# Alignment of Reference Frames in Dialogue

**Matthew E. Watson (s0129081@sms.ed.ac.uk)**
Department of Psychology, 7 George Square
Edinburgh, EH8 9JZ, UK

**Martin J. Pickering (Martin.Pickering@ed.ac.uk)**
Department of Psychology, 7 George Square
Edinburgh, EH8 9JZ, UK

**Holly P. Branigan (holly.branigan@ed.ac.uk)**
Department of Psychology, 7 George Square
Edinburgh, EH8 9JZ, UK

**Abstract**

Previous research has shown that interlocutors in a dialogue align their utterances at several levels of representation. This paper reports two experiments that use a confederate-priming paradigm to examine whether interlocutors also align their spatial representations during dialogue. Experiment 1 showed a significant reference frame priming effect: Speakers tended to use the same reference frame to locate an object in a scene as the frame that they had just heard their interlocutor use. Experiment 2 demonstrated the same pattern even when the speaker's description and their partner's previous description involved different prepositions. Hence the effect cannot be explained in terms of lexical priming of a particular preposition. Our results are strong evidence that interlocutors in a dialogue align non-linguistic as well as linguistic representations.

Research on dialogue has suggested that the traditional methods employed in psycholinguistics may not give a true, or at least complete, account of human language. The traditional approach focuses largely on monologue and involves investigating single word utterances in isolated controlled circumstances, e.g. the picture naming paradigm, or the lexical decision task. However, Clark (1996) pointed out that the natural setting for language is dialogue, and that language does not normally occur in these isolated circumstances, thus questioning the ecological validity of traditional methods. The realization of this has led to a research program into how language is used in dialogue (e.g., Clark & Wilkes-Gibbs, 1986; Horton & Keysar, 1996; Garrod & Anderson, 1987). Research in this framework has shown that interlocutors in a dialogue tend to align their utterances: Over the course of a conversation participants will come to communicate in a similar fashion to each other. This occurs at several levels of communication, including the conceptual (Garrod & Anderson, 1987), lexical (Clarke & Wilkes-Gibbs, 1986) and syntactic (Branigan, Pickering, & Cleland, 2000) levels. In these experiments, participants usually achieved alignment without resorting to overt negotiation. In the case of syntactic alignment at least, many subjects were not aware that they were aligning.

Pickering and Garrod (in press) proposed a mechanism for how alignment is achieved between interlocutors. According to this theory, alignment is the basis for successful dialogue; misunderstanding occurs when alignment is not achieved. Alignment occurs when the two interlocutors employ equivalent representations at different levels, and arises from an automatic priming mechanism. This allows alignment to be achieved quickly and efficiently without reliance upon time-consuming strategies of open negotiation. Indeed, such strategies are only employed when the primitive mechanisms fail. To prevent unnecessary negotiation Pickering and Garrod suggest a second primitive mechanism that allows repair of representations when misalignment occurs; see Garrod and Pickering (2004) for a summary.

Dialogue research has shown alignment of linguistic representations, but alignment is hypothesized also to occur for conceptual representations, such as those associated with object location. A speaker's conceptual representation of where objects are located is reliant upon an overall spatial representation, which underpins the use of spatial language. In order to describe object locations effectively it is important that both interlocutors take the same perspective (Levelt, 1989) concerning the objects they are locating. For example, an addressee must understand whose left a speaker is talking about. In the same way that interlocutors align on which lexical terms should be used to describe a scene, it would be advantageous for interlocutors to align on which perspective a scene should be described from.

The perspective that is taken depends upon the reference frame that is applied to a spatial representation of a scene. A reference frame is an axial co-ordinate system that defines regions extending from the origin, whose axes are labelled with directional terms. The object to be located (figure object) can then be located in relation to another object (reference object) based upon the directional axes of the reference frame. However, there are three different types of reference frame (at least in English; other languages use

only two or even one; Levinson 2003) that a speaker can employ in order to locate an object: absolute, relative, and intrinsic. It is important that the addressee knows which of these the speaker is using in order to successfully understand an utterance.

The absolute reference frame locates an object based upon salient, stable features of the environment, for example, the cardinal directions. The dot in Figure 1 can be described as *west of the chair* if the page is held horizontally with the top of the page facing north.

The intrinsic reference frame locates an object based upon the directional features of the reference object. The dot in Figure 1 can be described as *above the chair* because it is in alignment with the top of the chair.

