2

Why Dialogue Methods are Important for Investigating Spatial Language

MATTHEW E. WATSON, MARTIN J. PICKERING, and HOLLY P. BRANIGAN

2.1 Introduction

There is now a voluminous literature on the use of spatial language (e.g. Bloom, Peterson, Nadel, and Garrett, 1996; Carlson, 2003; Coventry and Garrod, 2004; Landau and Jackendoff, 1996; Levinson, 2003). However, it is notable that most experimental research has investigated spatial language in a monologue context. This is of course no different from other areas of psycholinguistics. An isolated context makes experimental control much more straightforward. To the extent that most psycholinguists have shown any interest in dialogue, they tend to assume that language in dialogue can be understood from studying language in monologue (see Clark, 1996; Pickering and Garrod, 2004). Under this assumption, understanding how people produce and comprehend utterances in isolation will lead to the understanding of how people communicate. However, there is more to dialogue than just the production of an utterance and its passive comprehension. Dialogue also does not divide interlocutors neatly into a speaker and an addressee and there are few occasions when participants produce a long monologue. Similarly, many utterances produced in a dialogue are only interpretable in the context in which they arise. Dialogue often exhibits what, in isolation, would be considered less than perfect examples of language. Many utterances in a conversation are not even grammatically well-formed sentences, yet people still manage to understand each other. There is therefore a necessity to investigate language in dialogue as well as in traditional psycholinguistic environments.

Research using dialogue paradigms has shown that language use is a highly interactive activity. Thus, addressees play a role in aiding the speaker to produce utterances, rather than alternating between production and passive comprehension (e.g. Bavelas, Coates and Johnson, 2000). Together interlocutors produce descriptions of objects or situations which are exclusive to a given interaction.

For example, Clark and Wilkes-Gibbs (1986) had one participant, the director, describe tangram figures (abstract geometric shapes) to another participant, the matcher, who had to put her tangrams into an order described by the director. Over the course of the conversation the director's descriptions of repeated tangram figures became shorter, more definite, and quite idiosyncratic. In addition, interlocutors tend to converge on similar descriptions, as we shall see.

2.2 Interactive Alignment Model

Pickering and Garrod (2004, see also Garrod and Pickering, 2004) have proposed the interactive alignment model of dialogue. Under this account alignment is the basis for successful communication, meaning that successful communication is characterized by interlocutors having sufficiently similar representations. Over the course of a conversation interlocutors align their linguistic representations, which leads to alignment of relevant aspects of the situation model, so that they are both using the same underlying representations to produce utterances. In this case, situation models are a multidimensional representation of the situation under discussion (Zwaan and Radvansky, 1998). The situation model is active in working memory and is a dynamic representation of information about the main characters under discussion, time, space, causality, and so on. Over time the situation models of interlocutors converge and the interlocutors become aligned. The speaker's situation model is primarily a representation of his or her own information, not a representation of what the addressee is likely to know or what is shared between the speaker and addressee. However, when the speaker and addressee are well aligned, it is also a good representation of the addressee's state of knowledge.

Alignment is achieved via three processes: (1) an automatic mechanism of alignment involving priming at all levels of linguistic representation and percolation between these levels; (2) a mechanism that repairs alignment failure; (3) alignment via explicit reasoning and modelling of a partner's mental state. This last process is used as a last resort when the automatic alignment processes fail. The interactive alignment model therefore acknowledges that interlocutors use low level automatic methods of aligning their representations (in process 1) as well as explicit modelling (in process 3), but the emphasis is on the use of automatic priming mechanisms to build up implicit common ground, which is the information that is shared between the interlocutors (see also Pickering & Garrod, 2006; Garrod & Pickering, 2004, 2007).

The interactive alignment model contrasts with theories of dialogue which argue that interlocutors model (full) common ground, which is the information that the interlocutors believe is mutually known (e.g., Clark 1996). On such accounts, the interlocutors regularly update common ground (thus carefully

demarcating what is mutually known and what is exclusively known). In this account a speaker introduces information which the addressee tacitly accepts either through saying things such as *yeah*, *ok* or by providing a new contribution. If the information is accepted then it enters common ground. If the addressee queries the information, then the two work together to resolve the issue, so that the original information or its replacement can enter common ground.

