1. INTRODUCTION

2. THE NATURE OF DIALOGUE AND THE ALIGNMENT OF REPRESENTATIONS
   2.1 Alignment of situation models is central to successful dialogue
   2.2 Achieving alignment of situation models
   2.3. Achieving alignment at other levels
   2.4 Alignment at one level leads to alignment at another
   2.5 Recovery from misalignment

3. THE INTERACTIVE ALIGNMENT MODEL OF DIALOGUE PROCESSING
   3.1. Interactive alignment versus autonomous transmission
   3.2. Channels of alignment
   3.3. Parity between comprehension and production

4. COMMON GROUND, MISALIGNMENT AND INTERACTIVE REPAIR
   4.1 Common Ground versus implicit common ground
   4.2 Limits on common ground inference
   4.3 Interactive repair using implicit common ground
   4.4 Interactive repair using full common ground

5. ALIGNMENT AND ROUTINIZATION
   5.1 Speaking: Not necessarily from intention to articulation.
   5.2 The production of routines
Short Abstract
Traditional mechanistic accounts of language processing derive almost entirely from the study of monologue. By contrast we propose a mechanistic account of dialogue, the interactive alignment account. It assumes that, in dialogue, the linguistic representations employed by the interlocutors become aligned at many levels, as a result of a largely automatic process. The process greatly simplifies production and comprehension in dialogue. It makes use of a simple interactive inference mechanism, enables the development of local dialogue routines that greatly simplify language processing, and explains the origins of self-monitoring in production.

Long Abstract
Traditional mechanistic accounts of language processing derive almost entirely from the study of monologue. Yet, the most natural and basic form of language use is dialogue. As a result, these accounts may only offer limited theories of the mechanisms that underlie language processing in general. We propose a mechanistic account of dialogue, the interactive alignment account, and use it to derive a number of predictions about basic language processes. The account assumes that, in dialogue, the linguistic representations employed by the interlocutors become aligned at many levels, as a result of a largely automatic process. This process greatly simplifies production and comprehension in dialogue. After considering the evidence for the interactive alignment model, we concentrate on three aspects of processing that follow from it. It makes use of a simple interactive inference mechanism, enables the development of local dialogue routines that greatly simplify language processing, and explains the origins of self-monitoring in production. We consider the need for a grammatical framework that is designed to deal with language in dialogue rather than monologue, and discuss a range of implications of the account.

Keywords: dialogue, language processing, common ground, dialogue routines, language production, monitoring.

1. INTRODUCTION
Psycholinguistics aims to describe the psychological processes underlying language use. The most natural and basic form of language use is dialogue: Every language user, including young children and illiterate adults, can hold a conversation, yet reading, writing, preparing speeches and even listening to speeches are far from universal skills. Therefore a central goal
of psycholinguistics should be to provide an account of the basic processing mechanisms that are employed during natural dialogue.

Currently, there is no such account. Existing mechanistic accounts are concerned with the comprehension and production of isolated words or sentences, or with the processing of texts in situations where no interaction is possible, such as in reading. In other words, they rely almost entirely on monologue. Thus, theories of basic mechanisms depend on the study of a derivative form of language processing. We argue that such theories are limited and inadequate accounts of the general mechanisms that underlie processing. In contrast, this paper outlines a mechanistic theory of language processing that is based on dialogue, but which applies to monologue as a special case.

Why has traditional psycholinguistics ignored dialogue? There are probably two main reasons, one practical and one theoretical. The practical reason is that it is generally assumed to be too hard or impossible to study, given the degree of experimental control necessary. Studies of language comprehension are fairly straightforward in the experimental psychology tradition - words or sentences are stimuli that can be appropriately controlled in terms of their characteristics (e.g., frequency) and presentation conditions (e.g., randomized order). Until quite recently it was also assumed that imposing that level of control in many language production studies was impossible. Thus, Bock (1996) points to the problem of "exuberant responding" - how can the experimenter stop subjects saying whatever they want? However, it is now regarded as perfectly possible to control presentation so that people produce the appropriate responses on a high proportion of trials, even in sentence production (e.g., Bock, 1986a; Levelt & Maassen, 1981).

Contrary to many people’s intuitions, the same is true of dialogue. For instance, Branigan, Pickering, and Cleland (2000) showed effects of the priming of syntactic structure during language production in dialogue that were exactly comparable to the priming shown in isolated sentence production (Bock, 1986b) or sentence recall (Potter & Lombardi, 1998). In Branigan et al.’s study, the degree of control of independent and dependent variables was no different from in Bock’s study, even though the experiment involved two participants engaged in a dialogue rather than one participant producing sentences in isolation. Similar control is exercised in studies by Clark and colleagues (e.g., Brennan & Clark, 1996; Wilkes-Gibbs & Clark, 1992; also Brennan & Schober, 2001; Horton & Keysar, 1996). Well-controlled studies of language production in dialogue may require some ingenuity, but such experimental ingenuity has always been a strength of psychology.

The theoretical reason is that psycholinguistics has derived most of its predictions from generative linguistics, and generative linguistics has developed theories of isolated, decontextualized sentences that are used in texts or speeches - in other words, in monologue. In contrast, dialogue is inherently interactive and contextualized: Each interlocutor both speaks and comprehends during the course of the interaction; each interrupts both others and himself; on occasion two or more speakers collaborate in producing the same sentence (Coates, 1990). So it is not surprising that generative linguists commonly view dialogue as being of marginal grammaticality, contaminated by theoretically uninteresting complexities. Dialogue sits ill with the competence/performance distinction assumed by most generative linguistics (Chomsky, 1965), because it is hard to determine whether a particular utterance is "well-formed" or not (or even whether that notion is relevant to dialogue). Thus, linguistics has tended to concentrate on developing generative grammars and related theories for isolated sentences; and psycholinguistics has tended to develop processing theories that draw upon the rules and representations assumed by generative linguistics. So far as most psycholinguists have thought about dialogue, they have tended to assume that the results of experiments on monologue can be applied to the understanding of dialogue, and that it is more profitable to
study monologue because it is "cleaner" and less complex than dialogue. Indeed, they have commonly assumed that dialogue simply involves chunks of monologue stuck together.

The main advocate of the experimental study of dialogue is Clark. However, his primary focus is on the nature of the strategies employed by the interlocutors rather than basic processing mechanisms. Clark (1996) contrasts the "language-as-product" and "language-as-action" traditions. The language-as-product tradition is derived from the integration of information-processing psychology with generative grammar and focuses on mechanistic accounts of how people compute different levels of representation. This tradition has typically employed experimental paradigms and decontextualized language; in our terms, monologue. In contrast, the language-as-action tradition emphasizes that utterances are interpreted with respect to a particular context and takes into account the goals and intentions of the participants. This tradition has typically considered processing in dialogue using apparently natural tasks (e.g., Clark, 1992; Fussell & Krauss, 1992). Whereas psycholinguistic accounts in the language-as-product tradition are admirably well-specified, they are almost entirely decontextualized and, quite possibly, ecologically invalid. On the other hand, accounts in the language-as-action tradition rarely make contact with the basic processes of production or comprehension, but rather present analyses of psycholinguistic processes purely in terms of their goals (e.g., the formulation and use of common ground; Clark, 1985; Clark, 1996; Clark & Marshall, 1981).

This dichotomy is a reasonable historical characterization. Almost all mechanistic theories happen to be theories of the processing of monologue; and theories of dialogue are almost entirely couched in intentional non-mechanistic terms. But this need not be. The goals of the language-as-product tradition are valid and important; but researchers concerned with mechanisms should investigate the use of contextualized language in dialogue.

In this paper we propose a mechanistic account of dialogue and use it to derive a number of predictions about basic language processing. The account assumes that in dialogue, production and comprehension become tightly coupled in a way that leads to the automatic alignment of linguistic representations at many levels. We argue that the interactive alignment process greatly simplifies language processing in dialogue as compared to monologue. It does so (1) by supporting a straightforward interactive inference mechanism, (2) by enabling interlocutors to develop and use routine expressions, and, (3) by supporting a system for monitoring language processing.

The first part of the paper presents the main argument (sections 2-6). In section 2 we show how successful dialogue depends on alignment of representations between interlocutors at different linguistic levels. In section 3 we contrast the interactive alignment model developed in section 2 with the autonomous transmission account that underpins current mechanistic psycholinguistics. Section 4 describes a simple interactive repair mechanism that supplements the interactive alignment process. We argue that this repair mechanism can re-establish alignment when interlocutors’ representations diverge without requiring them to model each other’s mental states. Thus interactive alignment and repair enable interlocutors to get around many of the problems normally associated with establishing what Stalnaker (1978) called common ground. The interactive alignment process leads to the use of routine or semi-fixed expressions. In section 5 we argue that such ‘dialogue routines’ greatly simplify language production and comprehension by short-circuiting the decision making processes. Finally in section 6 we discuss how interactive alignment enables interlocutors to monitor dialogue with respect to all levels at which they can align.

The second part of the paper explores implications of the interactive alignment account. In section 7 we discuss implications for linguistic theory. In section 8 we argue for a graded distinction between dialogue and monologue in terms of different degrees of coupling between speaker and listener. In section 9 we argue that the interactive alignment account
may have broader implications in terms of current developments in areas such as social interaction, language acquisition, and imitation more generally. Finally, in section 10 we enumerate the differences between the interactive alignment model developed in the paper and the more traditional autonomous transmission account of language processing.

2. THE NATURE OF DIALOGUE AND THE ALIGNMENT OF REPRESENTATIONS
Table 1 shows a transcript of a conversation between two players in a co-operative maze game (Garrod & Anderson, 1987). In this extract one player A is trying to describe his position to his partner B who is viewing the same maze on a computer screen in another room. The maze is shown in Figure 1.

Table 1. Example dialogue taken from Garrod and Anderson (1987).

1-----B: .... Tell me where you are?
2-----A: Ehm : Oh God (laughs)
3-----B: (laughs)
4-----A: Right : **two along from the bottom one up:**
5-----B: Two along from the bottom, which side?
6-----A: The left : going from left to right in the second box.
7-----B: You're in the second box.
8-----A: One up :(1 sec.) I take it we've got identical mazes?
9-----B: Yeah well : right, starting from the left, **you're one along:**
10-----A: Uh-huh:
11-----B: **and one up?**
12-----A: Yeah, and I'm trying to get to ...
13 [ 28 utterances later ]
41-----B: You are starting from the left, **you're one along, one up?** *(2 sec.)*
42-----A: **Two along**: I'm not in the first box, I'm in the second box:
43-----B: You're **two along**:
44-----A: **Two up** *(1 sec.)* counting the : if you take : the first box as being one up :
45-----B: *(2 sec.)* Uh-huh :
46-----A: Well : I'm **two along, two up** *(1.5 sec.)*
47-----B: Two up ? :
48-----A: Yeah *(1 sec.)* so I can move down one:
49-----B: Yeah I see where you are:

* The position being described in the utterances shown in bold is highlighted with an arrow in Figure 1. Colons mark noticeable pauses of less than 1 second.
At first glance the language looks disorganized. Many of the utterances are not grammatical sentences (e.g., only one of the first six contains a verb). There are occasions when production of a sentence is shared between speakers, as in (7-8) and (43-44). It often seems that the speakers do not know how to say what they want to say. For instance, A describes the same position quite differently in (4), "two along from the bottom one up," and (46), "two along, two up."

In fact the sequence is quite orderly so long as we assume that dialogue is a joint activity (Clark, 1996; Clark & Wilkes-Gibbs, 1986). In other words, it involves cooperation between interlocutors in a way that allows them to sufficiently understand the meaning of the dialogue as a whole; and this meaning results from these joint processes. In Lewis’s (1969) terms, dialogue is a game of cooperation, where both participants "win" if both understand the dialogue, and neither "wins" if one or both do not understand.

Conversational analysts argue that dialogue turns are linked across interlocutors (Sacks, Schegloff & Jefferson, 1974; Schegloff & Sacks, 1973). A question, such as (1) "Tell me where you are?", calls for an answer, such as (3) "Two along from the bottom and one up." Even a statement like (4) "Right, two along from the bottom two up," cannot stand alone. It requires either an affirmation or some form of query, such as (5) "Two along from the bottom, which side?" (Linnell, 1998). This means that production and comprehension processes become coupled. B produces a question and expects an answer of a particular type; A hears the question and has to produce an answer of that type. For example, after saying "Tell me where you are?" in (1), B has to understand "two along from the bottom one up" in (4) as a reference to A’s position on the maze; any other interpretation is ruled out. Furthermore, the meaning of what is being communicated depends on the interlocutors’ aggreement or consensus rather than on dictionary meanings (Brennan & Clark, 1996) and is subject to negotiation (Linnell, 1998, p.74). Take for example utterances (4-11) in the fragment shown above. In utterance (4), A describes his position as "Two along from the bottom and one up," but the final interpretation is only established at the end of the first exchange when consensus is reached on a rather different description by B (9-11) "You're one

Figure 1. Schematic representation of the maze being described in the conversation shown in Table 1. The arrow points to the position being described by the utterances marked in bold in the table.
along … and one up?" These examples demonstrate that dialogue is far more coordinated than it might initially appear.