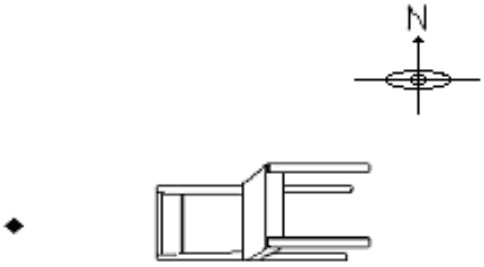


Figure 1. The dot can be described as west using an absolute reference frame, above using an intrinsic reference frame or left using a relative reference frame.

The relative reference frame locates an object in relation to the viewpoint of an observer. The axes of the reference frame are labelled based upon the features of the person upon whose viewpoint the location is based. In Figure 1 the dot would be described as *left of the chair* using a relative reference frame. (In many cases the relative reference frame is used from the viewpoint of a speaker or an addressee, but it can also be from a third person perspective.)

The above tripartite classification of reference frames follows that proposed by Levinson (1996, 2003), and is distinct from the classification traditionally employed in the psycholinguistic literature, which identified absolute, deictic, and intrinsic reference frames, all defined on the basis of their origin. (Deictic reference frames are all reference frames with an egocentric origin.) Levinson pointed out that this traditional system is not an appropriate way to categorize reference systems because it is possible to have a non-deictic relative reference frame, such as *The ball is to the right of the tree as you look at it*, and a deictic intrinsic reference frame such as *The ball is in front of me*.

When describing an object's location, an individual has to select one of these reference frames to use in preference to either of the other two reference frames. Carlson-Radvansky and Jiang (1998) showed that reference-frame selection is achieved via inhibition of non-selected reference frames. When participants used a relative reference frame to identify an object's location, they were slower to describe an object's location using an intrinsic reference frame immediately afterwards. Inhibition operates not only on the endpoint of an axis, but on at least the entire axis, e.g. if *left* (intrinsic) is inhibited then using *right* or *left* (intrinsic) in the subsequent description will take longer than using a relative reference frame.

The findings of Carlson-Radvansky and Jiang (1998) suggest that reference frames are influenced by low-level priming. However, the results do not establish whether or not this occurs during dialogue: Reaction time was used as a measure of cognitive effort in trials whereas in dialogue any effect of priming must manifest itself by a change in the person's linguistic behaviour. Furthermore, Carlson-Radvansky and Jiang's (1998) experiment only investigated inhibition of the endpoint of an axis and the inhibition of the axis itself. If interlocutors align reference frames, we would expect them to align the entire reference frame rather than just part of it. Therefore it is unclear whether this kind of priming is enough to cause the alignment of reference frames between interlocutors in the manner described by Pickering and Garrod (in press).

In two series of experiments Schober (1993, 1995) showed that the reference frame which an individual selects is affected by their partner in a conversation. Individuals who described the location of an object to a partner who viewed the scene from a different perspective were more likely to describe the location from their partner's perspective. When the partner queried such descriptions, they used their own perspective to describe object location. Schober concluded that interlocutors use conscious strategies to collaborate in ways of describing object location.

Schober's results suggest that interlocutors may align reference frames. However, it is not clear that this is necessarily the case. In his experiments, two participants interacted freely, allowing little control over what was said by each pair. This means that pairs of participants may be reverting to default reference frames. Furthermore, in a large proportion of trials participants located objects using terms that required no reference frames (e.g. *next to*, *between* and so on).

The present work is an experimental investigation to discover whether or not interlocutors align reference frames. The investigation uses a confederate-priming paradigm (e.g. Branigan et al., 2000) where a naïve participant and a participant who is - unknown to the naïve participant - a confederate of the experimenter and who is following a script, communicate during the experiment. If interlocutors do align reference frames then they will use a reference frame significantly more when they have just heard an utterance using that reference frame than when they have just heard an utterance using an alternative reference frame. Alternatively interlocutors may select a reference frame based solely upon the perceptual properties of the spatial

array, in which case they should be unaffected by the reference frame just used by their partner. Our experiments also set out to separate priming for reference frames from lexical priming. If priming of reference frames exists separately from lexical priming, we can expect subjects to use a reference frame significantly more if they have just heard an utterance using that reference frame even if the same spatial term is not applicable to both utterances.

## Experiment 1

### Method

**Participants** 12 students of the University of Edinburgh were paid volunteers in the experiment, which lasted 20 minutes. All were native English speakers.

**Materials** The experiment was run on two computers positioned back to back, using E-prime software. One program was created for the confederate and consisted of sentences positioned in the centre of the screen of the form "The dot above the chair". This formed the script for the experiment. The second program was for the participant and displayed pictures for the match and describe phases of the experiment.

12 monochrome objects were used as reference objects, all fitting into a rectangle 93 pixels high and 121 pixels wide. Two versions of each object were used, one rotated 90° clockwise and one rotated 90°anti-clockwise.