Clark's (1996) concept of common ground between interlocutors contrasts with Pickering and Garrod's (2004) concept of shared information (or implicit common ground). Common ground refers to a representation that each interlocutor has which contains the information that is available to both. It is explicitly marked as common ground and is explicitly maintained and updated as a conversation progresses. Common ground is a separate representation of knowledge independent of each interlocutor's own knowledge. Conversely, shared information, as suggested by Pickering and Garrod (2004), is not maintained and updated in its own right. Instead, shared information is simply the part of the situation model that overlaps between two (or more) interlocutors. As a conversation progresses the amount of shared information increases through the automatic priming mechanism.

2.3 Alignment of Representations

Pickering and Garrod (2004) argued that interlocutors largely achieve alignment of their situation models via alignment of linguistic representations and that this is achieved via an automatic priming mechanism. There is a great deal of evidence that interlocutors do mimic each other and tend to produce utterances that are similar to their partner's utterances and that this occurs at many levels of representation (e.g. lexical, syntactic, or semantic). Garrod and Anderson (1987) had pairs of participants play a cooperative maze game, in which they took turns to describe their positions to each other. The descriptions that the participants produced suggested that they were aligning situation models via the alignment of linguistic representations. Although there were several different ways in which participants could describe their position in the maze, for example a path description such as *I'm one along, and five up*, or a coordinate system such as *I'm at B5* (the maze was presented in a grid format), they tended to use the same description scheme as each other. That is, there was much more consistency within than between pairs in the method of location description. Participants also aligned on other aspects of their descriptions—for example, how they named different components of the maze. Some pairs referred to the locations in the maze (which were squares on the screen) as *nodes* whilst others referred to them as *boxes*. Once again there was greater consistency within than between pairs. The assumption is that the different ways of referring to the maze reflect differences in the underlying situation models

of the interlocutors and that, by aligning their language, pairs aligned the situation models that they were using to produce utterances about the maze (see also Garrod and Doherty, 1994).

Alignment is also evident in interlocutors' use of syntactic structures. A speaker's choice of syntactic structure has been argued to reflect an individual's situation model (Goldberg, 1995) and the evidence shows that interlocutors also align syntactic structures. Branigan, Pickering, and Cleland (2000) had participants describe pictures showing a dative event (the transfer of an object from an agent to a patient) and showed that a participant's use of a double object phrase (DO; e.g. *The waiter giving the customer the food*) or prepositional object phrase (PO; e.g. *The waiter giving the food to the customer*) to describe a target picture was influenced by the structure used by a scripted confederate to describe a prime picture: participants were more likely to use a DO construction after hearing the confederate use a DO construction than after hearing the confederate use a PO construction. This sort of alignment has also been shown for other types of syntactic structure, such as complex noun phrases (e.g. *The sheep that is red* vs. *The red sheep*; Cleland and Pickering, 2003) and has been shown when alignment leads to the production of syntactically ambiguous phrases (Haywood, Pickering, and Branigan, 2005). There is also evidence that bilingual interlocutors align syntax between L1 and L2, suggesting that syntax is shared between languages (Hartsuiker, Pickering, and Veltkamp, 2004; Schoonbaert, Hartsuiker, and Pickering, 2007).

Interlocutors have also been shown to align syntactically in spontaneous speech. Gries (2005) analysed a speech corpus and found that different dative constructions were more likely to be preceded by dative constructions of the same form, even when many contributions intervened between the two dative constructions. This shows that alignment does occur in natural speech, and also that the alignment effect persists over a relatively long time (see also Bock, Dell, Chang, and Onishi, 2007)

Evidence from syntactic alignment studies has also demonstrated another important feature of the interactive alignment model, namely that alignment percolates between levels of representation. This is shown by increased levels of syntactic alignment when there is lexical repetition across the prime and the target. Branigan *et al.* (2000) found that participants were more likely to use the same syntactic structure (DO or PO) if the verb used in the target was the same as the verb used by the confederate (e.g. *give* and *give* vs. *give* and *pass*). For example, participants were more likely to say *The waiter giving the customer the food* after hearing the confederate say *The nurse giving the patient the pill* than after *The nurse passing the patient the pill*. Similarly, Cleland and Pickering (2003) found that the semantic relatedness of the nouns in the prime and targets affected the level of alignment for complex noun phrases (e.g. *sheep* and *goat* vs. *sheep* and *car*). For example, participants were more likely to describe a picture as *The sheep that's red* after hearing the confederate say *The goat that's red* than after hearing *The car*

that's red. In both of these cases, lexical alignment enhanced syntactic alignment (see also Schoonbaert *et al.*, 2007).