At this point, we should distinguish two notions of coordination that have become rather confused in the literature. According to one notion (Clark, 1985), interlocutors are coordinated in a successful dialogue just as participants in any successful joint activity are coordinated (e.g., ballroom dancers, lumberjacks using a two-handed saw). According to the other notion, coordination occurs when interlocutors share the same representation at some level (Branigan et al., 2000; Garrod & Anderson, 1987). To remove this confusion, we refer to the first notion as coordination and the second as alignment. Specifically, alignment occurs at a particular level when interlocutors have the same representation at that level. Dialogue is a coordinated behavior (just like ballroom dancing). However, the linguistic representations that underlie coordinated dialogue come to be aligned, as we claim below.

We now argue six points: (1) that alignment of situation models (Zwaan & Radvansky, 1998) forms the basis of successful dialogue; (2) that the way that alignment of situation models is achieved is by a primitive and resource-free priming mechanism; (3) that the same priming mechanism produces alignment at other levels of representation, such as the lexical and syntactic; (4) that interconnections between the levels mean that alignment at one level leads to alignment at other levels; (5) that another primitive mechanism allows interlocutors to repair misaligned representations interactively; and (6) that more sophisticated and potentially costly strategies that depend on modeling the interlocutor’s mental state are only required when the primitive mechanisms fail to produce alignment. On this basis, we propose an interactive alignment account of dialogue in the next section.

2.1 Alignment of situation models is central to successful dialogue

A situation model is a multi-dimensional representation of the situation under discussion (Johnson-Laird, 1983; Sanford & Garrod, 1981; van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998). According to Zwaan and Radvansky, the key dimensions encoded in situation models are space, time, causality, intentionality, and reference to main individuals under discussion. They discuss a large body of research which demonstrates that manipulations of these dimensions affect text comprehension (e.g., people are faster recognizing that a word had previously been mentioned when that word referred to something that was spatially, temporally, or causally related to the current topic). Such models are assumed to capture what people are "thinking about" while they understand a text, and therefore are in some sense within working memory (they can be contrasted with linguistic representations on the one hand and general knowledge on the other).

Most work on situation models has concentrated on comprehension of monologue (normally, written texts) but they can also be employed in accounts of dialogue, with interlocutors developing situation models as a result of their interaction (Garrod & Anderson, 1987). More specifically, we assume that in successful dialogue, interlocutors develop aligned situation models. For example, in Garrod and Anderson, players aligned on particular spatial models of the mazes being described. Some pairs of players came to refer to locations using expressions like right turn indicator, upside down T shape, or L on its side. These speakers represented the maze as an arrangement of patterns or figures. In contrast, the pair illustrated in the dialogue in Table 1 aligned on a spatial model in which the maze was represented as a network of paths linking the points to be described to prominent positions on the maze (e.g., the bottom left corner). Pairs often developed quite idiosyncratic spatial models, but both interlocutors developed the same model (Garrod & Anderson, 1987; Garrod & Doherty, 1994; see also Markman & Makin, 1998).

Alignment of situation models is not necessary in principle for successful communication. It would be possible to communicate successfully by representing one’s interlocutor’s situation model, even if that model were not the same as one’s own. For instance, one player could
represent the maze according to a figure scheme but know that their partner represented it according to a path scheme, and vice versa. But this would be wildly inefficient as it would require maintaining two underlying representations of the situation, one for producing one’s own utterances and the other for comprehending one’s interlocutor’s utterances. Even though communication might work in such cases, it is unclear whether we would claim that the people understood the same thing. More critically, it would be computationally very costly to have fundamentally different representations. In contrast, if the interlocutors’ representations are basically the same, there is no need for listener modeling.

Under some circumstances storing the fact that one’s interlocutors represent the situation differently than oneself is necessary (e.g., in deception, or when trying to communicate to one interlocutor information that one wants to conceal from another). But even in such cases, many aspects of the representation will be shared (e.g., I might lie about my location, but would still use a figural representation to do so if that was what you were using). Additionally, it is clearly tricky to perform such acts of deception or concealment (Clark & Schaefer, 1987). These involve sophisticated strategies that do not form part of the basic process of alignment, and are difficult, because they require the speaker to concurrently develop two representations.

Of course, interlocutors need not entirely align their situation models. In any conversation where information is conveyed, the interlocutors must have somewhat different models, at least before the end of the conversation. In cases of partial misunderstanding, conceptual models will not be entirely aligned. In (unresolved) arguments, interlocutors have representations that cannot be identical. But they must have the same understanding of what they are discussing in order to disagree about a particular aspect of it (e.g., Sacks, 1987). For instance, if two people are arguing the merits of the Conservative versus the Labour parties for the U.K. government, they must agree about who the names refer to, roughly what the politics of the two parties are, and so on, so that they can disagree on their evaluations. In Lewis’s (1969) terms, such interlocutors are playing a game of cooperation with respect to the situation model (e.g., they succeed insofar as their words refer to the same entities), even though they may not play such a game at other "higher" levels (e.g., in relation to the argument itself). Therefore, we assume that successful dialogue involves approximate alignment at the level of the situation model at least.

2.2 Achieving alignment of situation models

In theory, interlocutors could achieve alignment of their models through explicit negotiation, but in practice they normally do not (Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986; Garrod & Anderson, 1987; Schober, 1993). It is quite unusual for people to suggest a definition of an expression and obtain an explicit assent from their interlocutor. Instead, "global" alignment of models seems to result from "local" alignment at the level of the linguistic representations being used. We propose that this works via a priming mechanism, whereby encountering an utterance that activates a particular representation makes it more likely that the person will subsequently produce an utterance that uses that representation. (On this conception, priming underpins the alignment mechanism and should not simply be regarded as a behavioral effect.) In this case, hearing an utterance that activates a particular aspect of a situation model will make it more likely that the person will use an utterance consistent with that aspect of the model. This process is essentially resource-free and automatic.

This was pointed out by Garrod and Anderson (1987) in relation to their principle of output/input coordination. They noted that in the maze game task speakers tended to make the same semantic and pragmatic choices that held for the utterances that they had just encountered. In other words, their outputs tended to match their inputs at the level of the situation model. As the interaction proceeded, the two interlocutors therefore came to align
the semantic and pragmatic representations used for generating output with the representations used for interpreting input. Hence, the combined system (i.e., the interacting dyad) is completely stable only if both subsystems (i.e., speaker A’s representation system and speaker B’s representation system) are aligned. In other words, the dyad is only in equilibrium when what A says is consistent with B’s currently active semantic and pragmatic representation of the dialogue and vice versa (see Garrod & Clark, 1993). Thus, because the two parties to a dialogue produce aligned language, the underlying linguistic representations also tend to become aligned. In fact, the output/input coordination principle applies more generally. Garrod and Anderson also assumed that it held for lexical representations. We argue that alignment holds at a range of levels, including the situational model and the lexical level, but also at other levels, such as the syntactic, as discussed in section 2.3, and “percolates” between levels, as discussed in section 2.4.

Other work suggests that specific dimensions of situation models can be aligned. With respect to the spatial dimension, Schober (1993) found that interlocutors tended to adopt the same reference frame as each other. When interlocutors face each other, terms like on the left are ambiguous depending on whether the speaker takes what we can call an egocentric or an allocentric reference frame. Schober found that if, for instance, A said on the left meaning on A’s left (i.e., an egocentric reference frame), then B would subsequently describe similar locations as on the right (also taking an egocentric frame of reference). Other evidence for priming of reference frames comes from experiments conducted outside dialogue (which involve the same priming mechanism on our account). Thus, Carlson-Radvansky and Jiang (1998) found that people responded faster on a sentence-picture verification task if the reference frame (in this case, egocentric vs. intrinsic to the object) used on the current trial was the same as the reference frame used on the previous trial.

So far we have assumed that the different components of the situation model are essentially separate (in accord with Zwaan & Radvansky, 1998), and that they can be primed individually. But in a particularly interesting study, Boroditsky (2000) found that the use of a temporal reference frame can be primed by a spatial reference frame. Thus, if people had just verified a sentence describing a spatial scenario that assumed a particular frame of reference (in her terms, ego moving or object moving), they tended to interpret a temporal expression in terms of an analogous frame of reference. Her results demonstrate priming of a structural aspect of the situation model that is presumably shared between the spatial and temporal dimensions at least. Indeed, work on analogy more generally suggests that it should be possible to prime abstract characteristics of the situation model (e.g., Markman & Gentner, 1993; Gentner & Markman, 1997), and that such processes should contribute to alignment in dialogue.

There is some evidence for alignment of situation models in comprehension. Garrod and Anderson (1987) found that players in the maze game would query descriptions from an interlocutor that did not match their own previous descriptions (see section 4 below). Recently, Brown-Schmitt, Campana, and Tanenhaus (in press) have provided direct and striking evidence for alignment in comprehension. Previous work has shown that eye movements during scene perception are a strong indication of current attention, and that they can be used to index the rapid integration of linguistic and contextual information during comprehension (Chambers, Tanenhaus, Eberhard, Filip, & Carlson, 2002; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Brown-Schmitt et al. monitored eye movements during unscripted dialogue, and found that the entities considered by the listener directly reflected the entities being considered by the speaker at that point. For instance, if the speaker used a referring expression that was formally ambiguous but which the speaker used to refer to a specific entity (and hence regarded as disambiguated), the listener also looked at that
entity. Hence, whatever factors were constraining the speaker’s situation model were also constraining the listener’s situation model.

2.3 Achieving alignment at other levels

Dialogue transcripts are full of repeated linguistic elements and structures indicating alignment at various levels in addition to that of the situation model (Aijmer, 1996; Schenkein, 1980; Tannen, 1989). Alignment of lexical processing during dialogue was specifically demonstrated by Garrod and Anderson (1987) as in the extended example in Table 1 (see also Garrod & Clark, 1993; Garrod & Doherty, 1994), and by Clark and colleagues (Brennan & Clark, 1996; Clark & Wilkes-Gibbs, 1986; Wilkes-Gibbs & Clark, 1992). These latter studies show that interlocutors tend to develop the same set of referring expressions to refer to particular objects and that the expressions become shorter and more similar on repetition with the same interlocutor and are modified if the interlocutor changes.

Levett and Kelter (1982) found that speakers tended to reply to "What time do you close?" or "At what time do you close" (in Dutch) with a congruent answer (e.g., "Five o’clock" or "At five o’clock"). This alignment may be syntactic (repetition of phrasal categories) or lexical (repetition of at). Branigan et al. (2000) found clear evidence for syntactic alignment in dialogue. Participants took it in turns to describe pictures to each other (and to find the appropriate picture in an array). One speaker was actually a confederate of the experimenter and produced scripted responses, such as "the cowboy offering the banana to the robber" or "the cowboy offering the robber the banana." The syntactic structure of the confederate’s description strongly influenced the syntactic structure of the experimental subject’s description. Their work extends "syntactic priming" work to dialogue: Bock (1986b) showed that speakers tended to repeat syntactic form under circumstances in which alternative non-syntactic explanations could be excluded (Bock, 1989; Bock & Loebell, 1990; Bock, Loebell, & Morey, 1992; Hartsuiker & Westenberg, 2000; Pickering & Branigan, 1998; Potter & Lombardi, 1998; cf. Smith & Wheeldon, 2001, and see Pickering & Branigan, 1999, for a review).

Branigan et al.’s (2000) results support the claim that priming activates representations and not merely procedures that are associated with production (or comprehension) - in other words, that the explanation for syntactic priming effects is closely related to the explanation of alignment in general. This suggests an important "parity" between the representations used in production and comprehension (see Section 3.2). Interestingly, Branigan et al. (2000) found very large priming effects compared to the syntactic priming effects that occur in isolation.

There are two reasons why this might be the case. First, a major reason why priming effects occur is to facilitate alignment, and therefore that they are likely to be particularly strong during natural interactions. In Branigan et al. (2000), participants respond at their own pace, which should be "natural," and hence conducive to strong priming. Second, we would be expect interlocutors to have their production systems highly activated even when listening, because they have to be constantly prepared to become the speaker, whether by taking the floor or simply making a backchannel contribution.