The figure object was an 11x11 pixel square rotated so that its vertices were the top, bottom, leftmost, and rightmost points. The figure object was located above, below, left, or right (in a relative reference frame) of the reference object. The centre of the figure object was positioned between 125 and 130 pixels from the centre of the reference object.

**Design** There were 3 within-participants and within-items factors: Prime Reference Frame (Relative vs. Intrinsic); Preposition (Same Preposition vs. Different Preposition); and Target Plane (Vertical vs. Horizontal). These are exemplified in Figure 2. The prime scene in Figure 2 can either be described as *The dot above the camera* (relative reference frame) or *The dot right of the camera* (intrinsic reference frame). In the top diagram of Figure 2, alignment requires using the same preposition (either *above* or *right of*); in the bottom diagram, alignment require using a different preposition (either *left* or *below*). Finally, the top target scene is aligned vertically whilst the bottom target scene is aligned horizontally.

Two lists of 96 trials were constructed, with each trial consisting of a match phase and a describe phase. The reference objects in each list were rotated clockwise and anti-clockwise on half of the trials each. Reference frame was counterbalanced across list and rotation. Preposition overlap was counterbalanced across rotation in each list. Participants saw 12 trials in each of the 8 conditions formed by crossing the three factors. The trials were presented in a fully randomized order, which was different for each participant.

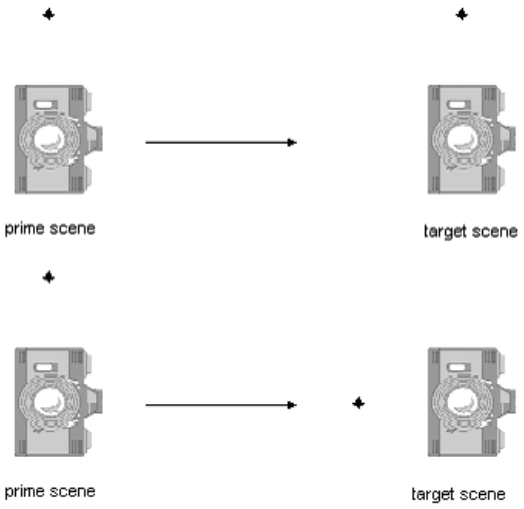


Figure 2: The top diagram shows the 'same preposition' condition. The bottom diagram shows the different preposition condition.

**Procedure** The two participants were introduced to each other (throughout the experiment, the experimenter treated the confederate as if she was a naïve participant).

The participant and confederate each sat at a computer each. The computers were situated back to back so that neither could see each other, or the other's screen. After hearing instructions, participants pressed the space bar to begin a practice session of 8 trials, one trial corresponding to each of the 8 conditions. Instructions then appeared on the screen signalling the end of the practice session and the start of the experiment. Each trial proceeded as follows: After participants pressed <space> to begin, the match screen appeared. The match screen contained two examples of a reference object (both the same, with one on the left and one on the right) and a dot located above, below, left or right of each one. The confederate gave a description of the location of the dot in relation to the object. The participant then chose which of the two examples on the screen matched the confederate's description of the dot location accurately, pressing the M key if it was the right-hand example and the Z key if it was the left-hand example. Participants were told that if they were not sure which picture matched their partner's description to pick the one they thought matched most closely.

After selection the match scenes disappeared (no feedback was given) and a fixation cross appeared in the centre of the screen. This remained on screen for 1000ms. The fixation cross was then replaced by a reference object in the centre of the screen with a dot above, below, left, or right of it. Participants then described the location of the dot in relation to the object. After describing this they pressed space and the scene disappeared. It was replaced by a fixation cross in

the centre of the screen for 500ms. This then disappeared and the next trial began with a match task.

## Results

For the analysis participants' first responses were used. The percentage of intrinsic responses were then analyzed using two 2x2x2 repeated measures ANOVAs (by participants (F1) and by items (F2)), with Prime reference frame (intrinsic or relative), preposition (same or different), and Target plane (horizontal or vertical) as factors.

Table 1 shows the mean number of intrinsic responses used by subjects in each of the 8 conditions. There was a significant main effect of Prime reference frame (62.9% vs. 53.3%; $F1(1,11) = 26.86$, $p < .01$; $F2(1,11) = 9.35$, $p < .05$). That is, participants were significantly more likely to use an intrinsic reference frame after the confederate had used an intrinsic reference frame, compared to when the confederate had used a relative reference frame.

