which assigned the FoR in equal proportions. In Experiment II we want to test how adaptable is the system to the human and hence priming was a task of the human. Their preferences are summarised in Table 4. These probabilities can be used for initialising the conversation (cf. Section 2) and also tell us the preference of humans for FoR in the chosen visual and dialogue contexts; other contexts may lead to different preferences. The figures confirm the general tendencies already described in Section 2. There is a clear hierarchy of the FoR choice to start a conversation, which is I $>$ H $>$ S. However, one confounding factor impacting on this result is the fact that relationship between the FoRs and the spatial descriptions in Game 1 of the evaluation was kept constant across all participants. In particular, I was always associated with describing the blue box as being “in front of” the chair. Several researchers, for example (Logan, 1995; Franklin and Tversky, 1990), have reported results that humans find it easier to use and generate “front” and “back” descriptions rather than “left” and “right”. Consequently, this preference for I, although consistent with other research (Kelleher and Costello, 2005; Johannsen and de Ruiter, 2013), may be the result of an interaction with the relative ease of using “front” and “back”. In future work we intend to study this confounding factor in more detail.

Game	H	S	I
1	0.4068	0.0508	0.5424

Table 4: The likelihood of human selecting a FoR given the beginning of the conversation.

Moratz and Tenbrink (2006) report that humans prefer to use addressee-centred FoR and therefore adapt to their partner rather than take their own perspective which appears to be contradicted by our results as S is rarely used in comparison to H. When describing scenes humans prefer to use their own perspective over the perspective of the addressee, the system. However, speakers in Experiment II are performing different speech acts than those in (Moratz and Tenbrink, 2006): in the former they are providing a description and in the latter they are issuing a command to a person operating a robot. In (Moratz and Tenbrink, 2006) the hearer of the utterance is much more marked than in Experiment II which may count as a possible explanation for different experimental observations.

4.2 Human agreement with the strategies

As shown in Conversational Games II, in Games 2 and 3 the system used the FoR assignment strategies defined in Section 3 to predict the most likely FoR to generate a description and in Game 4 to make an assumption about the FoR in the description made by its human partner. Table 5 shows a confusion matrix between a system-predicted FoR and a human-chosen FoR. In Games 2 and 3 the human made a corrective description *after* they had heard the system's description. In Game 4 each made their choice independently. The term agreement may be interpreted as a satisfaction of a human with the system's generation in Games 2 and 3 and as a match in their predictions in Game 4. Note that the S is rarely chosen. This is because this FoR was disfavoured by humans in the priming step as shown in Table 4.

		Human		
Game	System	H	S	I
2	H	**22**	0	2
	S	0	**2**	1
	I	0	0	**32**
	Agreement	94.92%		
3	H	**22**	0	2
	S	0	**2**	1
	I	1	0	**31**
	Agreement	93.22%		
4	H	**18**	3	6
	S	0	0	0
	I	0	1	**30**
	Agreement	82.76%		

Table 5: Confusion matrix for the FoR chosen by the system and humans.

Overall, there is a high agreement of humans with the generations of the system: 94.92% in Game 2 and 93.22% in Game 3. The system does slightly less well predicting the FoR assumed for the subsequent generation of a human (82.76%). However, here both were "blind" to each others choice and hence the figure excludes the effect of a potential secondary FoR priming of a human in Games 2 and 3. The system and humans most disagree when the former predicts H but a human chooses S or I. Again, this variability of choice may be explained by the fact that the speaker-hearer roles have reversed and therefore the linguistic alignment is less stable in this new conversational context.

5 Discussion

The results from both experiments show that conversational partners act in a cooperative manner and they align to the linguistically primed perspective. This is the most frequently chosen strategy in this restricted scenario. However, linguistic priming is not the only strategy that they can use for FoR assignment: they may associate FoR with a salient centrally located reference objects (visual priming) or with the speaker or the addressee of the utterance depending on the utterance's speech act (priming by the participant's role in conversation). Both strategies exhibited a secondary effect in our experimental environment.