If syntactic alignment is due, in part, to the interactional nature of dialogue, then the degree of syntactic alignment should reflect the nature of the interaction between speaker and listener. As Clark and Schaeffer (1987; see also Schober & Clark, 1989; Wilkes-Gibbs & Clark, 1992) have demonstrated, there are basic differences between addressees and other listeners. So we might expect stronger alignment for addressees than for other listeners. To test for this, Branigan, Pickering, and Cleland (2002) had two speakers take it in turns to describe cards to a third person, so the two speakers heard but did not speak to each other. Priming occurred under these conditions, but it was weaker than when two speakers simply responded to each other. Hence, syntactic alignment is affected by speaker participation in dialogue. Although, we would claim, the same representations are activated under these conditions as during
dyadic interaction, the closeness of dyadic interaction means that it leads to stronger priming. For instance, we assume that the production system is active (and hence is ready to produce an interruption) when the addressee is listening to the speaker. By contrast, Branigan et al.’s (2002) side participant is not in a position to make a full contribution, and hence does not need to activate his production system to the same extent.

Alignment also occurs at the level of articulation. It has long been known that as speakers repeat expressions, articulation becomes increasingly reduced (i.e., the expressions are shortened and become more difficult to recognize when heard in isolation; Fowler & Housum, 1987). However, Bard et al. (2000) found that reduction was just as extreme when the repetition was by a different speaker in the dialogue as it was when the repetition was by the original speaker. In other words, whatever is happening to the speaker’s articulatory representations is also happening to their interlocutor’s. There is also evidence that interlocutors align accent and speech rate (Giles, Coupland, & Coupland, 1992; Giles & Powesland, 1975).

Finally, there is some evidence for alignment in comprehension. Levelt and Kelter (1982, Experiment 6) found that people judged question-answer pairs involving repeated form as more natural than pairs that did not; and that the ratings of naturalness were highest for the cases where there was the strongest tendency to repeat form. This suggests that speakers prefer their interlocutors to respond with an aligned form.

2.4 Alignment at one level leads to alignment at another

So far, we have concluded that successful dialogue leads to the development of both aligned situation models and aligned representations at all other linguistic levels. There are good reasons to believe that this is not coincidental, but rather that aligned representations at one level lead to aligned representations at other levels.

Consider the following two examples of influences between levels. First, Garrod and Anderson (1987) found that once a word had been introduced with a particular interpretation it was not normally used with any other interpretation in a particular stretch of maze-game dialogue. For instance, the word row could refer either to an implicitly ordered set of horizontal levels of boxes in the maze (e.g., with descriptions containing an ordinal like "I’m on the fourth row") or to an unordered set of levels (e.g., with descriptions that do not contain ordinals like "I’m on the bottom row"). Speakers who had adopted one of these local interpretations of row and needed to refer to the other would introduce a new term, such as line or level. Thus, they would talk of the fourth row and the bottom line, but not the fourth row and the bottom row (see Garrod & Anderson, 1987 p. 202). In other words, aligned use of a word seemed to go with a specific aligned interpretation of that word. Restricting usage in this way allows dialogue participants to assume quite specific unambiguous interpretations for expressions. Furthermore, if a new expression is introduced they can assume that it would have a different interpretation from a previous expression, even if the two expressions are "dictionary synonyms." This process leads to the development of a lexicon of expressions relevant to the dialogue (see section 5). What interlocutors are doing is acquiring new senses for words or expressions. To do this, they use the principle of contrast just like children acquiring language (e.g., E.V. Clark, 1993).

Second, it has been shown repeatedly that priming at one level can lead to more priming at other levels. Specifically, syntactic alignment (or "syntactic priming") is enhanced when more lexical items are shared. In Branigan et al.’s (2000) study, the stooge produced a description using a particular verb (e.g., the nun giving the book to the clown). Some experimental subjects then produced a description using the same verb; whereas other subjects produced a description using a different verb. Syntactic alignment was considerably enhanced if the verb was repeated (as also happens in monologue; Pickering & Branigan, 1998). Thus, interlocutors do not align representations at different linguistic levels independently.
Likewise, Cleland and Pickering (2002) found people tended to produce noun phrases like the sheep that’s red as opposed to the red sheep more often after hearing the goat that’s red than after the book that’s red. This demonstrates that semantic relations between lexical items enhance syntactic priming. These effects can be modeled in terms of a lexical representation outlined in Pickering and Branigan (1998). A node representing a word (i.e., its lemma; Levelt, Roelofs, & Meyer, 1999; cf. Kempen & Huijbers, 1983) is connected to nodes that specify its syntactic properties. So, the node for give is connected to a node specifying that it can be used with a noun phrase and a prepositional phrase. Processing giving the book to the clown activates both of these nodes and therefore makes them both more likely to be employed subsequently. However, it also strengthens the link between these nodes, on the principle that coactivation strengthens association. Thus, the tendency to align at one level, such as the syntactic, is enhanced by alignment at another level, such as the lexical. Cleland and Pickering’s (2002) finding demonstrates that exact repetition at one level is not necessary: the closer the relationship at one level (e.g., the semantic), the stronger the tendency to align at the other (e.g., the syntactic). Note that we can make use of this tendency to determine which specific levels are linked.

In comprehension, there is evidence for parallelism at one level occurring more when there is parallelism at another level. Thus, pronouns tend to be interpreted as coreferential with an antecedent in the same grammatical role (e.g., "William hit Oliver and Rod slapped him" is interpreted as Rod slapping Oliver) (Sheldon, 1974; Smyth, 1994). Likewise, the likelihood of a gapping interpretation of an ambiguous sentence is greater if the relevant arguments are parallel (e.g., "Bill took chips to the party and Susan to the game" is often given an interpretation where Susan took chips to the game) (Carlson, 2001). Finally, Gangé and Shoben (2002; cf. Gangé, 2001) found evidence that interpreting a compound as having a particular semantic relation (e.g., type of doctor in adolescent doctor) was facilitated by prior interpretation of a compound containing either the same noun or adjective which used the same relation (e.g., adolescent magazine or animal doctor). These effects have only been demonstrated in reading, but we would also expect them to occur in dialogue.

The mechanism of alignment and in particular the percolation of alignment between levels has a very important consequence that we discuss in section 5. Interlocutors will tend to align expressions at many different levels at the same time. When all levels are aligned, interlocutors will repeat each others’ expressions in the same way (e.g., with the same intonation). Hence, dialogue should be highly repetitive, and should make extensive use of fixed expressions. Importantly, fixed expressions should be established during the dialogue, so that they become dialogue routines.

### 2.5 Recovery from misalignment

Of course, these primitive processes of alignment are not foolproof. For example, interlocutors might align at a "superficial" level but not at the level of the situation model (e.g., if they both refer to John but do not realize that they are referring to different Johns). In such cases, interlocutors need to be able to appeal to other mechanisms in order to establish or reestablish alignment. The account is not complete until we outline such mechanisms, which we do in section 4 below. For now, we simply assume that such mechanisms exist and are needed to supplement the basic process of alignment.

### 3. THE INTERACTIVE ALIGNMENT MODEL OF DIALOGUE PROCESSING

The interactive alignment model assumes that successful dialogue involves the development of aligned representations by the interlocutors. This occurs by priming mechanisms at each level of linguistic representation, by percolation between the levels so that alignment at one level enhances alignment at other levels, and by repair mechanisms when alignment goes awry. Figure 2 illustrates the process of alignment in fairly abstract terms. It shows the levels...
of linguistic representation computed by two interlocutors and ways in which those representations are linked. Critically, Figure 2 includes links between the interlocutors at multiple levels.

Figure 2. A and B represent two interlocutors in a dialogue in this schematic representation of the stages of comprehension and production processes according to the interactive alignment model. The details of the various levels of representation and interactions between levels are chosen to illustrate the overall architecture of the system rather than to reflect commitment to a specific model.

In this section, we elucidate the figure in three ways. First, we contrast it with a more traditional "autonomous transmission" account, as represented in Figure 3, where multiple links between interlocutors do not exist. Second, we interpret these links as corresponding to channels whereby priming occurs. Finally, we argue that the bi-directional nature of the links means that there must be parity between production and comprehension processes.
Figure 3. A and B represent two interlocutors in a dialogue in this schematic representation of the stages of comprehension and production processes according to the autonomous transmission account. The details of the various levels of representation and interactions between levels are chosen to illustrate the overall architecture of the system rather than to reflect commitment to a specific model.

3.1. Interactive alignment versus autonomous transmission
In the autonomous transmission account, the transfer of information between producers and comprehenders takes place via decoupled production and comprehension processes that are "isolated" from each other (see Fig. 3). The speaker (or writer) formulates an utterance on the basis of his representation of the situation. Crudely, a non-linguistic idea or "message" is converted into a series of linguistic representations, with earlier ones being syntactic, and later ones being phonological. The final linguistic representation is converted into an articulatory program, which generates the actual sound (or hand movements) (e.g., Levelt, 1989). Each intermediate representation serves as a "way station" on the road to production — its significance is internal to the production process. Hence there is no reason for the listener to be affected by these intermediate representations.

In turn, the listener (or reader) decodes the sound (or movements) by converting the sound into successive levels of linguistic representation until the message is recovered (if the communication is successful). He then infers what the speaker (or writer) intended on the basis of his autonomous representation of the situation. So, from a processing point of view, speakers and listeners act in isolation. The only link between the two is in the information conveyed by the utterances themselves (Cherry, 1956). Each act of transmission is treated as a discrete stage, with a particular unit being encoded into sound by the speaker, being transmitted as sound, and then being decoded by the listener. Levels of linguistic representation are constructed during encoding and decoding, but there is no particular association between the levels of representation used by the speaker and listener. Indeed, there is even no reason to assume that the levels will be the same, nor that assumptions about the levels involved in comprehension should constrain those assumed in production, or vice versa. Hence, Figure 3 could just as well involve different levels of representation for speaker and listener.

The autonomous transmission model is not appropriate for dialogue because, in dialogue, production and comprehension processes are coupled (Garrod, 1999). In formulating an utterance the speaker is guided by what has just been said to him and in comprehending the utterance the listener is constrained by what he has just said, as in the example dialogue in Table 1. The interlocutors build up utterances as a joint activity (Clark, 1996), with interlocutors often interleaving production and comprehension tightly. They also align at many different levels of representation, as discussed in Section 2. Thus, in dialogue each level of representation is causally implicated in the process of communication and these intermediate representations are retained implicitly. Because alignment at one level leads to alignment at others, the interlocutors come to align their situation models and hence are able to understand each other. This follows from the interactive alignment model described in Figure 2, but is not reflected in the autonomous transmission account in Figure 3.

3.2. Channels of alignment
The horizontal links in Figure 2 correspond to channels by which alignment takes place. The communication mechanism used by these channels is priming. Thus, we assume that lexical priming leads to the alignment at the lexical level, syntactic priming leads to alignment at the syntactic level, and so on. Although fully specified theories of how such priming operates are not available for all levels, sections 2.2 and 2.3 described some of the evidence to support...
priming at these levels, and detailed mechanisms of priming are proposed in many of the papers referred to there. As an example, Branigan et al. (2000) provided an account of syntactic alignment in dialogue that involved priming of syntactic information at the lemma stratum. Because channels of alignment are bi-directional, the model predicts that if evidence is found for alignment in one direction (e.g., from comprehension to production) it should also be found for alignment in the other (e.g., from production to comprehension). Of course, the linguistic information conveyed by the channels is encoded in sound.

Critically, these channels are direct and automatic (as implied by the term "priming"). In other words, the activation of a representation in one interlocutor leads to the activation of the matching representation in the other interlocutor directly. There is no intervening "decision box" where the listener makes a decision about how to respond to the "signal." Although such decisions do of course take place during dialogue (see Section 4 below), they do not form part of the basic interactive alignment process which is automatic and largely unconscious. We assume that such channels are similar to the direct and automatic perception-behavior link that has been proposed to explain the central role of imitation in social interaction (Bargh & Chartrand, 1999; Dijksterhuis & Bargh, 2001).

Figure 2 therefore indicates how interlocutors can align in dialogue via the interactive alignment model. It does not of course provide an account of communication in monologue, but the goal of monologue is not to get to aligned representations. Instead, the listener attempts to obtain a specific representation corresponding to the speaker’s message, and the speaker attempts to produce the appropriate sounds that will allow the listener to do this. Moreover, in monologue (including writing), the speaker’s and the listener’s representations can rapidly diverge (or never align at all). The listener then has to draw inferences on the basis of his knowledge about the speaker, and the speaker has to infer what the listener has inferred (or simply assume that the listener has inferred correctly). Of course, either party could easily be wrong, and these inferences will often be costly. In monologue, the automatic mechanisms of alignment are not present (the consequences for written production are demonstrated in Traxler & Gernsbacher, 1992, 1993). It is only when regular feedback occurs that the interlocutors can control the alignment process.

The role of priming is very different in dialogue from monologue. In monologue, it can largely be thought of as an epiphenomenal effect, which is of considerable use to psycholinguists as a way of investigating representation and process, but of little importance in itself. However, our analysis of dialogue demonstrates that priming is the central mechanism in the process of alignment and mutual understanding. Thus dialogue indicates the important functional role of priming. In conclusion, we regard priming as underlying the links between the two sides of Figure 2, and hence the mechanism that drives interactive alignment.