Table 1:
Mean percentage of intrinsic responses in Experiment 1.

| | Relative Prime | | Intrinsic Prime | |
|---|---|---|---|---|
| | Same | Diff | Same | Diff |
| Vertical | 32.3 | 52.8 | 54.9 | 52.3 |
| Horizontal | 63.6 | 64.4 | 73.3 | 70.9 |

When the figure and reference objects were aligned vertically, participants used an intrinsic reference frame 48% of the time compared to 68% when the alignment was horizontal. This difference was significant ($F1(1,11) = 8.07$; $p < .05$; $F2(1,11) = 101.17$; $p < .01$), showing that participants were significantly more likely to use an intrinsic reference frame when the objects were aligned horizontally than when they were aligned vertically.

As expected there was no effect of preposition ($p > .05$): Participants used an intrinsic reference frame as much when the prepositions were the same as when they were different. This is regardless of which reference frame the confederate used.

There was a significant two-way interaction between Prime Reference Frame and Preposition ($F1(1,11) = 13.07$; $p < .01$; $F2(1,11) = 6.19$; $p < .05$). All other two-way interactions were non-significant ($p > .05$). Post-hoc analyses showed that these interactions occurred because of a difference between two of the eight conditions, relative, same, vertical and relative, different, vertical: The former yielded 32.3% intrinsic responses whereas the latter yielded 52.8% intrinsic responses ($t(23) = -2.91$; $p = .01$).This means that participants were more likely to use a relative reference frame when the reference and figure object were aligned vertically (i.e. they would use *above* and *below* to describe the dots' location) following the confederate using a relative reference frame when there was preposition overlap (i.e. the confederate used above or below) than when there was no preposition overlap (i.e. the confederate used left or right).

## Discussion

The results of Experiment 1 show an effect of alignment of reference frames. Participants were more likely to use an intrinsic reference frame after the confederate had used an intrinsic reference frame.

The significant effect of Target Plane indicates that participants preferred to use the lexical terms *above* or *below* to *left* or *right,* regardless of reference frame. This was expected because the top/bottom axis is easier to identify (than the left/right axis) due to asymmetries of the reference objects along this axis (Bryant & Wright, 1999).

One of the important goals of the experiment was to distinguish lexical priming effects from reference-frame priming effects. A sole effect of reference frame priming would have meant that participants aligned reference frames as much when the prime and target scenes were the same (represented in the upper portion of Figure 2) as when the prime and target scenes were different (represented in the lower portion of Figure 2). However, the presence of a significant interaction between Prime reference frame and preposition condition meant that this was not the case. This interaction was caused by two of the conditions; the other three pairs of same/different conditions yielded no significant differences between them. This indicates that the apparent lexical priming effect was evident only when the relative reference frame was used and the figure and reference objects were aligned vertically. Such a situation would seem unusual, because it should be the case that lexical priming is evident for all same/different pairs of conditions.

However, there is an alternative explanation for this pattern of data that does not rely upon lexical priming. We noted that participants used intrinsic left and right differently (in fact, inversely) to the confederate. Thus, participants would describe the prime scenes in Figure 2 as *the dot left of the camera*, whereas the confederate described them as *the dot right of the camera.* Therefore for half of the match tasks in the relative, vertical, different condition, the non-matching scene also provided a match to the confederate's description if an intrinsic reference frame was applied (according to the participant's interpretation). This would be the only condition in which potential confusion could arise. Therefore, for this condition, if participants chose the non-matching scene in the match task they would effectively be primed to use the intrinsic reference frame rather than the intended relative reference frame.

In Experiment 2, we therefore made the confederate describe intrinsic left and right in the way that participants had done in Experiment 1, in order to see whether the observed interaction was due to lexical priming, or was instead an artefact of the participants' misinterpretation of what the confederate was describing as intrinsically left and right.

## Experiment 2

Experiment 2 was a replication of Experiment 1, with the exception that what was described as left and right intrinsic was reversed in accordance with participants' interpretations from Experiment 1. 16 further students from the University of Edinburgh were paid volunteers in the experiment, which lasted 20 minutes.

### Results

The analyses were conducted in the same fashion as Experiment 1. Table 2 shows the mean number of intrinsic responses used by subjects in each of the 8 conditions. As in Experiment 1, there was a significant main effect of Prime Reference Frame ($F1(1,15) = 6.79$; $p < .05$; $F2(1,11) = 24.36$; $p < .01$): Participants used an intrinsic reference frame more often following an intrinsic description by the confederate than following a relative description by the confederate.

