BEHAVIORAL AND BRAIN SCIENCES (2004) **27,** 169–226
Printed in the United States of America

Toward a mechanistic psychology of dialogue

Martin J. Pickering
Department of Psychology, University of Edinburgh, Edinburgh EH8 9JZ, United Kingdom
Martin.Pickering@ed.ac.uk
http://www.psy.ed.ac.uk/Staff/academics.html#PickeringMartin

Simon Garrod
Department of Psychology, University of Glasgow, Glasgow G12 8QT, United Kingdom
simon@psy.gla.ac.uk **http://staff.psy.gla.ac.uk/~simon/**

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: common ground; dialogue; dialogue routines; language comprehension; language production; monitoring; perception-behavior link

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, whereas 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.* Hence, 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 that 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 responsing" – how can the experimenter stop subjects from 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 experimental control is possible with dialogue. For example, Branigan et al. (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 that 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 col-

© 2004 Cambridge University Press 0140-525X/04 $12.50

leagues (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 why psycholinguistics has ignored dialogue 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, that is, 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; 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. 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 (sects. 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 reestablish 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,

Martin J. Pickering is Professor of Psychology at the University of Edinburgh and has previously held lectureships and research fellowships at the Universities of Edinburgh, Glasgow, and Arizona. He has published over fifty journal articles, two edited volumes, and numerous other papers on topics in language comprehension, language production, reading, and dialogue. Some particular areas of interest are syntactic ambiguity resolution, semantic interpretation, the processing of unbounded dependencies, the use of eye-tracking methods, syntactic priming in language production, and aspects of the relationship between linguistics and psycholinguistics.

Simon Garrod is currently Professor of Cognitive Psychology at the University of Glasgow. Between 1989 and 1999, he was also Deputy Director of the ESRC Human Communication Research Centre. He has published two books, *Understanding Written Language* (with A. J. Sanford) and *Seeing, Saying and Acting: The Psychological Semantics of Spatial Prepositions* (with K. Coventry). He has also published over one hundred research papers on various aspects of the psychology of language. His special interests include discourse processing, language processing in dialogue, psychological semantics, and graphical communication. He is a Fellow of the Royal Society of Edinburgh.

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 cooperative 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.[1]

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."

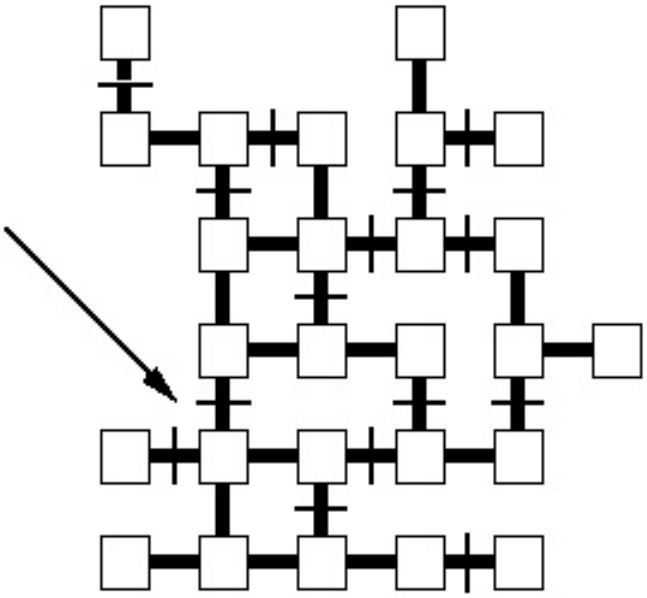

Figure 1. Schematic representation of the maze being described in the conversation shown in Table 1. The crossed bars indicate closed paths. The arrow points to the position being described by the utterances marked in boldface in the table.

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.

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 . . .
[*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 identified with an arrow in Figure 1. Colons mark noticeable pauses of less than 1 second.