3.3 Parity between comprehension and production

On the autonomous transmission account, the processes employed in production and comprehension need not draw upon the same representations (see Fig. 3). By contrast, the interactive alignment model assumes that the processor draws upon the same representations (see Fig. 2). This parity means that a representation that has just been constructed for the purposes of comprehension can then be used for production (or vice versa). This straightforwardly explains, for instance, why we can complete one another’s utterances (and get the syntax, semantics, and phonology correct; see section 7.1). It also serves as an explanation of why syntactic priming in production occurs when the speaker has only heard the prime (Branigan et al., 2000; Potter & Lombardi, 1998), as well as when he has produced the prime (Bock, 1986b; Pickering & Branigan, 1998).

The notion of parity of representation is controversial but has been advocated by a wide range of researchers working in very different domains (Calvert et al., 1997; Liberman & Whalen, 2000; MacKay, 1987; Mattingly & Liberman, 1988). For example, Goldinger (1998)
demonstrated that speech ‘shadowers’ imitate the perceptual characteristics of a shadowed word (i.e., their repetition is judged acoustically more similar to the shadowed word than to another production of the same word by the shadower). He argued that this vocal imitation in shadowing strongly suggests an underlying perception-production link at the phonological level.

Parity is also increasingly advocated as a means of explaining perception/action interactions outside language (Hommel, Müsseler, Aschersleben, & Prinz, 2001). We return to this issue in section 8. Note that parity only requires that the representations be the same. The processes leading to those representations need not be related (e.g., there is no need for the mapping between representations to be simply reversed in production and comprehension).

4. COMMON GROUND, MISALIGNMENT, AND INTERACTIVE REPAIR
In current research on dialogue, the key conceptual notion has been "common ground" which refers to background knowledge shared between the interlocutors (Clark & Marshall, 1981). Traditionally, most research on dialogue has assumed that interlocutors communicate successfully when they share a common ground, and that one of the critical preconditions for successful communication is the establishment of common ground (Clark & Wilkes-Gibbs, 1986). Establishment of common ground involves a good deal of modeling of one’s interlocutor’s mental state. In contrast, our account assumes that alignment of situation models follows from lower-level alignment, and is therefore a much more automatic process. We argue that interlocutors align on what we term an implicit common ground, and only go beyond this to a (full) common ground when necessary. In particular, interlocutors draw upon common ground as a means of repairing misalignment when more straightforward means of repair fail.

4.1 Common ground versus implicit common ground
Alignment between interlocutors has traditionally been thought to arise from the establishment of common, mutual, or joint knowledge (Lewis, 1969; McCarthy, 1990; Schiffer, 1972). Perhaps the most influential example of this approach is Clark and Marshall’s (1981) argument that successful reference depends on the speaker and the listener inferring mutual knowledge about the circumstances surrounding the reference. Thus, for a female speaker to be certain that a male listener understands what is meant by "the movie at the Roxy," she needs to know what he knows and what he knows that she knows, and so forth. Likewise, for him to be certain about what she means by "the movie at the Roxy," he needs to know what she knows and what she knows that he knows, and so forth. However, there is no foolproof procedure for establishing mutual knowledge expressed in terms of this iterative formulation, because it requires formulating recursive models of interlocutors’ beliefs (see Barwise, 1989; Clark, 1996, ch. 4; Halpern & Moses, 1990; Lewis, 1969). Therefore, Clark and Marshall (1981) suggested that interlocutors instead infer what Stalnaker (1978) called the common ground. Common ground reflects what can reasonably be assumed to be known to both interlocutors on the basis of the evidence at hand. This evidence can be non-linguistic (e.g., if both know that they come from the same city they can assume a degree of common knowledge about that city; if both admire the same view and it is apparent to both that they do so, they can infer a common perspective), or can be based on the prior conversation.

Even though inferring common ground is computationally more feasible than inferring the iterative formulation of mutual knowledge, it still requires the interlocutor to maintain a very complex situation model that reflects both his own knowledge and the knowledge that he assumes to be shared with his partner. To do this, he has to keep track of the knowledge state of the interlocutor in a way that is separate from his own knowledge state. This is a very stringent requirement for routine communication, in part because he has to make sure that this model is constantly updated appropriately (e.g., Halpern & Moses, 1990).
In contrast, the interactive alignment model proposes that the fundamental mechanism that leads to alignment of situation models is automatic. Specifically, the information that is shared between the interlocutors constitutes what we call an implicit common ground. When interlocutors are well aligned, the implicit common ground is extensive. Unlike common ground, implicit common ground does not derive from interlocutors explicitly modeling each other’s beliefs. Implicit common ground is therefore built up automatically and is used in straightforward processes of repair. Interlocutors do of course make use of (full) common ground on occasion, but it does not form the basis for alignment.

Implicit common ground is effective because an interlocutor builds up a situation model that contains (or at least foregrounds) information that the interlocutor has processed (either by producing that information or comprehending it). But since the other interlocutor is also present, he comprehends what the first interlocutor produces and vice versa. This means that both interlocutors foreground the same information, and therefore tend to make the same additions to their situation models. Of course each interlocutor’s situation model will contain some information that he is aware of but the other interlocutor is not, but as the conversation proceeds and more information is added, the amount of information that is not shared will be reduced. Hence the implicit common ground will be extended. Notice that there is no need to infer the situation model of one’s interlocutor.

This account predicts that speakers only automatically adapt their utterances when the information can be accessed from their own situation model. However, because access is from aligned representations which reflect the implicit common ground these adaptations will normally be helpful incidentally for the listener. This point was first made by Brown and Dell (1987), who noted that if speaker and listener have very similar representations of a situation, then most utterances that appear to be sensitive to the mental state of the listener may in fact be produced without reference to the listener. This is because what is easily accessible for the speaker will also be easily accessible for the listener. In fact, the better aligned speaker and listener are, the closer such an implicit common ground will be to the full common ground, and the less effort need be exerted to support successful communication.

Hence, we argue that interlocutors do not need to monitor and develop full common ground as a regular, constant part of routine conversation, as it would be unnecessary and far too costly. Establishment of full common ground is, we argue, a specialized and non-automatic process that is used primarily in times of difficulty (when radical misalignment becomes apparent). We now argue that speakers and listeners do not routinely take common ground into account during initial processing. We then discuss interactive repair, and suggest that full common ground is only used when simpler mechanisms are ineffective.

4.2 Limits on common ground inference

Studies of both production and comprehension in situations where there is no direct interaction (i.e., situations which do not allow feedback) indicate that language users do not always take common ground into account in producing or interpreting references. For example, Horton and Keysar (1996) found that speakers under time pressure did not produce descriptions that took advantage of what they knew about the listener’s view of the relevant scene. In other words, the descriptions were formulated with respect to the speaker’s current knowledge of the scene rather than with respect to the speaker and listener’s common ground. Keysar, Barr, Balin, and Paek (1998) found that in visually searching for a referent for a description listeners are just as likely to initially look at things that are not part of the common ground as things that are, and Keysar, Barr, Balin, and Bruner (2000) found that listeners initially considered objects that they knew were not visible to their conversational partner. In a similar vein, Brown and Dell (1987) showed that apparent listener-directed ellipsis was not modulated by information about the common ground between speaker and listener, but rather was determined by the accessibility of the information for the speaker alone (though cf.
Lockridge & Brennan, 2002, and Schober & Brennan, in press, for reservations). Finally, Ferreira and Dell (2000) found that speakers did not try to construct sentences that would make comprehension easy (i.e., by preventing syntactic misanalysis on the part of the listener).

Even in fully interactive dialogue it is difficult to find evidence for direct listener modeling. For example, it was originally thought that articulation reduction might reflect the speaker’s sensitivity to the listener’s current knowledge (Lindblom, 1990). However, Bard et al. (2000) found that the same level of articulation reduction occurred even after the speaker encountered a new interlocutor. In other words, degree of reduction seemed to be based only on whether the reference was given information for the speaker and not on whether it was part of the common ground. Additionally, speakers will sometimes use definite descriptions (to mark the referent as given information; Haviland & Clark, 1974) when the referent is visible to them, even when they know it is not available to their interlocutor (Anderson & Boyle, 1994).

Nevertheless, under certain circumstances interlocutors do engage in strategic inference relating to (full) common ground. As Horton and Keysar (1996) found, with less time pressure speakers often do take account of common ground in formulating their utterances. Keysar et al. (1998) argued, that listeners can take account of common ground in comprehension under circumstances in which speaker/listener perspectives are radically different (see also Brennan & Clark, 1996; Schober & Brennan, in press), though they proposed that this occurs at a later monitoring stage, in a process that they called perspective adjustment. More recently, Hanna, Tanenhaus, and Trueswell (2001) found that listeners looked at an object in a display less if they knew that the speaker did not know of the object’s existence than otherwise (see Nadig & Sedivy, 2002, for a related study with 5-6 year old children). These differences emerged during the earliest stages of comprehension, and therefore suggest that the strongest form of perspective adjustment cannot be correct. However, their task was repetitive and involved a small number of items, and listeners were given explicit information about the discrepancies in knowledge. Under such circumstances, it is not surprising that listeners develop strategies that may invoke full common ground. During natural dialogue, we predict that such strategies will not normally be used.

In conclusion, we have argued that performing inferences about common ground is an optional strategy that interlocutors employ only when resources allow. Critically, such strategies need not always be used, and most "simple" (e.g., dyadic, non-didactic, non-deceptive) conversation works without them most of the time.

4.3 Interactive repair using implicit common ground

Of course, the automatic process of alignment does not always lead to appropriately aligned representations. When interlocutors’ representations are not properly aligned, the implicit common ground is faulty. We argue that they employ an interactive repair mechanism which helps to maintain the implicit common ground. The mechanism relies on two processes: (1) checking whether one can straightforwardly interpret the input in relation to one’s own representation, and (2) when this fails, reformulating the utterance in a way that leads to the establishment of implicit common ground. Importantly, this mechanism is iterative, in that the original speaker can then pick up on the reformulation and, if alignment has not been established, reformulate further.

Consider again the example in Table 1. Throughout this section of dialogue A and B assume subtly different interpretations for two along. A interprets two along by counting the boxes on the maze, whereas B is counting the links between the boxes (see Fig. 1). This misalignment arises because the two speakers represent the meaning of expressions like two along differently in this context. In other words, the implicit common ground is faulty.
Therefore, the players engage in interactive repair, first by determining that they cannot straightforwardly interpret the input, and then by reformulation. The reformulation can be a simple repetition with rising intonation (as in 7), a repetition with an additional query (as when B says "two along from the bottom, which side?" in 5), or a more radical restatement (as when A reformulates "two along" as "second box" in 6). Such reformulation is very common in conversation and is described by some linguists as clarification request (See Ginzburg, 2001). None of these reformulations requires the speaker to take into account the listener’s situation model. They simply reflect failures to understand what the speaker is saying in relation to the listener’s own model.

They serve to throw the problem back to the interlocutor who can then attempt a further simple reformulation if he still fails to understand the description. For example, B says "…you’re one along, one up?" (41), which A reformulates as "Two along …" (42). Probably because of this reformulation, A then asks the clarification request "You’re two along." The cycle continues until the misalignment has been resolved in (44) when A is able to complete B’s utterance without further challenge (for discussion of such embedded repairs see also Jefferson, 1987). This repair process can be regarded as involving a kind of dialogue inference, but notice that it is externalized, in the sense that it can only operate via the interaction between the interlocutors. It contrasts with the kind of discourse inference that occurs during text comprehension (or listening to a speech), where the reader has to mentally infer the writer’s meaning (e.g., via a bridging inference; Haviland & Clark, 1974).

4.4 Interactive repair using full common ground
Interactive repair using implicit common ground is basic because it only relies on the speaker checking the conversation in relation to his own knowledge of the situation. Of course there will be occasions when a more complicated and strategic assessment of common ground may be necessary, most obviously when the basic mechanism fails. In such cases, the listener may have to draw inferences about the speaker (e.g., "She has referred to John; does she mean John Smith or John Brown? She knows both, but thinks I don’t know Brown, hence she probably means Smith."). Such cases may of course involve internalized inference, in a way that may have more in common with text comprehension than with most aspects of everyday conversation. But interlocutors may also engage in explicit negotiation or discussion of the situation models. This appears to occur in our example when A says "I take it we’ve got identical mazes" (8).