Table 2:
Mean percentage intrinsic responses in Experiment 2.

| | Relative Prime | | Intrinsic Prime | |
|---|---|---|---|---|
| | Same | Diff | Same | Diff |
| Vertical | 32.1 | 26.6 | 41.2 | 39.1 |
| Horizontal | 39.6 | 38.1 | 52.1 | 48.9 |

However, the interaction between Prime Reference Frame and Preposition did not reach significance, indicating that there was no effect of using the same lexical item for the prime and target ($F1(1,15) = .018$; $p > .05$; $F2(1,11) = 3.02$; $p > .05$). All other interactions were non-significant (all $p > .05$).

## General Discussion

The results of this study show that interlocutors align reference frames when describing objects' locations. Importantly, the results indicate that alignment is not due to lexical priming caused by the experimental participant repeating the preposition just used by the confederate.

The apparent lexical priming effect shown in Experiment 1 was due to the participants interpreting left and right intrinsic differently to what was intended by the confederate. When the source of this difficulty was addressed in Experiment 2, this effect was not evident. The results showed no difference in the proportion of reference-frame alignment when the naïve participant used the same preposition as the confederate, as when a different preposition was used.

Our results support the hypothesis that interlocutors align at many levels of representation when conversing (Pickering & Garrod, in press). Furthermore, it extends this alignment beyond linguistic representations and into an aspect of conceptual representation, i.e., the spatial domain. These results, however, do not precisely determine the mechanism by which alignment is achieved. In particular it is not clear whether participants make some use of a deliberate strategy to make the task easier for their partner. For example, it is possible that participants may be partly aware of the importance of aligning without realizing exactly what they are aligning on.

What is surprising about these results is that there was no cumulative effect of lexical priming and reference frame priming. Other studies have shown a larger alignment effect when more factors are common between the prime and the target (e.g. Branigan et al., 2000; Cleland & Pickering, 2003). The lack of a cumulative effect may be due to the nature of the lexical items used in this experiment. The prepositions were used to refer to both their intrinsic relation and relative relation, and so held little meaning independent of the reference frame.

The results also support the work of Carlson-Radvansky and Jiang (1998) who showed that reference frames were subject to negative priming. Their investigation only focused upon inhibition along a single axis of a representation. The results of this study extend these findings and show that activation of one axis of a reference frame activates the whole reference frame (at least in 2 dimensions), indicating that reference frames are a holistic representation.

Previous work (Schober, 1993, 1995) has shown that interlocutors will co-ordinate the reference object and origin of a reference frame to the matcher in a match-and-describe task. However, this did not show that interlocutors were aligning reference frames; as Levinson (2003) has argued, it is possible to have a non-egocentric relative reference frame and an egocentric intrinsic reference frame. The results presented here provide strong evidence that interlocutors do align reference frames. Ongoing work is investigating the predictions made by Levinson's definitions of reference frames that an egocentric/intrinsic description (e.g. *the ball in front of me*) can prime the use of an allocentric/intrinsic description (e.g. *the ball in front of the car*).

Previous work (Branigan et al., 2000; Clark & Wilkes-Gibbs, 1986; Garrod & Anderson, 1987) has shown that interlocutors align representations during dialogue. The results of these experiments extend this body of evidence to show that independent of lexical priming, alignment extends beyond the language faculty and that interlocutors also align reference frames to describe objects' locations in a scene.

## References


Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000) Syntactic co-ordination in dialogue. *Cognition,* 75, B13-B25.

Bryant, D. J., & Wright, W. G. (1999). How bodily asymmetries determine accessibility in spatial frameworks. *Quarterly Journal of Experimental Psychology,* 52A, 487-508.

Carlson-Radvansky, L. A., & Jiang, Y. (1998). Inhibition accompanies reference frame selection. *Psychological Science*, 9, 386-391.

Clark, H. H. (1996). *Using language.* Cambridge: Cambridge University Press.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition,* 22, 1-39.

Cleland, A. A., & Pickering, M. J. (2003). The use of lexical and syntactic information in language production: Evidence from the priming of noun phrase structure. *Journal of Memory and Language, 49*, 214-230.

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition,* 27, 181-218.

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*, *8*, 8-11.

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition,* 59, 91-117.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge MA: MIT Press.

Levinson, S. C. (1996). Frames of reference and Molyneux's question: Cross-linguistic evidence. In P. Bloom, M. Peterson, L. Nadel, & M. Garrett (Eds.), *Language and space,* pp. 109-169. Cambridge MA: MIT Press.

Levinson, S. C. (2003). *Space in language and cognition: Explorations in cognitive diversity.* Cambridge: Cambridge University Press.

Pickering, M. J., & Garrod, S. (in press). Towards a mechanistic theory of dialogue. *Behavioural and Brain Sciences.*

Schober, M. F. (1993). Spatial perspective-taking in conversation. *Cognition,* 47, 1-24.

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