Investigating the role of priming and alignment of perspective in dialogue

Simon Dobnik and **Christos Koniaris**
University of Gothenburg, Sweden
{simon.dobnik,christos.koniaris}@gu.se

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

Abstract

We examine the alignment of the primed frame of reference (FoR) for spatial descriptions over several utterances of a situated dialogue. We confirm the tendency of FoR alignment and that the *intrinsic* FoR is the most popular one independent of the priming.

1 Introduction

Typically, speakers use projective spatial descriptions such as "to the left of" or "behind" without a specification of the frame of reference (FoR) or perspective according to which the hearers should interpret the scene. For example, they can be interpreted relative to any of the discourse agents or the reference object itself if it can ground orientation of the FoR. The latter is a contextual variable which must be resolved from the visual or discourse context in order for the description to be grounded properly. The resolution of the FoR from the visual context succeeds if a given description can be unambiguously satisfied in it, i.e., a given pair of objects and the relation can be interpreted only according to that perspective for the utterance to be true. FoR can be resolved explicitly by linguistic discourse if the speaker describes it, or if, as we argue in this paper, the perspective is primed (combining observations from both perceptual and linguistic discourse) and aligned over several utterances of a situated conversation. (Carlson-Radvansky and Logan, 1997) is an early example of experimental research that examined the influence of FoR ambiguity on spatial term semantics. A finding from this study was that for vertically aligned prepositions, e.g., *above*, there was a preference for the *hearer/viewer-centric* FoR. Later work, (Kelleher and Costello, 2005) examined the impact of FoR ambiguity with respect to horizontally aligned prepositions, e.g., *in front of*. Interesting, this research reported a preference for the *intrinsic* FoR. Neither of these studies explicitly considered the effect of priming on FoR selection. More recently, (Li et al., 2011) studied the impact of FoR preference on object selection from an array of objects. The results from this study indicate that the intrinsic FoR of the object array was preferred. Again the study did not examine any priming effects. Finally, (Duran et al., 2011) examined the effect of social factors – such as the presence of a social partner and their ability to use a FoR – on reference frame selection. In contrast with the previous studies where the *intrinsic* FoR was preferred, this study found that participants invested in either an other-centric (*speaker-relative* in the terminology used in this paper) or egocentric (*hearer-relative*) mode of responding. In this study we are interested in mechanisms of such priming and alignment of FoR over several utterances in a way that they could be implemented as a model of a dialogue manager of a situated conversational agent (Trafton et al., 2005).

2 Experiment

(Dobnik, 2012) identifies the strategies of reference alignment and coordination by examining a small corpus of situated conversations between two human agents. Here, we build on this work by constraining the scenarios in such a way so that we can study under what conditions the identified strategies are applied and how are they followed. We replace one of the conversational partners with a pre-scripted virtual agent and restrict her utterances to particular scene configurations. In each turn, the agent generates a spatial description of a scene from which the hearer may or may not resolve the reference frame. The human must click on the object referred to by the description and so confirms their interpretation. During the priming step a description and a scene are chosen so that

"I chose the blue box to the left of the chair."

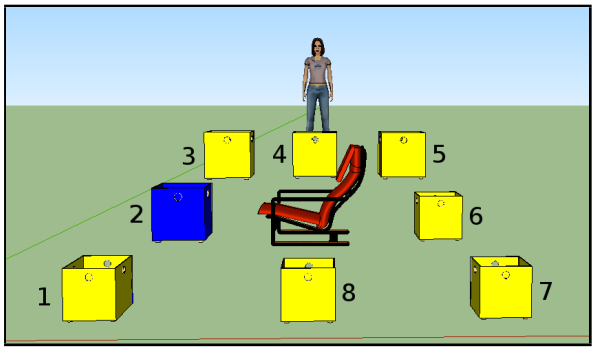

Figure 1: The description and the scene uniquely ground the reference frame in the priming turn of the conversation to H. If boxes 6 and 8 were also blue, then the description would also be interpretable under S and I reference frames. Object numbers were hidden from participants.

only one object matches as the description's reference and hence the description can only be interpreted according to one perspective (speaker-relative (S), hearer-relative (H) and intrinsic (I)) as shown in Figure 1. In the second turn, the system generates another description but in this case it matches the scene ambiguously in respect to all three reference frames, i.e., there would be three objects matching the description, one for each FoR interpretation. The human now has a choice to follow the primed FoR or choose a different one. Following a successful interpretation, the system generates yet another similar description. Finally, in the fourth turn the floor is handed to the human and they are invited to describe the location of an object indicated by the system. The purpose of this turn is to see whether the priming would also be preserved when the speaker–hearer roles change.