Use of full common ground is particularly likely when one speaker is trying to deceive the other or to conceal information (e.g., Clark & Schaefer, 1987), or when interlocutors deliberately decide not to align at some level (e.g., because each interlocutor has a political commitment to a different referring expression; Jefferson, 1987). Such cases may involve complex (and probably conscious) reasoning, and there may be great differences between people’s abilities (e.g., between those with and without an adequate ‘theory of mind’; Baron-Cohen, Tager-Flusberg, & Cohen, 2000). For example, Garrod and Clark (1993) found that younger children could not circumvent the automatic alignment process. Seven year old maze game players failed to introduce new description schemes when they should have done so, because they could not overcome the pressure to align their description with the previous one from the interlocutor. By contrast, older children and adults were twice as likely to introduce a new description scheme when they had been unable to understand their partner’s previous description. Whereas the older children could adopt a strategy of non-alignment when appropriate, the younger children seemed unable to do so. Our claim is that these strategic processes are overlaid on the basic interactive alignment mechanism. However, such strategies are clearly costly in terms of processing resources and may be beyond the abilities of less skilled language users.
The strategies discussed above relate specifically to alignment (either avoiding it or achieving it explicitly), but of course many aspects of dialogue serve far more complicated functions. Thus, a speaker can attempt to produce a particular emotional reaction in the listener by an utterance, or persuade the listener to act in a particular way or to think in depth about an issue (e.g., in expert-novice interactions). Likewise, the speaker can draw complex inferences about the mental state of the listener and can try to probe this state by interrogation. Thus, it is important to stress that we are proposing interactive alignment as the primitive mechanism underlying dialogue, not a replacement for the more complicated strategies that conversationalists may employ on occasion.

Nonetheless, we claim that normal conversation does not routinely require modeling the interlocutor’s mind. Instead, the overlap between interlocutors’ representations is sufficiently great that a specific contribution by the speaker will either trigger appropriate changes in the listener’s representation, or will bring about the process of interactive repair. Hence, the listener will retain an appropriate model of the speaker’s mind, because, in all essential respects, it is the listener’s representation as well.

Processing monologue is quite different in this respect. Without automatic alignment and interactive repair the listener can only resort to costly bridging inferences whenever he fails to understand anything. And, to ensure success, the speaker will have to design what he says according to what he knows about the audience (see Clark & Murphy, 1982). In other words, he will have to model the mind or minds of the audience. Interestingly, Schober (1993) found that speakers in monologue were more likely to adopt a listener-oriented reference frame than speakers in dialogue, and that this was costly. Because adopting the listener’s perspective can be very complex (e.g., if different members of the audience are likely to know different amounts), it is not surprising that people’s skill at public speaking differs enormously, in sharp contrast to everyday conversation.

5. ALIGNMENT AND ROUTINIZATION

The process of alignment means that interlocutors draw upon representations that have been developed during the dialogue. Thus it is not always necessary to construct representations that are used in production or comprehension from scratch. This perspective radically changes our accounts of language processing in dialogue. One particularly important implication is that interlocutors develop and use routines (set expressions) during a particular interaction. Most of this section addresses the implications of this perspective for language production, where they are perhaps most profound. We then turn more briefly to language comprehension.

5.1. Speaking: Not necessarily from intention to articulation.

The seminal account of language production is Levelt's (1989) book Speaking, which has the informative subtitle From intention to articulation. Chapter by chapter, Levelt describes the stages involved in the process of language production, starting with the conceptualization of the message, through the process of formulating the utterance as a series of linguistic representations (representing grammatical functions, syntactic structure, phonology, metrical structure, etc.), through to articulation. The core assumption is that the speaker necessarily goes through all of these stages in a fixed order. The same assumption is common to more specific models of word production (e.g., Levelt et al., 1999) and sentence production (e.g., Bock & Levelt, 1994; Garrett, 1980). Experimental research is used to back up this assumption. In most experiments concerned with understanding the mechanisms underlying language production, the speaker is required to construct the word or utterance from scratch, or from a pre-linguistic level at least. For example, a common method is picture description (e.g., Bock, 1986b; Schriefers, Meyer, & Levelt, 1990). These experiments therefore employ methods that reinforce the ideomotor tradition of action research that underlies Levelt’s framework (see Hommel et al., 2001).
It appears to be universally agreed that this exhaustive process is logically necessary because speakers have to articulate the words. Indeed, a common claim in work on language production is that, while comprehenders can sometimes "short-circuit" the comprehension process by taking into account the prior context (e.g., guessing thematic roles without actually parsing), producers always have to go through each step from beginning to end. To quote Bock and Huitema (1999, p.385):

"… there may be times when just knowing the words in their contexts is enough to understand the speaker, without a complete syntactic analysis of the sentence. But in producing a sentence, a speaker necessarily assigns syntactic functions to every element of the sentence; it is only by deciding which phrase will be the subject, which the direct object, and so on that a grammatical utterance can be formed — there is no way around syntactic processing for the speaker."

In fact, this assumption is wrong: It is logically just as possible to avoid levels of representation in production as in comprehension. Although we know that a complete output normally occurs in production, we do not know what has gone on at earlier stages. Thus, it is entirely possible, for example, that people do not always retrieve each lexical item as a result of converting an internally generated message into linguistic form (as assumed by Levelt et al, 1999, for example), but rather that people draw upon representations that have been largely or entirely formed already. Likewise, sentence production need not go through all the representational stages assumed by Garrett (1980), Bock and Levelt (1994), and others. For instance, if one speaker simply repeated the previous speaker's utterance, the representation might be taken "as a whole," without lexical access, formulation of the message, or computation of syntactic relations.

Repetition of an utterance may seem unnatural or uncertainly related to normal processing, but in fact, as we have noted, normal dialogue is highly repetitive (e.g., Tannen, 1989). This is of course different from carefully crafted monologue where — depending to some extent on the genre — repetition is regarded as an indication of poor style (see Amis, 1997, pp. 246-250). In our example dialogue 82% of the 127 words are repetitions; in this paragraph only 25% of the 125 words are repetitions. (Ironically, we — the authors — have avoided repetition even when writing about it.) In fact, the assumption that repetition is unusual or special is a bias probably engendered by psychologists' tendency to spend much of their time reading formal prose and designing experiments using decontextualized "laboratory" paradigms like picture naming.

So it is possible that people can short-circuit parts of the production process just as they may be able to short-circuit comprehension. Moreover, this may be a normal process that occurs when engaged in dialogue. We strongly suspect (see below) that phrases (for instance) are not simply inserted as a whole, but that the true picture is rather more complicated. But it is critical to make the logical point that the stages of production are not set in stone, as previous theories have assumed.

5.2. The production of routines

A routine is an expression that is "fixed" to a relatively great extent. First, the expression has a much higher frequency than the frequency of its component words would lead us to expect (e.g., Aijmer, 1996). (In computational linguistics this corresponds to having what is called a high "mutual information" content; Charniak, 1993.) Second, it has a particular analysis at each level of linguistic representation. Thus, it has a particular meaning, a particular syntactic analysis, a particular pragmatic use, and often particular phonological characteristics (e.g., a fixed intonation). Extreme examples of routines include repetitive conversational patterns such as How do you do? and Thank you very much. Routines are highly frequent in dialogue: Aijmer estimates that up to 70% of words in the London-Lund speech corpus occur as part of recurrent word combinations (see Altenberg, 1990). However, different expressions can be
routines to different degrees, so actual estimates of their frequency are somewhat arbitrary. Some routines are idioms, but not all (e.g., *I love you* is a routine with a literal interpretation in the best relationships; see Nunberg, Sag, & Wasow, 1994; Wray & Perkins, 2001). Most discussion of routines focuses on phrases whose status as a routine is pretty stable. Although long-term routines are important, we also claim that routines are set up "on the fly" during dialogue. In other words, if an interlocutor uses an expression in a particular way, it may become a routine for the purposes of that conversation alone. We call this process routinization. Here we consider why routines emerge and why they are useful. The next section considers how they are produced (in contrast to non-routines). This, we argue, leads to a need for a radical reformulation of accounts of sentence production. Finally, we consider how the comprehension of routines causes us to reformulate accounts of comprehension.

5.2.1. Why do routines occur?

Most stretches of dialogue are about restricted topics and therefore have quite a limited vocabulary. Hence, it is not surprising that routinization occurs in dialogue. But monologue can also be about restricted topics, and yet all indications suggest it is much less repetitive and routinization is much less common. The more interesting explanation for routinization in dialogue is that it is due to interactive alignment. A repeated expression (with the same analysis and interpretation) is of course aligned at most linguistic levels. Thus, if interlocutors share highly activated semantic representations (what they want to talk about), lexical representations (what lexical items are activated) and syntactic representations (what constructions are highlighted), they are likely to use the same expressions, in the same way, to refer to the same things. The contrast with most types of monologue occurs (in part, at least) because the producer of a monologue has no-one to align his representations with (see section 2). The use of routines contributes enormously to the fluency of dialogue in comparison to most monologue — interlocutors have a smaller space of alternatives to consider and have ready access to particular words, grammatical constructions, and concepts.

Consider the production of expressions that keep being repeated in a dialogue, such as "the previous administration." When first used, this expression is presumably constructed by accessing the meaning of "previous" and combining it with the meaning of "administration." The speaker may well have decided "I want to refer to the Conservative Government, but want to stress that they are no longer in charge, etc. so I’ll use a circumlocution." He will construct this expression by selecting the words and the construction carefully. Likewise, the listener will analyze the expression and consider alternative interpretations. Both interlocutors are therefore making important choices about alternative forms and interpretations. But if the expression is repeatedly used, the interlocutors do not have to consider alternatives to the same extent. For example, they do not have to consider that the expression might have other interpretations, or that "administration" is ambiguous (e.g., it could refer to a type of work). Instead, they treat the expression as a kind of name that refers to the last Conservative government. Similar processes presumably occur when producing expressions that are already frozen (Pinker & Birdsong, 1979; see also Aijmer, 1996). Generally, the argument is that people can "shortcircuit" production in dialogue, by removing or drastically reducing the choices that otherwise occur during production (e.g., deciding which synonym to use, or whether to use an active or a passive). Why might this happen? The obvious explanation is that routines are in general easier to produce than non-routines. Experimental work on this is lacking, but an elegant series of field studies by Kuiper (1996) suggest that this explanation is correct. He investigated the language of sports commentators and auctioneers, who are required to speak extremely quickly and fluently. For example, radio horse-racing commentators have to produce a time-locked and accurate monologue in response to rapidly changing events. This monologue is highly repetitive and stylized, but quite remarkably fluent. He argued that the commentators achieve
this by storing routines, which can consist of entirely fixed expressions (e.g., they are coming round the bend) or expressions with an empty slot that has to be filled (e.g., X is in the lead), in long-term memory, and then access these routines, as a whole, when needed. Processing load is thereby greatly reduced in comparison to non-routine production. Of course, this reduction in load is only possible because particular routines are stored; and these routines are stored because the commentators repeatedly produce the same small set of expressions in their career.

Below, we challenge his assumption that routines are accessed "as a whole," and argue instead that some linguistic processing is involved. But we propose a weaker version of his claims, namely that routines are accessed telegraphically, in a way that is very different from standard assumptions about language production (as in, e.g., Levelt, 1989). Moreover, we argue that not all routines are learned over a long period, but that they can instead emerge "on the fly," as an effect of alignment during dialogue.

5.2.2. Massive priming in language production

Contrary to Kuiper (1996), some compositional processes take place in routines, as we know from the production of idiom blends (e.g., that's the way the cookie bounces) (Cutting & Bock, 1997). However, there are good reasons to assume that production of idioms and other routines may be highly telegraphic. The normal process of constructing complex expressions involves a large number of lexical, syntactic, and semantic choices (why choose one word or form rather than another, for instance). In contrast, when a routine is used, most of these choices are not necessary. For example, speakers do not consider the possibility of passivizing an idiom that is normally active (e.g., The bucket was kicked), so there is no stage of selection between active and passive. Likewise, they do not consider replacing a word with a synonym (e.g., kick the pail), as the meaning would not be preserved. Similarly, a speech act like I name this ship X is fixed, insofar as particular illocutionary force depends on the exact form of words (cf. I give this ship the name X). Also, flat intonation suggests that no choices are made about stress placement (Kuiper, 1996).

Let us expand this by extending some of the work of Potter and Lombardi to dialogue (Lombardi & Potter, 1992; Potter & Lombardi, 1990, 1998). They address the question of how people recall sentences (see also Bock, 1986b, 1996). Recall differs from dialogue in that (1) the same sentences are perceived and produced; and (2) there is only one participant, acting as both comprehender and producer. Potter and Lombardi had experimental subjects read and then recall sentences whilst performing concurrent tasks. They found that a "lure" word sometimes intruded into the recalled sentence, indicating that subjects did not always store the surface form of the sentence; that these lure words caused the surface syntax of the sentence to change if they intruded and did not fit with the sentence that was read; and that other clauses could syntactically prime the target sentence so that it was sometimes misremembered as having the form of the prime sentence. They argued that people did not remember the surface form of the sentence but rather remembered its meaning and had the lexical items and syntactic constructions primed during encoding. Recall therefore involved converting the meaning into the surface form using the activation of lexical items and syntax to cause a particular form to be regenerated. In normal sentence recall, this is likely to be the form of the original sentence.