Interlocutors also align lexically, in that they tend to use the same word as each other to refer to the same item. Brennan and Clark (1996) showed that interlocutors align or converge on the use of the same word to describe an object (e.g. *pennyloafer* vs. *docksider*). Garrod and Anderson (1987) showed that players in their maze game tended to describe their positions in mazes by using the same terms as each other, and giving those terms the same meaning (for example, *level* to mean *row*, counting from the bottom). In addition interlocutors also align on several other aspects including accent and speech rate (Giles and Powesland, 1975) and non-linguistic factors such as foot rubbing, face touching and body posture (Chartrand and Bargh, 1999; Shockley, Santana, and Fowler, 2003).

2.4 Partner Specificity

Brennan and Clark (1996) argued that interlocutors form conceptual pacts, whereby they tacitly agree to use a specific term to refer to a particular entity. It follows that the pact may exist only for that pair of interlocutors (or even only for that conversation). Therefore, when people switch conversation partners they have to form new conceptual pacts. In accord with this, Brennan and Clark found that speakers were more likely to retain a particular term when they subsequently interacted with the same partner than with a different partner.

However, the extent to which partner-specificity occurs is controversial, at least in comprehension. Barr and Keysar (2002) found that participants in an experiment were faster to look at objects the second time the object was mentioned, but that it did not matter who mentioned the object the second time (that is, whether it was a new or the old speaker); see also Kronmüller and Barr (2007). In contrast, Metzing and Brennan (2003) used a similar method that measured eye movements and did find partner-specific effects. Their experiment included a condition where a new term was used to refer to a previously mentioned object. Metzing and Brennan found that the addressee took longer to look at the object when an old speaker used a new term than when a new speaker used a new term. They argued that an old speaker's use of a new term to refer to a previously mentioned object broke a conceptual pact. The addressee, therefore, assumed that the new term referred to a new object, which caused processing difficulty. However, a new speaker's use of a new term to refer to a previously mentioned object did not break a conceptual pact, and so the addressee did not experience difficulty.

It therefore appears that partner-specificity does affect comprehension and production to some extent. According to the simplest automatic account, linguistic representations are activated by comprehension or production, and thus any subsequent act of comprehension or production is equally facilitated. However,

Garrod and Pickering (2007) point out that the interactive alignment model can account for partner-specificity by drawing on Horton and Gerrig's (2005) claim that people associate the use of terms with a particular speaker via implicit memory. If a speaker uses a particular expression, the expression and the speaker become associated in the addressee's memory. Therefore, the mention of that expression also activates information about the person who used that expression; and talking with the person who originally used that expression activates that expression in memory. Such a mechanism does not require conceptual pacts to explain partner-specificity.

2.5 Perspective Alignment

Producing or comprehending a spatial description requires an individual to adopt a perspective on a scene. For example, if two people are viewing a scene of a knife and a fork from opposite sides, then from one person's perspective the knife is left of the fork, whereas from the other person's perspective the knife is right of the fork. However, the speakers are not restricted to taking their own perspective: both can adopt the other person's perspective, and produce a description which corresponds to that person's perspective. For communication to be successful, it is important that the addressee adopts the same perspective as the speaker. Thus, we would expect perspective alignment to occur, as part of the general process of alignment of situation models (e.g. Garrod and Anderson, 1987).

In fact, people appear to readily adopt another person's perspective when describing the location of objects. Schober (1993) used a task in which a director and a matcher each viewed, from different perspectives, a large circle containing two smaller circles. On the director's scene one of the smaller circles was marked. The director had to describe which of the two was marked so that the matcher could mark that circle on his or her scene. Because the director and the matcher were viewing the scene from different positions the director could choose to describe the scene from his or her own perspective (i.e. egocentrically) or from the matcher's perspective (i.e. allocentrically). Schober found that directors were more likely to use an allocentric (i.e. matcher-centred) perspective than an egocentric one, suggesting that they were trying to minimize effort for their partner. In addition, because feedback from the matchers was allowed in the experiment it was possible to identify which perspective the matchers used when they questioned the director further about which circle to mark. The results indicated that the matchers used predominantly a director-centred perspective, which for the matcher was also an allocentric perspective. Therefore, when the directors spoke they tended to use the perspective of their addressee (i.e. the matcher) and when the matchers spoke they also tended to use the perspective of their addressee (i.e. the director). In this experiment the participants tended to align on using their addressee's perspective. However, this meant that the person whose perspective

was used switched depending upon whether the matcher or director was speaking. Directors used an allocentric perspective more when they were describing object location to a partner who was not present and so the director could not receive any feedback.