Since priming is given for all three FoRs and there are dialogue segments of 4 turns, this gives us "conversations" that contain totally 12 turns per participant (75 participants, 51 complete trials used). In this paper we concentrate only on the results from the second and third turn of each primed dialogue segment ($3 \times 2 \times 51 = 306$ utterances). The experiment was implemented as a web page and the results are from both supervised lab sessions and anonymous online contributions.

3 Results and discussion

Our findings are presented in Table 1. The first row shows the number of trials over all participants for each of the three primed turns. The second row shows the hearers' own preference for FoR. The third gives us the number of trials where the priming succeeded into alignment and the fourth shows utterances where although a FoR was primed in the previous utterance it was not applied by the hearer in this utterance. The rows 58 list the breakdown of FoR usage for the cases where the user adopted a different FoR to the primed one. Finally, the last row shows the number of utterances a particular FoR was used instead of the primed one.

	Hearer (H)	Intrinsic (I)	Speaker (S)
Primed with	102	102	102
Used by hearer	74	157	75
Priming succeeded	52	78	43
Priming failed	50	24	59
H priming followed by	–	32	18
I priming followed by	10	–	14
S priming followed by	12	47	–
Used instead of prime	22	79	32

Table 1: Summary of the number of utterances/trials according to the FoR assignment

The results in Row 3 show that priming has an effect on the choice of the FoR in the subsequent utterances (baseline count per primed FoR is 102/3 = 34). The results also show that there is a clear preference for using intrinsic FoR as shown previously in the literature (see Section 1): Row 2 and the breakdown of non-alignment in rows 5–8. Our impression from short discussions with some of the participants is that this choice may be seen as a convenient way of setting a "neutral", objective reference that both the hearer and the speaker can easily refer to during their communication. Furthermore, the chair which is setting the intrinsic FoR is also providing additional *perceptual priming* which may be another contributing factor: the chair is large, red and in the centre of the scene.

4 Future work

In the immediate future work, we will analyse the distribution of alignment between the Turn 2 and 3, and the behaviour of the hearer when they take on the role of a speaker (data from Turn 4) and also extend our experiments to include conditions under which reference objects receive different visual priming.

References

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.

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 others perspective despite cognitive cost. *Cognition*, 121(1):22–40.

John Kelleher and F Costello. 2005. 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*.

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.

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.

Investigating the role of priming and alignment of perspective in dialogue

Simon Dobnik[1], *Christos Koniaris,*[1] *and John D. Kelleher*[2]

[1]*University of Gothenburg, Sweden*
[2]*Dublin Institute of Technology, Ireland*

Aims

How do conversational participants resolve the frame of reference (FoR) ambiguity in situated conversation involving spatial descriptions?
What is the effect of priming on FoR use in situated conversation involving spatial descriptions?

The underspecification of FoR

- The FoR may be resolved from the visual context.

"I chose the blue box to the left of the chair."

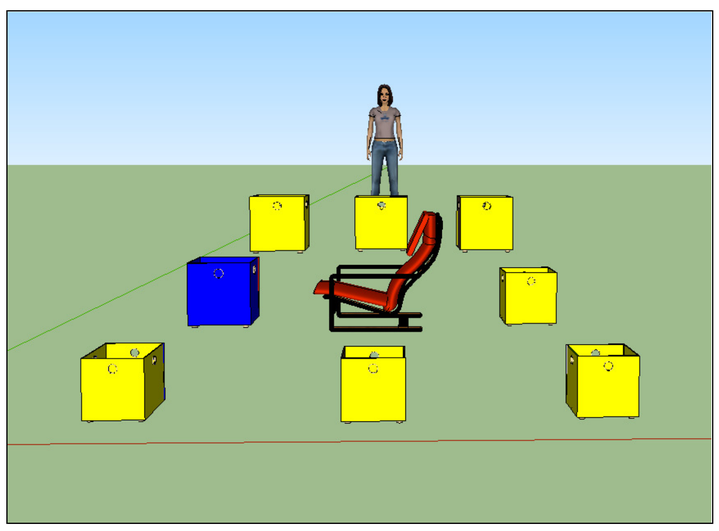

Figure 1: Which blue box did the speaker refer to? 2: hearer-relative (H) FoR; speaker is the person facing us in the scene.