This suggests that language production can be greatly enhanced by the prior activation of relevant linguistic representations (in this case, lexical and syntactic representations). In dialogue, speakers do not normally aim simply to repeat their interlocutors’ utterances. However, production will be greatly enhanced by the fact that previous utterances will activate their syntactic and lexical representations. Hence, they will tend to repeat syntactic and lexical forms, and therefore to align with their interlocutors. These arguments suggest why sentence recall might actually present a reasonable analogue to production in naturalistic
dialogue; and why it is probably a better analogue than, for instance, isolated picture
description. In both sentence recall and production in dialogue, very much less choice needs
to be made than in monologue. The decisions that occur in language production (e.g., choice
of word or structure) are to a considerable extent driven by the context and do not need to be a
burden for the speaker. Thus, they are at least partly stimulus-driven rather than entirely
internally generated, in contrast to accounts like Levelt (1989).

However, our account differs from Potter and Lombardi's in one respect. They assume no
particular links between the activation of syntactic information, lexical information, and the
message. In other words, the reason that we tend to repeat accurately is that the appropriate
message is activated, the appropriate words are, and the appropriate syntax is. But we have
already argued that alignment at one level leads to more alignment at other levels (e.g.,
syntactic priming is enhanced by lexical overlap; Branigan et al., 2000). The alignment model
assumes interrelations between all levels, so that a meaning, for instance, is activated at the
same time as a word. This explains why people not only repeat words but also repeat their
senses in a dialogue (Garrod & Anderson, 1987). In other words, what actually occurs in
dialogue is lots of lexical, syntactic, and semantic activation of various tokens at each level,
and activation of particular links between the levels. This leads to a great deal of alignment,
and hence the production of routines. It also means that the production of a word or utterance
in dialogue is only distantly related to the production of a word or utterance in isolation.

Kuiper (1996) assumes that most routines are stored after repeated use, in a way that is not
directly related to dialogue. However, he considers an example of how an auctioneer creates a
"temporary formula" by repeating a phrase (p. 62). He regards this case as exceptional and
does not employ it as part of his general argument. In contrast, we assume that the
construction of temporary formulae is the norm in dialogue. Many studies show how new
descriptions become established for the dialogue (e.g., Brennan & Clark, 1996; Clark &
Wilkes-Gibbs, 1986; Garrod & Anderson, 1987). In general it is striking how quickly a novel
expression can be regarded as entirely normal, whether it is a genuine neologism or a novel
way of referring to an object (Gerrig & Bortfeld, 1999).

In situations in which a community of speakers regularly discuss the same topic we might
expect the transient routines that they establish to eventually become fixed within that
community. In fact, Garrod and Doherty (1994) demonstrated that an experimentally
established community of maze-game players quickly converged on a common description
scheme. They also found that the scheme established by the community of players was used
more consistently than schemes adopted by isolated pairs of players over the same period.
This result points to the interesting possibility that the interactive alignment process can be
responsible for fixing routines in the language or dialect spoken by a community of speakers
(see Clark, 1998).

5.2.3 Producing words and sentences

Most models of word production assume that the apparent fluency of production hides a
number of stages that lead from conceptual activation to articulation. In Levelt et al. (1999) a
lexical entry consists of sets of nodes at different levels (or strata): a semantic representation,
a syntactic (or lemma) representation, a phonological representation, a phonetic
representation, and so on. Each level is connected to the one after it, so that the activation of a
semantic representation (e.g., for cat) leads to the activation of its syntactic representation (the
"cat" lemma plus syntactic information specifying that it is a singular count noun), which in
turn leads to the access of the phonological representation /k/\a/t/. Evidence for the sequential
nature of activation comes from time-course data (Schriefers et al., 1990; van Turennout,
Hagoort, & Brown, 1998), "tip-of-the-tongue" data (Vigliocco, Antonini, & Garrett, 1997),
and so on. Alternative accounts question the specific levels assumed by Levelt et al. and the
mechanisms of activation but do not question the assumption that earlier levels become
activated before later ones (Caramazza, 1997; Dell, 1986). Notice that the data used to derive these accounts is almost entirely based on paradigms that require generation from scratch (e.g., picture naming) or from linguistic information with a very indirect relationship to the actual act of production required (e.g., responding with the object of a definition). We do not contend that the dialogical perspective leads us to a radically different view of word production. More specifically, we have no reason to doubt that the same levels of representation are accessed in the same order during production in dialogue (though this question has not been addressed by mainstream psycholinguistic research). For example, Potter and Lombardi’s data suggest that even in repetition of a word, it is likely that lexical access occurs (and that there is no direct access of the word-form, for example). However, contextual activation is likely to have some effects on the time-course of production, particularly in relation to the decisions at different stages in the production process. For example, a choice between two synonyms might normally involve some processing difficulty, but if one has been established in the dialogue (e.g., by lexical entrainment), no meaningful process of selection is needed.

The situation is very different with isolated sentence production. Models of production assume that a speaker initially constructs a message, then converts this message into a syntactic representation, then into a phonological representation, and then into sound (Bock & Levelt, 1994; Garrett, 1980; Levelt, 1989). Normally, they also assume that the syntactic level involves at least two stages: a functional representation, and a constituent-structure representation. It is accepted that cascading may happen, so that the complete message does not need to be computed before syntactic encoding can begin (e.g., Meyer, 1996). But ordering is assumed, so that, for instance, a word cannot be uttered until it is assigned a functional role and a position within a syntactic representation. However, we propose that it may be possible to break this rigid order of sentence production, and instead to build a sentence "around" a particular phrase if that phrase has been focused in the dialogue. In accord with this, context can affect sentence formulation in monologue, so that a focused phrase is produced first (Bock, 1986a; Prat-Sala & Branigan, 2000). Prat-Sala and Branigan, in particular, found effects of focus on word order that were not due to differences in grammatical role. Hence it may be possible to utter a phrase before assigning it a grammatical role. For instance, in Pictures, I think you like, and Pictures, I think please you, the meaning of Pictures does not vary but its grammatical role (subject or object) does vary. Assuming that production is at least partially incremental, people can therefore utter Pictures before deciding which role it should be given. This would of course not be possible within traditional models where phonological representations and acoustic form cannot be constructed before grammatical role is assigned (e.g., Bock & Levelt, 1994). So the effects of strong context, in either dialogue or monologue, may be to change the process of sentence production quite radically.

5.3 Alignment in comprehension

The vast literature on lexical comprehension is almost entirely concerned with monologue (e.g., reading words in sentential or discourse contexts) or isolated words. But the alignment model suggests that lexical comprehension in dialogue is very different from monologue. A major consequence of alignment at a lexical level is that local context becomes central. Listeners, just like speakers, should be able to select words from a set that have been central to that dialogue — a "dialogue lexicon."

One of the most universally accepted phenomena in experimental psychology, which is enshrined in all classic models (e.g., Morton, 1969) is the word frequency effect: More frequent words are understood and produced faster than less frequent words. Of course, processing is affected by repetition but this is normally regarded as only modulating the underlying frequency effect. However, in dialogue, local context is so central that the
frequency of an expression (or, e.g., its age of acquisition) should become far less important. To a large extent, frequency is replaced by accessibility with respect to the dialogue context. In contrast, the analogous context in monologue does not lead to alignment and there is a strong tendency to avoid repetition in many genres (e.g., formal writing) so the value of local context will be much less. Frequency is central to comprehension of monologue because it is what people fall back on if they have no strong context. So a prediction of our account is that frequency effects will be dramatically reduced in dialogue.

With respect to lexical ambiguity, we predict that context will have a very strong role, so that effects of meaning frequency can be overridden. Most current theories of lexical ambiguity resolution follow Swinney (1979) in assuming that multiple meanings of an ambiguous word are accessed in a bottom-up manner, largely irrespective of context. Similarly, differences in frequency do not affect access, unless perhaps one meaning is highly infrequent (see Balota, Paul, & Spieler, 1999 and Moss & Gaskell, 1999, for discussion). But in dialogue, only the contextually relevant meaning may be activated (or, in a modular account, the irrelevant meaning may always be suppressed rapidly). Hence, an interlocutor will straightforwardly adopt the appropriate meaning. An implication is that dialogue context should allow "subordinate bias effect" to be overridden (Duffy, Morris, & Rayner, 1988). According to Duffy et al., context can support the less frequent meaning and make it as accessible as the more frequent meaning, but it cannot cause the less frequent meaning to become more accessible than the more frequent meaning (Binder & Rayner, 1998; cf. Kellas & Vu, 1999; Rayner, Pacht, & Duffy, 1994). Although this may be true for reading (and monologue processing generally), it may not hold for dialogue.

The comprehension of routines is in a sense like lexical comprehension, in that their "frequency" and interpretation is set by the dialogue. However, this effect is in fact so strong that it appears to occur in monologue comprehension as well. A great deal of work is concerned with the comprehension of novel compounds in isolation (e.g., Murphy, 1988; Wisniewski, 1996), and the interpretations assigned depend on specific aspects of the words combined. Strong discourse contexts appear to enable direct access to infrequent interpretations of compound nouns such as baseball smile in reference to the smile of a boy given a baseball (Gerrig & Bortfeld, 1999). This would indicate that people can also "shortcircuit" the normal access to the individual nouns in a compound when there is a restricted meaning available from the immediate context.

6. SELF-MONITORING

The autonomous transmission model assumes that the speaker constructs a message, formulate an utterance as a series of linguistic representations and then articulate it as sound; and the listener then hears the message, converts it into linguistic representations and then comprehends it. The interlocutors (ideally) end up with the same semantic representation, and alignment at other levels is a derivative process (if it ever occurs at all). In contrast, Figure 2 proposes that interlocutors align themselves at different levels simultaneously via the automatic channels, and the parity assumption insures that the same representations are used in production and comprehension. Self-monitoring uses the same mechanism of alignment, but within the speaker.

All models assume that speakers monitor their own output, so that, for instance, they are able to interrupt their productions in order to change what they say (Hartsuiker & Kolk, 2001; Levelt, 1983, 1989). This can occur either before or after they start to produce a word. According to Levelt, speakers monitor their own productions by using the comprehension system (cf. Postma, 2000, for discussion of alternatives). They can monitor their actual outputs, in which case comprehension proceeds in an essentially normal way. According to a model that only contained this outer loop, monitoring would fit straightforwardly into the
autonomous transmission model shown in Figure 3. The only difference would be that both interlocutors were the same person. However, Levelt assumed the existence of an inner loop as well, which acts upon the phonological representation according to Wheeldon and Levelt (1995). Additionally, Levelt assumes that monitoring can occur within conceptualization, in order to make so-called "appropriateness repairs," for instance. It is impossible to include "inner" monitoring straightforwardly within the autonomous transmission model, because the monitor acts upon a representation that the interlocutor cannot act upon. From another perspective, it is unclear how the inner loop or the loop within the conceptualizer should have developed, given that they bear no relationship to any process involved in comprehending one’s interlocutor. The postulation of a monitor that uses the comprehension system is parsimonious (and it is easy to see how it could have evolved), but the postulation of special routes from production to comprehension that serve no other purpose is not.

In contrast, the inner loop and the loop within the conceptualizer fit straightforwardly into the interactive alignment model. Interlocutors are affected by each others' semantic and phonological representations via the channels of alignment represented in Figure 2. Hence a speaker can also be affected by his own representations at these levels. Self-monitoring is therefore compatible with Figure 2, except that A and B now refer to the same person (regarded as producer and comprehender). However, there is an important difference between interacting with oneself and interacting with an interlocutor. When interacting with an interlocutor, the information conveyed by the channels is encoded as sound. But when interacting with oneself, there is no need to encode the information as sound (indeed, the existence of internal monitoring proves that this is not necessary).

Given the existence of such levels of representation, there is no reason why the speaker should not automatically monitor at these levels. We propose that the speaker performs monitoring at these different levels in a way that leads to self-alignment. When the speaker produces an error at (say) the syntactic level (e.g., by selecting the wrong lemma), the result is a lack of alignment between the intended representation and the representation available to the monitor. This will become apparent as the levels of representation are traversed. For example, if a speaker accesses the semantic and syntactic forms of "dog" in order to utter it but wrongly accesses the phonological form of "cat," he will monitor this form, and then access its syntactic and semantic representations. Because these do not match the representation that he has accessed during production, the speaker will realize his error and (normally) attempt to correct himself. If he detects the mismatch and begins to correct himself before articulation begins, the repair will be covert; if not, some or all of "dog" will be produced. Self-correction involves a repair process that is essentially similar to the straightforward repair process used during interaction (see section 4.3). As the speaker’s production and comprehension systems draw upon the same implicit common ground, this repair process will tend to be successful, and hence there is normally no need to make reference to full common ground in self-monitoring.5

The interactive alignment model makes the very interesting prediction that monitoring can occur at any level of linguistic representation that can be aligned. For example, we predict the existence of syntactic monitoring. Consider the misassignment of syntactic gender and its subsequent detection. Speakers clearly can begin to say *Le tête* and then correct to *La tête*. This detection could occur externally or via the phonological channel. But an important prediction of this account is that monitoring (and the correction of errors) can also occur at the syntactic level (e.g., correcting gender, count/mass errors, errors of auxiliary selection, or errors of subcategorization), and at other levels as well. One reason for suspecting that this might be correct is that "other monitoring" (i.e., detecting errors in others’ speech) appears faster for phonological than syntactic errors (Oomen & Postma, 2002). If self-monitoring of syntax occurred via the phonological loop, we would predict that it would be slow in
comparison to self-monitoring of phonological errors. But we know of no evidence for this claim.