Directionals are a clear example that the meaning of linguistic expressions is dynamic and consistently changes through updates from the contexts in which the words are used (Larsson, 2007). Applying them in our constrained scenario demonstrates the plasticity of their meaning. An expression like "the box is to the left of the chair" is not only ambiguous in the assignment of the FoR but also in terms of the spatial template projected within the FoR, depending on the arrangement of the scene and the presence of distractor objects (Costello and Kelleher, 2006; Brenner et al., 2007). It follows that the meaning of directionals (and many other kinds of descriptions) relies on both the discourse and perceptual contexts in which they are used. If the meanings of words are dynamic and adaptable to contexts, it must be the case that there exist invariances within the contexts that are stable enough over time to be suitable referents. For example, reference objects in spatial descriptions ("the chair" in the example above) must not change size, shape and location in order to be good landmarks for "the box". The same holds for the discourse context where stability is achieved through alignment. If conversational participants choose the FoR randomly for each utterance, the information that is in the common ground of the dialogue (the sequence of the assigned FoRs) is not a reliable predictor of the forthcoming FoR choices. Participants would have to opt for some other strategy. This would be uncooperative given that linguistic interaction is the primary activity that they are engaged in. Grounding a different FoR in the common ground could also be due to miscommunication (the disagreement in Table 5) which is resolved between participants through alignment (see Mills and Healey (2008)). We hope to study the convergence of participants to a common FoR in case of miscommunication in our future work.

An important question we need to address is how well the strategies that we observe in the constrained scenario generalise to real situated dialogue. There are at least three issues at stake.

In real situated scenes there may be additional invariances in both linguistic and visual contexts that our experimentation did not take into account. This has been addressed extensively in previous research (cf. Section 1) and no doubt will be further investigated. Another question is how these invariances would be used for FoR assignment in cases where all of them are available. Our results suggest that linguistic priming may be stronger than visual priming which may be stronger than speaker priming. For example, the maximum probabilities for selecting each FoR in Game 2 in Table 1 tend to go with the linguistically primed FoR (in a diagonal) rather than visually primed FoR (column I) or speaker primed FoR (column S in Games 1–3 and column H Game 4). It is true that in the subsequent turns the linguistic priming degrades slightly but still has a considerable effect. Notice that in the absence of linguistic priming visual priming takes the lead (Table 4). Thirdly, real conversations may not consist of exactly four conversational games. The states that we explore in our constructed dialogues represent the key transitions between conversational games where the FoR is at stake and the speaker and the hearer must make a choice, namely at the beginning of the conversation, at a continuation of the conversation and at the change of the speaker-hearer roles. Hence, one could apply individual parts of the network to the relevant transitions in a dialogue. Finally, in a real scenario the sequences of conversational games that we explored may be interpreted by intermediate dialogue games that do not involve spatial reasoning. Would linguistic priming degrade in such cases and if so after what length of interruption? Does priming from an intermediary non-spatial dialogue game interfere with priming in a spatial game? This question would have to be answered by further experimental work.

6 Conclusions and future work

We established and tested strategies of perspective taking of conversational participants in a constrained situated dialogue where we focused on linguistic priming. From the collected dataset we can conclude that (i) in the absence of linguistic priming there exist preferences for the assignment of FoR in this scenario, namely *Intrinsic* $>$ *Speaker* $>$ *Hearer* (naming FoR after the conversational roles); (ii) the linguistic priming of FoR at the beginning of a conversation by one participants develops into alignment of both participants in the subsequent games, even when, but to a lesser degree, the speaker-hearer roles change; and (iii) visual properties of scenes and shifts in the speaker-hearer roles also exert priming and consequently affect the alignment to linguistic priming. Through the application of the FoR assignment strategies, we have demonstrated that humans evaluate them favourably, and the properties of the FoR assignment (i–iii) also hold. We additionally demonstrate that a model of interpretative judgements can be used for generating descriptions and vice versa. We expect that the user adaptation of the system would facilitate more effective spatial communication.

We chose a scenario with constrained visual and dialogue contexts to study the strategies of linguistic priming and alignment of FoR with an intention of formulating them as dialogue manager rules. In such a system the FoR assignment model would be part of a larger spatial cognition model which would also include a model for spatial templates and a model of world knowledge for prepositional use. An important part of the investigation would be how to make these models interact with each other aiming at the system to behave in a more cognitively plausible manner. An evaluation of the performance of such a situated agent by human observers would tell us how well the strategies identified in the present work generalise to new and less constrained situations.