- FoR may be primed by perceptual and linguistic discourse.
- FoR may be integrated in the common ground of the dialogue and aligned over several utterances. (Dobnik, 2012)

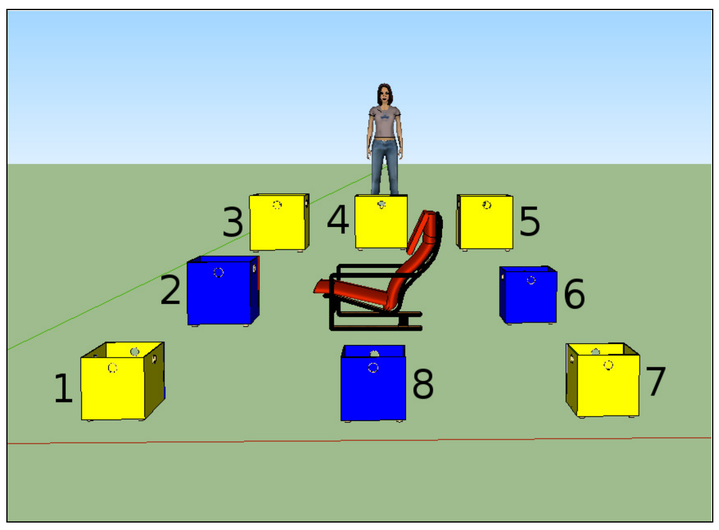

Figure 2: Which blue box did the speaker refer to? 2: hearer-relative (H) FoR; 6: speaker-relative (S) FoR; 8: chair-intrinsic (I) FoR.

- For some tasks, situation context or spatial relations there may exist a preference for a particular FoR in addition to discourse priming, see (Kelleher and Costello, 2005) and cf. (Duran, Dale, and Kreuz, 2011).

Experiment

- Virtual environment and dialogue interface implemented as a webpage.
- One of the conversational partners is a prescripted virtual agent and her utterances are restricted to particular scene configurations.
- The human must click on the object referred to by the description and confirm their interpretation or describe an object to the agent indicated by an arrow.
- Each dialogue segment is primed with one of 3 FoRs and consists of 4 turns (=12 turns per participant).

Turn 1 (Priming): a description and a scene are generated so that only one object matches as the description's reference and hence the description can only be interpreted according to one FoR (Figure 1).

Turn 2 (Does priming succeed?): The system generates another description (with a different spatial description) that matches the scene ambiguously in respect to all three FoRs (Figure 2).

Turn 3 (Does alignment persist?): Same as Turn 2 but with a different spatial description.

Turn 4 (Does priming preserve if speaker-hearer roles change?)

"Tell me: which box did you choose?"

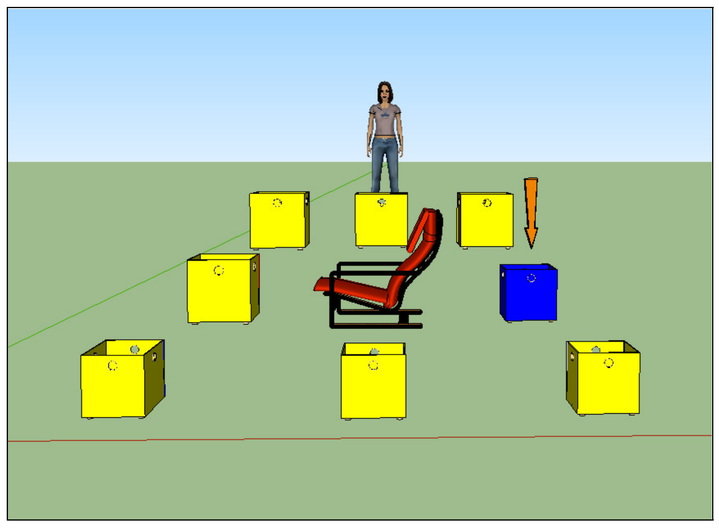

Figure 3: it is now the user's turn to describe.

Dataset

- 75 participants/trails (both from supervised lab sessions and anonymous online contributions) of which 51 complete trials were used.
- Here we consider data from Turn 2 and 3: the effect of priming on FoR alignment in language understanding.
- $3 \times 2 \times 51 = 306$ utterances

Results

	H	I	S
Primed with	102	102	102
Used by hearer	74	157	75
Priming succeeded	52	78	43
Priming failed	50	24	59
H priming followed by	–	32	18
I priming followed by	10	–	14
S priming followed by	12	47	–
Used instead of prime	22	79	32

Table 1: Summary of the number of utterances/trials according to the FoR assignment. **Primed with:** total number of primed utterance by FoR; **Used by hearer:** hearer's preference for FoR; **Priming succeeded/Priming failed:** the success of priming into alignment; **H/I/S followed by:** confusion matrix for the alignment; **Used instead of prime:** confusion matrix summary.

Results

- Priming has an effect on the choice of the FoR in the subsequent utterances: baseline count per primed FoR is 102/3 = 34 (cf. Row 3).
- There is a clear preference for intrinsic FoR as shown previously in the literature: Row 2 and the breakdown of non-alignment in rows 5-8.
- It is a convenient way of setting a "neutral", objective reference that both the hearer and the speaker can easily refer to during their communication.
- The chair (large, red and in the centre of the scene) setting the intrinsic FoR is also providing additional visual priming.

References

Dobnik, Simon. 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.

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

Kelleher, John and F Costello. 2005. 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*.