More generally, the existence of monitoring appears to be a consequence of dialogue. In dialogue, interlocutors have to switch between speaking and listening rapidly and repeatedly, and interlocutors have to be able to listen and plan their next utterance at the same time (otherwise the lack of pauses, for instance, could not be explained). The obvious way in which this can occur is for interlocutors to be listening at all times, with that listening involving aligning one's representations with the input. If interlocutor A is speaking, then B is listening to A and thus aligning with A. But if A is speaking, then A listens to himself through monitoring and thus aligns with himself. In other words, monitoring is a by-product of a language processing system that is sufficiently flexible to allow comprehension and production to occur to some extent simultaneously in dialogue. This means, for example, that monitoring should tend to be hard during periods of overlapping speech. Furthermore, monitoring is a key part of the checking and interactive repair process discussed in section 4.3. As a speaker you have to monitor your own contributions with respect to the implicit common ground and as a listener you have to monitor your partner’s contributions with respect to the same implicit common ground.

7. DIALOGUE AND LINGUISTIC REPRESENTATION

In the introduction, we noted that the main theoretical reason why mechanistic psycholinguistics has largely ignored dialogue is that formal linguistics has largely failed to address dialogue. We cannot of course rectify this situation here, but it is important to provide some sketch of how linguistic theory could support the study of dialogue, just as it has so far provided support for the study of monologue. Rather than attempt to address all relevant phenomena, we restrict ourselves to the discussion of two important general issues: the analysis of linked utterances and the architecture of the language system.

7.1 Dealing with linked utterances

As noted in section 2, dialogue turns are not isolated utterances, but are linked across interlocutors. However, traditional linguistics is based on monologue, and therefore treats the contribution of a single speaker as the unit of analysis. Even when the contributions are linked fragments, each contribution is treated on its own. However, this is clearly wrong. As long ago as 1973, Morgan demonstrated that there were syntactic restrictions on well-formed exchanges between interlocutors. For example, in A: What does Tricia enjoy most? B: Being called 'your highness'/*To be called ‘your highness’, the grammatical form of the answer is constrained by the subcategorization requirements of the verb in the question (see also Ross, 1969). Likewise, if A utters Is Jack in town? and B replies Jack?, B’s clarification request can only be analyzed with respect to A’s utterance (Ginzburg, 2001). The syntactic form of such elliptical requests is determined by the context (e.g., who? is also a possible response because it is a noun phrase like Jack). Hence this demonstrates a syntactic parallelism constraint between turns in dialogue.

The meaning of dialogue turns is also heavily constrained by context. If produced in isolation, the meaning of Jack? would be unclear; as a reply to Is Jack in town?, it means either "are you asking if Jack is in town?" (the clausal reading) or "who is the person named Jack you were referring to?" (the constituent reading). On both readings, some syntactic parallelism is required (e.g., he but not him can be used to clarify Is he in town?). The constituent reading employs phonological (or perhaps phonetic) parallelism, as it actually requires "echoing" of the exact form used (Ginzburg, 2001). A satisfactory linguistic account of dialogue should provide an account of how the form and interpretation of such short answers is constrained by the linguistic context. In part, this is because they are very common: According to Fernandez and Ginzburg (2002), non-sentential utterances constitute over 11% of dialogue turns in their
sample of the British National Corpus (Burnard, 2000), and clarification ellipses constitute nearly 9% of these. Ginzburg and Sag (2001) offer a linguistic account of such phenomena by incorporating context into linguistic representations.\(^7\) The interactive alignment model predicts parallelism in general and hence it is not surprising that parallelism emerges as a linguistic constraint in linked dialogue turns. Thus, Goldinger’s (1998) finding of phonological echoing and the phonological restriction on the constituent reading of clarification ellipsis may not be coincidental. Note that an adequate theory of language production also needs to be able to account for the contextual dependency of such utterances. It is not clear that current theories can do this, because they are designed to account for the production of isolated (and "complete") sentences (e.g., Bock & Levelt, 1994; Garrett, 1980).

The linguistic analysis of linked contributions as a single unit means that the mechanisms used to produce and comprehend them can be narrowly linguistic, in the sense that there is no need to appeal to "bridging" inference. Let us consider this in relation to a particularly extreme example of joint construction, when one interlocutor completes the other’s fragment. For example, Clark and Wilkes-Gibbs (1986) cite the following exchange: A: That tree has, uh, uh, ... B: Tentworms. A: Yeah. B: Yeah. Here, A appears unable to utter the appropriate expression, and B helps out by making a suggestion (which is then accepted). Of course, B’s response is only felicitous because it is syntactically congruent with A’s fragment (has can take a noun-phrase complement such as Tentworms, but could not take a prepositional-phrase complement such as Of tentworms).

According to the orthodox (monological) view, B would have to parse A’s utterance and assign it a semantic interpretation. Presumably, the parser can interpret an input (That tree has, uh, uh,) which is ungrammatical and not even a traditional constituent (though how this can be done is rarely specified). Then B would have to access its syntax and semantics (at least) but suppress production of these words. Next B must “fill in” the missing noun phrase by accessing and producing Tentworms. A will in turn have to interpret B’s "degenerate" utterance, and then integrate these two fragments via a bridging inference (though note that neither fragment has a propositional interpretation). This should cause processing difficulty (Haviland & Clark, 1974), but does not appear to. If things are this complicated it is unclear why interruptions should occur at all,\(^8\) why they can occur so rapidly, or why producing language in such contexts is not manifestly harder, say, than monologue. It also predicts that elliptical responses to questions should be harder than non-elliptical ones. This is clearly incorrect (e.g., Clark, 1979, showed that full responses are complex and have special implicatures).

Contrast this with the claim of the interactive alignment model, in which B, as listener, activates the same representations as A. These representations can be used in production in just the same way as in comprehension. Thus, we predict that it should be more-or-less as easy to complete someone else’s sentence as one’s own, and this does appear to be the case. Similarly, interlocutors should be able to complete each others’ words (e.g., if one speaker has difficulty) by making use of shared phonological representations. One prediction is that speech errors could be induced through perception as well as production (e.g., if B finishes off A’s tongue twister, then B should be liable to produce errors).

The existence of non-sentential turns in dialogue suggests that any appropriate grammatical account needs to be able to deal with such fragments, and allow their interpretations to be integrated into the dialogue context (as in, e.g., Poesio & Traum, 1997). A reasonable assumption is that the grammar should treat all well-formed dialogue turns as constituents, with a semantic interpretation, so that their meaning can be combined with the meanings of other participants’ turns in a compositional manner. This would require a "flexible" notion of constituency, where many fragments that are traditionally not constituents are treated as constituents (e.g., The tree has). One linguistic approach that accords with this is
Combinatorial Categorial Grammar (Steedman, 2000; cf. Ades & Steedman, 1982; Pickering & Barry, 1993). It allows most (but not all) fragments to be constituents, and is therefore a plausible candidate for analyzing the syntax of dialogue (which can also deal with monologue). It also provides a natural account of routines, because these may be constituents within flexible categorial grammar but not traditional linguistics (e.g., He’s overtaking; Kuiper, 1996) (for other linguistic treatments, see Kempson, Meyer-Viol, & Gabbay, 2001; Phillips, in press). Such linguistic proposals have already had some impact on psycholinguistic accounts concerned primarily with monologue comprehension (e.g., Altmann & Steedman, 1988; Pickering & Barry, 1991), in part because they provide a natural account of incremental interpretation (e.g., Just & Carpenter, 1980; Marslen-Wilson, 1973). Of course, any appropriate account also has to treat some dialogue utterances as ill-formed, for instance when a speaker simply stops mid-utterance (Levelt, 1983). In general, we need a linguistic account of well-formed dialogue utterances, and this account cannot be derived straightforwardly from linguistic theories based on monologue or citation speech.

7.2 The architecture of the language system

The interactive alignment model assumes independent but linked representations for syntax, semantics, and phonology (at least), where each level of representation plays a causal role via alignment channels (see Fig. 2). This sits ill with a Chomskyan "transformational" theory, with a central generative syntactic component and peripheral semantic and phonological systems that are purely "interpretative." In Chomskyan approaches (whether Standard Theory, Government and Binding Theory, or Minimalism), syntax creates sentence structure, and sound and meaning are "read off" this structure (Chomsky, 1965, 1981, 1995). Instead, the interactive alignment model is compatible with constraint-based grammar approaches in which syntax, semantics, and phonology form separate but equal parts of a multidimensional sign (Kaplan & Bresnan, 1982; Gazdar, Klein, Pullum, & Sag, 1985; Pollard & Sag, 1994). Within this tradition, Jackendoff's (1997, 1999, 2002) framework forms a particularly appropriate linguistic basis for the interactive alignment model. He assumes that phonological, syntactic, and semantic formation rules generate phonological, syntactic and semantic structures respectively, and are brought into correspondence by interface rules which encode the relationship between different systems.9 In our terms, the alignment channels can affect the application of the formation rules, whereas the interface rules are encoded in the links between the levels.10 It also provides a natural account of idioms and other routines, since the lexicon includes complex expressions (2002, chapter 6). In contrast, it is much more difficult to see why alignment should occur at phonological and semantic levels if no generative component underlies these levels. Moreover, the correspondence between the Chomskyan architectures and models of production and comprehension has always been difficult to sustain (e.g., Bock et al., 1992; Fodor, Bever, & Garrett, 1974; Pickering & Barry, 1991). Thus, we see the integration of a framework incorporating multiple generative components with a grammar that has a flexible approach to constituency as forming the linguistic basis for a psycholinguistic account of dialogue.

8. DISTINGUISHING BETWEEN DIALOGUE AND MONOLOGUE

In this paper we have argued that dialogue is the primary setting for language use and hence that dialogue processing represents the basic form of language processing. Throughout, we have treated dialogue and monologue as distinct kinds of language use. But is there a clear-cut distinction between dialogue and monologue or do they range along a dialogic continuum?

8.1 Degree of coupling defines a dialogic continuum

Interactive activities vary according to the degree of coupling between the interacting agents. Whereas a tightly coupled activity such as ballroom dancing requires continuous coordination between partners, a loosely coupled activity such as golf only requires intermittent coordination (one may have to wait until one’s partner has struck the ball, quality of play may
be affected by how close the scores are etc.). Similarly, different styles of communication vary in the degree of coupling between communicators. Whereas holding a one-to-one intimate conversation may require precise and continuous coordination (e.g., interruption, joint construction of utterances, back-channeling), giving a lecture only requires intermittent coordination (e.g., altering one’s style according to visual or vocal feedback from the audience, or responding to an occasional question).

The interactive alignment model was primarily developed to account for tightly coupled processing of the sort that occurs in face-to-face spontaneous dyadic conversation between equals with short contributions. We propose that in such conversation, interlocutors are most likely to respond to each others’ contributions in a way that is least affected by anything apart from the need to align. Hence it is not surprising that such language use in such situations is often regarded as primitive or basic (Clark, 1996; Linnell, 1998). As the conversational setting deviates from this “ideal,” the process of alignment becomes less automatic. For example, video-mediated conversation, ritualized interactions, multi-party discussions, tutorials, and speeches during debates each deviate in different ways from the ideal. In such cases, interlocutors will be less able to rely on automatic alignment and repair, and will need to spend more time constructing models of their interlocutors’ mental states if they are to be successful.

For example, Doherty-Sneddon et al. (1997) found that interlocutors in a collaborative problem-solving task were more efficient when they could see and hear each other than when either they could only hear each other or interacted via a high-quality video link. Specifically, face-to-face participants employed fewer words and checked their interlocutors’ comprehension less often than participants in the other conditions. Likewise, Fay, Garrod, and Carletta (2000) compared discussions involving five- or ten-member groups. In the small groups, the pattern of interruptions and turn-taking were similar to those in dyadic dialogue. Most interestingly, speakers tended to align with the immediately preceding speaker (with respect to their opinions about what was most important). But in the large groups, speakers did not align with the preceding speaker, but rather with the dominant speaker in the group. Hence the interactive alignment model predicted behavior in small groups but not large groups, where speakers appeared to use "serial monologue."