In Schober's (1995) study, the results showed that directors often used the perspective of the matcher when describing the location of the relevant objects. Interestingly, however, the directors did not use their partner's perspective exclusively, but occasionally used their own perspective and most often used neutral descriptions that were not from any perspective (e.g. *near, between*). Even though directors changed perspectives during the experiment, matchers were able to successfully complete the task without difficulty, thereby indicating that the matchers were able to understand directors despite the varying perspectives that the latter adopted. Clearly, then, matchers were able to understand which perspective the director was adopting for a given utterance, and to interpret the utterance accordingly. But how is it that an addressee is able to understand which perspective a speaker is using to describe the location of objects? Similarly, how is it that a speaker decides which perspective to use when describing the location of an object to an addressee?

2.5.1 Reference frame parameters

Given that alignment occurs at multiple levels of representation, Watson, Pickering, and Branigan (2004) hypothesized that interlocutors should also align spatial representations, and specifically the perspective that they use to describe the locations of objects in a scene. The perspective that is used to describe an object's location is dependent upon the reference frame that a speaker imposes upon the scene. A reference frame defines an origin, orientation, direction, and scale (Logan and Sadler, 1996). The origin refers to the object (or part of object) that the reference frame is situated on; this is usually the reference object (or part of it). For example, if Figure 2.1 is described as *The dot is above the chair*, the origin of the reference frame is on the chair. The orientation parameter determines which axes of the reference frame are the top-bottom, left-right, and front-back axes. For example, the above description of Figure 2.1 uses an intrinsic reference frame (defined below), and therefore the reference frame is oriented so the top-bottom axis is aligned along the canonically vertical axis of the chair. In this case this is the axis which extends from the chair legs through the seat and up to the top of the back of the chair. The direction parameter is nested within the orientation parameter; it sets the directional endpoints of each of the axes after the orientation has been determined. Continuing with the above example, once the top-bottom axis has been oriented in reference to the chair, the direction of the axis is assigned. In this case, the end of the axis that coincides with the canonical top of the chair is labelled as the top and the opposite end is labelled as the bottom. Finally, the scale parameter sets the distance according to one of the components of the scene.

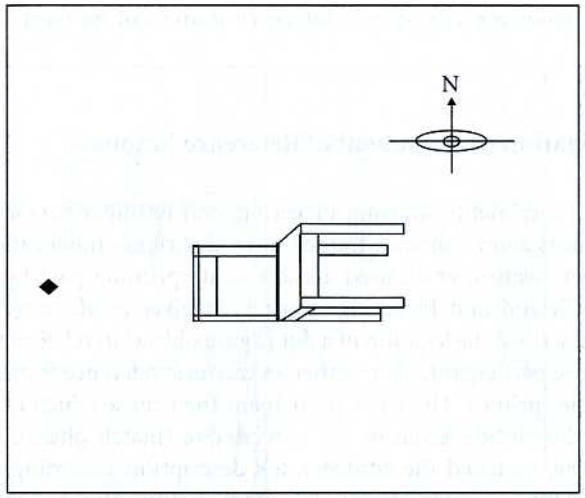

FIG. 2.1. The dot can be described as *west* of the chair using an absolute reference frame, *above* using an intrinsic reference frame, or *left* using a relative reference frame

The scale can be set according to the size of the chair (reference object), which is somewhat larger than the dot, or it can be set according to the size of the dot (figure object), which is the smaller of the two objects. Alternatively, both the figure and reference object may contribute to defining the scale parameter.

2.5.2 Reference frame types

The literature on spatial language has typically suggested there are three different types of reference frames: absolute, relative, and intrinsic (e.g. Levinson, 2003; Logan and Sadler, 1996). We illustrate these with respect to Figure 2.1.

Absolute reference frame. This refers to fixed features of the environment such as the points of the compass or gravity (or to directions such as downwind or inland in some cultures). In Figure 2.1, the dot can be described as *west of the chair.*

Intrinsic reference frame. The position of the figure in relation to the reference object is interpreted with respect to the actual orientation of the reference object. For example, the intrinsic meaning of *above* is (roughly) nearer to the top of the object than to any other part of it. As a chair has a top, the intrinsic reference frame allows the dot in Figure 2.1 to be described as *above the chair.*

Relative reference frame. The position of the figure object in relation to the reference object is interpreted with respect to the viewpoint of an observer. Using this reference frame, the dot in Figure 2.1 can be described as *to the left of the chair* (assuming the page is being held in a canonical fashion). In this case, the observer

is the reader. However, the relative reference frame can be used with different observers.