Throughout our analysis we have noted how the visual priming of the chair may have drawn the participant's attention to the chair's FoR and that the reverse was the case for the static avatar representing the system. In future studies we will investigate the interaction between object salience and the adoption of FoR. We will also investigate the effects of the description choice between "front"/"back" and "left"/"right" on the FoR assignment by varying the priming from the current front-back dimension for I and the lateral dimension for H and S to the opposite. Overall, varying the parameters of the linguistic and visual contexts reminds us of an important theoretical insight that the meaning of linguistic descriptions is highly dynamic and context relative.

Acknowledgements

The authors wish to thank all participants in experiments and three anonymous DialWatt reviewers.

References

Anne H. Anderson, Miles Bader, Ellen Gurman Bard, Elizabeth Boyle, Gwyneth Doherty, Simon Garrod, Stephen Isard, Jacqueline Kowtko, Jan McAllister, Jim Miller, et al. 1991. The HCRC map task corpus. *Language and speech*, 34(4):351–366.

Michael Brenner, Nick Hawes, John D. Kelleher, and Jeremy L. Wyatt. 2007. Mediating between qualitative and quantitative representations for task-orientated human-robot interaction. In *IJCAI 2007*, pages 2072–2077.

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 and Deanna Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition*, 22(1): 1–39.

Fintan J. Costello and John D. Kelleher. 2006. Spatial prepositions in context: the semantics of near in the presence of distractor objects. In *Proceedings of the Third ACL-SIGSEM Workshop on Prepositions*, Prepositions '06, pages 1–8, Stroudsburg, PA, USA. Association for Computational Linguistics.

Simon Dobnik. 2009. *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.

Simon Dobnik. 2012. Coordinating spatial perspective in discourse. In Pierre Nugues, editor, *Proceedings of the 4th Swedish Language Technology Conference (SLTC 2012)*, pages 21–22, Lund, October 24–26.

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.

Nancy Franklin and Barbara Tversky. 1990. Searching imagined environments. *Journal of Experimental Psychology: General*, 119(1):63–76.

Jonathan Ginzburg and Raquel Fernández. 2010. Computational models of dialogue. In Alexander Clark, Chris Fox, and Shalom Lappin, editors, *The handbook of computational linguistics and natural language processing*, Blackwell handbooks in linguistics, pages 429–481. Wiley-Blackwell, Chichester, United Kingdom.

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

John D. Kelleher and Fintan J. Costello. 2005. Cognitive representations of projective prepositions. In *Proceedings of the Second ACL-SIGSEM workshop on the linguistic dimensions of prepositions and their use in computational linguistics formalisms and applications*, pages 119–127, University of Essex, Colchester, United Kingdom. Association for Computational Linguistics.

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

Staffan Larsson. 2007. A general framework for semantic plasticity and negotiation. In *Proceedings of the 7th International Workshop on Computational Semantics (IWCS-7)*, Tilburg, The Netherlands.

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.

Willem J. M. Levelt. 1989. *Speaking: from intention to articulation*. MIT Press, Cambridge, Mass.

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.

Gordon D. Logan. 1995. Linguistic and conceptual control of visual spatial attention. *Cognitive Psychology*, 28(2):103–174.

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 J. Mills and Patrick G. T. Healey. 2008. Semantic negotiation in dialogue: the mechanisms of alignment. In *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue*, pages 46–53. Association for Computational Linguistics.

Reinhard Moratz and Thora Tenbrink. 2006. Spatial reference in linguistic human-robot interaction: Iterative, empirically supported development of a model of projective relations. *Spatial Cognition and Computation*, 6(1):63–107.

Martin J. Pickering and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(2):169–190.

Luc Steels and Martin Loetzsch. 2009. Perspective alignment in spatial language. In Kenny R. Coventry, Thora Tenbrink, and John. A. Bateman, editors, *Spatial Language and Dialogue*. Oxford University Press.

J. Gregory Trafton, Nicholas L. Cassimatis, Magdalena D. Bugajska, Derek P. Brock, Farilee E. Mintz, and Alan C. Schultz. 2005. Enabling effective human-robot interaction using perspective-taking in robots. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 35(4):460–470.

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*, Chicago, USA.