Whereas the prototypical form of dialogue involves tightly coupled contributions by interlocutors, the prototypical form of monologue involves one communicator making a single presentation without receiving any feedback. Good examples of this are speeches where there is no possibility of audience reaction (e.g., when speaking on the radio), and traditional written communication. In such cases, the communicator has to formulate everything on his own. He receives no help about what to produce, and cannot make use of an interlocutor’s contributions, because nothing from the addressee comes in through the alignment channels. (The only information that comes through the channels is via self-monitoring, and this is of much more limited use.) Hence, true monologue is very difficult, with successful communication often requiring very considerable planning (as in planning and rehearsing speeches) or use of very routinized speech (as in Kuiper’s sportscasters and auctioneers). However, much narrative is not as difficult as this, because the audience provides a considerable amount of feedback via backchannel and non-linguistic contributions (e.g., Bavelas, Coates, & Johnson, 2000). In cases where an interchange moves between highly interactive interchanges and long speeches by one interlocutor, we predict dynamic shifts in the difficulty of production.

In the comprehension of monologue, the listener will have to bring to bear appropriate inference skills. For example, he will often have to draw costly bridging inferences to help understand what the writer or speaker had really meant with a definite reference (Garrod & Sanford, 1977; Haviland & Clark, 1974), though again the difficulty is reduced if the listener
can give feedback (Schober & Clark, 1989). But in "passive" comprehension, there is no opportunity to call on aligned linguistic representations and no opportunity to resolve ambiguities using interactive alignment. Instead people have to fall back on the frequency of words, syntactic forms, and meanings in making comprehension decisions, as no other useful information is available. Therefore language users need to develop a whole range of elaborate strategies to become competent processors of monologue. Of course much of education involves training in writing essays and producing speeches, and the like, and a smaller part involves comprehension of monologue (e.g., in being able to identify the important arguments in a text). In contrast, people are very rarely taught how to hold conversations (except in some clinical circumstances). Without training in monologue, people are very likely to go off track during comprehension and production. Even after these strategies have been developed, people still find monologue far more difficult than dialogue.

9. IMPLICATIONS

The interactive alignment model is designed to account for the processing of dialogue, but we have already suggested that monologue can be regarded as an extreme case of non-interactive language use. This means that it can be harnessed into accounts of monologue processing as well. We shall briefly suggest its relevance to a range of other issues that extend beyond dialogue.

One interesting possibility is that it can serve as the basis for predominantly automatic accounts of social interaction more generally. There is considerable evidence that people imitate each other in non-linguistic ways, and hence alignment is presumably not purely linguistic. For example, Chartrand and Bargh (1999) demonstrated non-conscious imitation of such bodily movements as foot rubbing. Such findings, together with findings of the effects of the automatic activation of stereotypes on behavior, have led to the postulation of an automatic perception-behavior link that underlies such imitation (Bargh & Chartrand, 1999; Bargh, Chen, & Burrows, 1996; Dijsterhuis & Bargh, 2001; Dijsterhuis & Van Knipperberg, 1998). According to these researchers, the strength of this link means that the great majority of social acts do not involve a decision component. Our contention is somewhat related, in that we argue that the process of alignment allows the reuse of representations that are constructed during comprehension, in a way that removes the need to make complex decisions about how to represent the mental state of the interlocutor. Of course, there are still some conscious decisions about what one wants to talk about, but the computational burden is greatly reduced by making the process as automatic as possible. The social-psychological literature is fairly vague about precisely what is imitated; in contrast, our account assumes that people align on well-defined linguistic representations.

Indeed, the interactive-alignment account of dialogue meshes well with recent proposals about the central role of imitation within psychological and neuroscientific theorizing more generally (Heyes, 2001; Hurley & Chater, in press). The discovery of mirror neurons provides a reason to expect certain forms of imitation to be straightforward, and the finding that the same areas of the brain (Brodmann’s Areas 44 and 45) are involved in imitation as in language use (Iacoboni et al., 1999; Rizolatti & Arbib, 1998) provides support for the assumption that alignment constitutes a fundamental aspect of language use. To make these links more explicit, it would probably be necessary to perform the very difficult task of investigating brain activity during dialogue.

An obvious application of our account is to language acquisition, because alignment underlies imitative processes that occur as children acquire language. For instance, Brooks and Tomasello (1999) showed that 2-3 year olds could be trained to use passives by being presented with other passives. A prediction of the interactive-alignment model is that children will tend to repeat a construction that is novel to them to a greater extent when they also
repeat lexical items. From a rather different perspective, work on atypical language development might provide evidence for the circumstances under which the propensity for alignment might be disrupted. One would predict that this would be most likely when social functioning was impaired, and indeed there is evidence that imitation in general is impaired in autism (Williams, Whiten, Suddendorf, & Perrett, 2001). However, it is important to stress that alignment is unlikely to be directly linked to a complete "theory of mind," because it is not dependent on the modeling of the interlocutor’s mental state. Indeed findings such as Brooks and Tomasello’s speak against this account, on the grounds that such alignment occurs before most children pass "false belief" tasks (e.g., Baron-Cohen et al., 2000).

However, the model does not claim that assumptions about the mental state of one’s interlocutor are irrelevant to alignment. Presumably, one can decide whether one is interacting with an agent with which it is appropriate to align. Thus, we can consider the interesting case of human-computer interaction, where people may or may not align with computers’ utterances. If the conscious ascription of a mental state is necessary for alignment, then people will only align if they perform such ascriptions. But if people behave toward computers as "social agents," whatever they consciously believe about their mental states, then we predict unimpaired alignment will occur with computers, just as many other aspects of social behavior do (Reeves & Nass, 1996).

10. SUMMARY AND CONCLUSION

This paper presented a mechanistic model of language processing in dialogue, the interactive alignment model. The model assumes that as dialogue proceeds interlocutors come to align their linguistic representations at many levels ranging from the phonological to the syntactic and semantic. This interactive alignment process is automatic and only depends on simple priming mechanisms that operate at the different levels, together with an assumption of parity of representation for production and comprehension. The model assumes that alignment at one level promotes alignment at other levels including the level of the discourse model and hence acts as a mechanism to promote mutual understanding between interlocutors. The interactive alignment model was contrasted with an autonomous transmission account that represents the traditional psycholinguistic framework for language processing applied to dialogue. The main points of contrast between the two models are summarized in Table 2.

Table 2. Contrasts between autonomous transmission account of language processing in dialogue and the interactive alignment account.

<table>
<thead>
<tr>
<th>Autonomous transmission account</th>
<th>Interactive alignment account</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Linkage between interlocutors via sound alone — no direct links across other levels of representation.</td>
<td>1. Linkage between interlocutors Links across multiple levels of representation via ‘alignment channels’. Sound comes to encode words, linguistic information, and aspects of situational models.</td>
</tr>
<tr>
<td>2. Inference Internalised in the mind of speaker/listener: Speaker in terms of audience design; Listener in terms of bridging inference process.</td>
<td>2. Inference Externalised in the interaction between interlocutors via a basic interactive repair mechanism.</td>
</tr>
<tr>
<td>3. Routines Special case of language largely associated with idioms.</td>
<td>3. Routines Arise out of the application of the interactive alignment process. A high proportion of dialogue uses routines, which simplify both production and comprehension.</td>
</tr>
<tr>
<td>4. Self-monitoring Inner loop monitoring requires a special internal route from production to comprehension.</td>
<td>4. Self-monitoring Monitoring occurs at any level of representation that is subject to alignment as a consequence of the account.</td>
</tr>
<tr>
<td>5. Repair mechanisms</td>
<td>5. Repair mechanisms</td>
</tr>
</tbody>
</table>
First, according to the interactive alignment account the interaction between interlocutors supports direct channels between the linguistic representations that they use for language processing. In effect, the sounds come to directly encode words, meanings and even aspects of the situation model. Alignment occurs at different levels of representation and alignment at one level leads to further alignment at other levels. One of the mechanisms for this direct encoding is what we call routinization (see Table 2(5)): the setting up of semi-fixed complex expressions that directly encode complex meanings. A second contrast with the autonomous processing account relates to the nature of the inference processes associated with establishing the common ground in dialogue. Whereas inference in the traditional account is internalized in the minds of the speaker and listener, in the interactive alignment account it is externalized through an interactive repair mechanism that makes use of clarification requests. A third set of contrasts derives from the nature of the monitoring process assumed in the interactive alignment account. Whereas in the traditional account internal self-monitoring leads to the stipulation of a special mechanism in addition to the normal comprehension process, in the interactive alignment account it arises directly from the parity assumption. Monitoring output can occur at any level at which there is interactive alignment. Furthermore, there is a direct and simple relationship between self-repair processes and other repair processes in dialogue because the self-monitoring process is directly comparable to the other-monitoring process (see Table 2 (4)). Finally, the interactive alignment account challenges linguists to come up with a more flexible account of grammar capable of capturing linguistic constraints on linked sentence fragments.

Footnotes

1 In more detail, the procedure is as follows. Two players are confronted with two computer controlled mazes which do not differ in relevant ways. They are seated in different rooms but communicate via an audio link. The players each have a token representing their current position in their maze which is only visible to them and they take turns to move the tokens through the maze one position at a time until both players have reached their respective goal positions. At any time approximately half of the paths in each maze are closed. The closed paths are in different positions for each player and are only visible to that player. What makes the game collaborative is that the mazes are linked in such a way that when one player lands in a position where the other player’s maze has a ‘switch’ box, all of his closed paths open and open paths close. This means that the players have to keep track of each other’s positions in order to successfully negotiate their mazes. The dialogue shown in Table 1 is taken from a conversation which occurred at the beginning of a game. Garrodand Anderson (1987) analyzed transcripts from 25 pairs of players to see how location descriptions developed over the course of each game. Some of the results of this analysis are considered in more detail in Section 2.2.

2 Actually, Carlson-Radvansky and Jiang only found inhibition if the two trials used the same axis of the reference frame (e.g., the up-down axis). This limitation may be related to the fact
that priming was assessed outside a dialogue situation. An interesting prediction is that interlocutors would align on reference frames, not just axes.

Critically, ordinals such as 4th can only quantify over ordered sets of items, whereas locative adjectives such as top or bottom usually modify unordered sets of items. Therefore when speakers say 4th row..., they either have to give a post modifying phrase such as ...from the bottom, which imposes a particular ordering on the set of rows, or they have to assume that row denotes an element in an implicitly ordered set of rows. In other words, they assume that row in the bare 1st row is to be interpreted like storey of a building in 1st storey. (Notice that it is odd to talk of the 2nd storey from the bottom or even the bottom storey of a building, but fine to talk about the bottom floor.)

A very interesting issue occurs when alignment at one level conflicts with alignment at another. Perhaps the most obvious cases of this are when alignment at the situation model requires non-alignment at the lexical level. For example, in Schober’s (1993) example, two interlocutors who are facing each other use different terms to refer to similar locations (on the left vs. on the right) in order to maintain the same egocentric frame of reference. Likewise, Markman and Gentner (1993) show that successful use of analogy can require lexical misalignment. In Garrod and Anderson’s (1987) maze game, if one player uses second row to refer to the second row from the top in a five-row maze, then the other player will tend to use fourth row to refer to the second row from the bottom. The player could lexically align by using second row in this way, but of course this would involve misalignment of situation models, and would therefore be misleading. The implication is that normally alignment at the situation level overrides alignment at lower levels.

We assume that a case, for example, where the speaker could not remember who he meant by John (whilst speaking) would be pathological.

Most theories accept that a few dialogue phenomena do need to be explained. For instance, "binding" theory (Chomsky, 1981) can be evoked to explain why himself is coreferential with John in A: Who does John love? B: Himself, though see Ginzburg (1999) for evidence against an account in such terms. Rather than think of question-answer pairs as a marginal phenomenon that needs special explanation with a monological account, we regard them as a particularly orderly aspect of dialogue.

Roughly, Ginzburg and Sag assume HPSG-style feature structures (Pollard & Sag, 1994), in which context is incorporated into the representation of the fragments using the critical notion of QUDs ("questions under discussion").

Estimates from small group dialogues indicate that as many as 31% of turns are interrupted by the listener (Fay, Garrod, & Carletta, 2000).

Jackendoff uses the term conceptual structures instead of semantic structures, for reasons that we shall ignore for current purposes.

Note that Jackendoff (2002) assumes interface rules between semantic (conceptual) structures and phonological structures (p. 127, Fig. 5.5). If this is correct, it would suggest that Fig. 2 should incorporate such a link as well. He also suggests that the lexicon should be regarded as part of the interface components (p. 131).

The tendency might even be stronger for young children than adults, at least when it is the verb which is repeated. According to the "verb island hypothesis," syntactic information is more strongly associated with individual verbs in young children than it is in adults (e.g., children are often able to use a particular construction with some verbs but not others) (Tomasello, 2000).

ACKNOWLEDGEMENTS

We wish to thank Ellen Bard, Holly Branigan, Nick Chater, Herb Clark, Rob Hartsuiker, Tony Sanford, Philippe Schyns, Mark Steedman and Patrick Sturt for valuable comments, criticisms and helpful suggestions on earlier versions of this paper.
References