2.6 Investigation of Alignment of Reference Frames

In a series of experiments, Watson, Pickering, and Branigan (2004) have shown that interlocutors align reference frames when describing the location of objects to one another. Watson *et al.* used a confederate priming paradigm (Branigan *et al.*, 2000; Cleland and Pickering, 2003; Hartsuiker *et al.*, 2004) in which a confederate described the location of a dot (figure object) in relation to a reference object to a naïve participant, using either an intrinsic reference frame or a relative reference frame (prime). The naïve participant then chose which of two pictures matched the description given by the confederate (match phase). One picture, the *match scene*, matched the confederate's description according to either the intrinsic or relative reference frame; the other picture, the *distracter scene*, did not match the confederate's description according to either reference frame. The participant then described the location of a dot to the confederate (target), in the belief that the confederate had to choose which of two pictures matched their description. Several conditions for alignment have now been investigated in this way—of which two will be discussed here.

2.6.1 Within-axis alignment of reference frames

Experiment 1 investigated within-axis alignment of a reference frame between interlocutors (see also Carlson-Radvansky and Jiang, 1998). In this experiment the figure object was either in the same position on the target as the prime (same-preposition condition) or in the opposite position on the target as the prime (antonym-preposition condition). In the same-preposition condition, using the same reference frame as the confederate required the participant to use the same preposition as the confederate. For example, if the confederate described the match scene in Figure 2.2a as *The dot left of the chair* (using an intrinsic reference frame) and the participant used the same reference frame as the confederate, then the participant would describe the target as *The dot left of the chair.* In the antonym-preposition condition, using the same reference frame as the confederate required the participant to use the antonymous preposition. Therefore, if the confederate described the match scene in Figure 2.2b as *The dot right of the chair* (using an intrinsic reference frame) and the participant used the same reference frame as the confederate, then the participant would describe the target as *The dot left of the chair.* This situation was analogous for the other prepositions and the relative reference frame.

The results showed that participants used an intrinsic reference frame on 36.4% of trials after the confederate had used an intrinsic reference frame, but on only

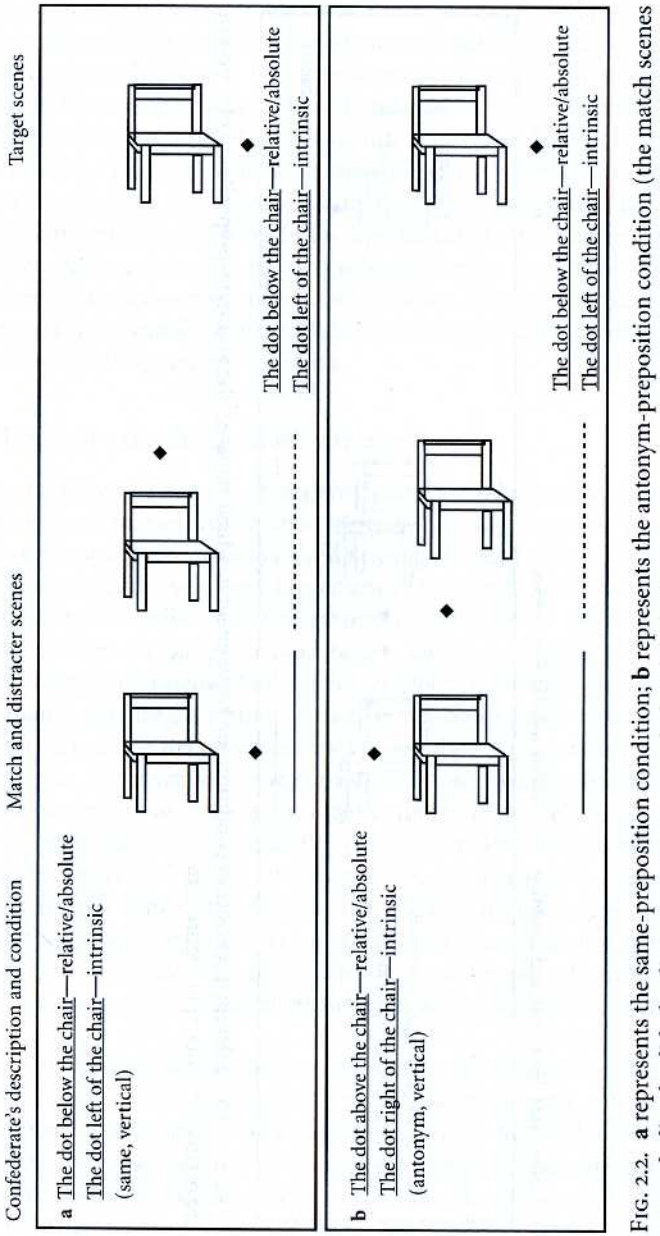

FIG. 2.2. **a** represents the same-preposition condition; **b** represents the antonym-preposition condition (the match scenes are underlined solid, the distracter scenes are underlined dashed)

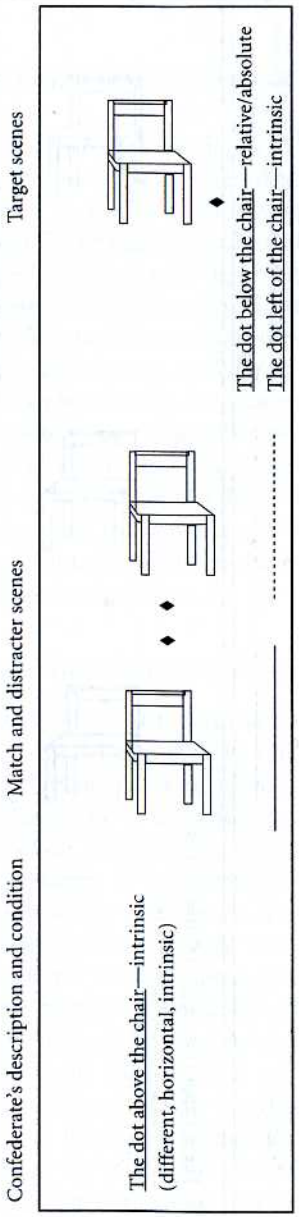

FIG. 2.3. The different-preposition condition for Experiment 2, which replaced the antonym-preposition condition in Experiment 1, shown in Figure 2.2b

26.9% of trials after the confederate had used a relative reference frame. This represents an alignment effect of approximately 10%. Hence participants were more likely to use a reference frame when it had just been used by the confederate than when the confederate had just used an alternative reference frame.

Experiment 1 only showed that interlocutors aligned single axes of reference frames (within-axis alignment). Thus, if one interlocutor used *above* in a relative reference frame, then a second interlocutor was more likely to use *above* in a relative reference frame or *below* in a relative reference frame; because *above* and *below* are on the same axis (vertical), alignment was within axes. It did not show whether alignment occurs between axes. Specifically, if one interlocutor uses *above* in a relative reference frame, would a second interlocutor be more likely to use *left* or *right* in a relative reference frame? Because *left* and *right* are located on a different axis (horizontal) to *above* and *below* (vertical), alignment in such a case would show that it occurs between axes.

2.6.2 Between-axis alignment of reference frames

Experiment 2 investigated whether participants would show between-axis alignment of reference frames, in addition to within-axis alignment of reference frames. In this experiment the confederate used intrinsic left and right in the same way as the participants in the previous experiment. One condition was identical to the same-preposition condition in Experiment 1, such that the figure object was in the same location on the prime and target (shown in Figure 2.2a). In another condition (different-preposition), the figure object was situated on different axes on the match and target scenes. Hence, if the figure object was located on a horizontal axis on the match scene, it was located on a vertical axis on the target, and vice versa. For example, if the confederate described the match in Figure 2.3 as *The dot above the chair* (using an intrinsic reference frame) and the participant used the same reference frame as the confederate, then the participant would describe the target as *The dot left of the chair*. If participants were more likely to use a reference frame after hearing the confederate use a reference frame in the different-preposition condition it would be strong evidence that interlocutors align the entire reference frame and not just the axes of a reference frame. This would also rule out lexical priming as an explanation of the alignment effect.

The results showed that participants used an intrinsic reference frame on 45.3% of trials after the confederate used an intrinsic reference frame, but on only 34.1% of trials after the confederate used a relative reference frame. As in Experiment 1, participants were more likely to use a reference frame if it had just been used by the confederate; furthermore, the effect was of a similar magnitude to Experiment 1. Hence, participants tended to align an entire reference frame. Importantly, in both experiments it was possible to establish that this was a true reference-frame priming effect and not a lexical priming effect. In both experiments, on half of the