Conversational analysts argue that dialogue turns are linked across interlocutors (Sacks et al. 1974; Schegloff & Sacks 1973). A question, such as (1) "Tell me where you are?", calls for an answer, such as (4) "Two along from the bottom one up." Even a statement like (4) "Right, two along from the bottom one 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' agreement 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 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 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) Alignment of situation models (Zwaan & Radvansky 1998) forms the basis of successful dialogue; (2) the way that alignment of situation models is achieved is by a primitive and resource-free priming mechanism; (3) the same priming mechanism produces alignment at other levels of representation, such as the lexical and syntactic; (4) interconnections between the levels mean that alignment at one level leads to alignment at other levels; (5) another primitive mechanism allows interlocutors to repair misaligned representations interactively; and (6) 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 that demonstrates that manipulations of these dimensions affect text comprehension (e.g., people are faster to recognize that a word has previously been mentioned when that word refers to something that is 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 they 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 his 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 from 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' (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 that alignment "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 in 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.[2]

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., Gentner & Markman 1997; Markman & Gentner 1993), 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 sect. 4). Recently, Brown-Schmidt et al. (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 et al. 2002; Tanenhaus et al. 1995). Brown-Schmidt 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 example, if the speaker used a referring expression which 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.

Levelt 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. Branigan et al.'s work extends "syntactic priming" 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 et al. 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 those used in comprehension (see sect. 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 they are likely to be particularly strong during natural interactions. In the Branigan et al. (2000) study, participants responded at their own pace, which should have made processing "natural," and hence conducive to strong priming. Second, we would 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 et al. (submitted) had two speakers take turns describing 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 (submitted) 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 his interlocutor's. There is also evidence that interlocutors align accent and speech rate (Giles & Powesland 1975; Giles et al. 1992).

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").[3] 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). 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 has 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 sect. 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., 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 confederate 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 (e.g., *the cowboy giving the banana to the burglar*), whereas other subjects produced a description using a different verb (e.g., *the cowboy handing the banana to the burglar*). 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 (2003) 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 et al. 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 (2003) 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, Gagné and Shoben (2002; cf. Gagné 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 that 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.[4] 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; cf. Garrod & Clark 1993). In such cases, interlocutors need to be able to appeal to other mechanisms 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.

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 bidirectional nature of the links means that there must be parity between production and comprehension processes.

3.1. Interactive alignment versus autonomous transmission

In the autonomous transmission account, the transfer of information between producers and comprehenders takes

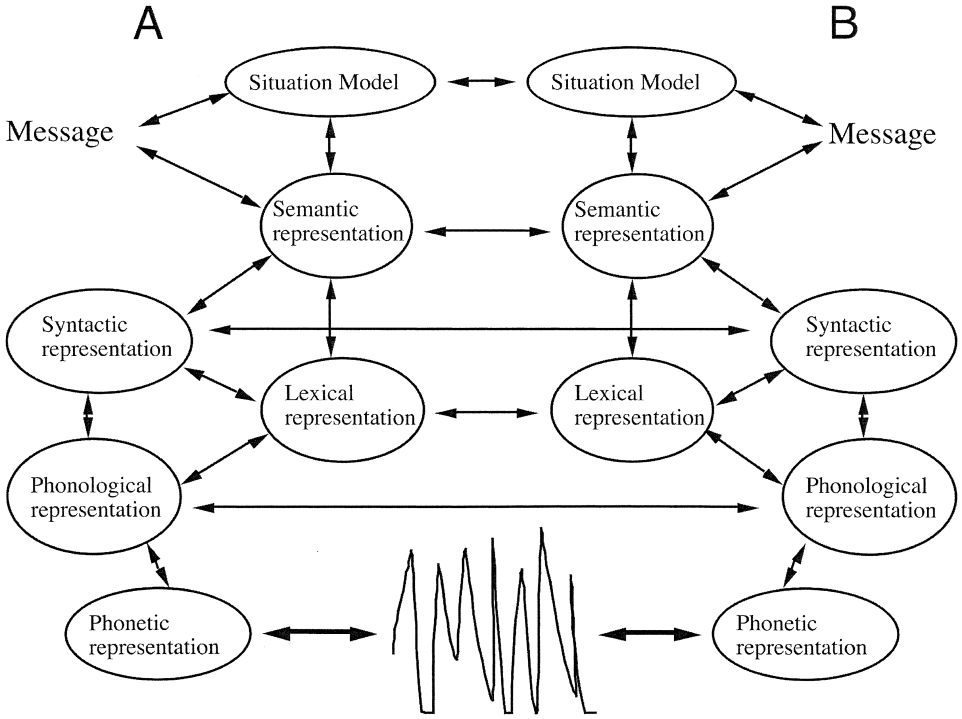

Figure 2. A and B represent two interlocutors in a dialogue in this schematic representation of the stages of comprehension and production processes according to the interactive alignment model. The details of the various levels of representation and interactions between levels are chosen to illustrate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

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 the levels involved in comprehension should constrain those 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 the speaker 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 prim-

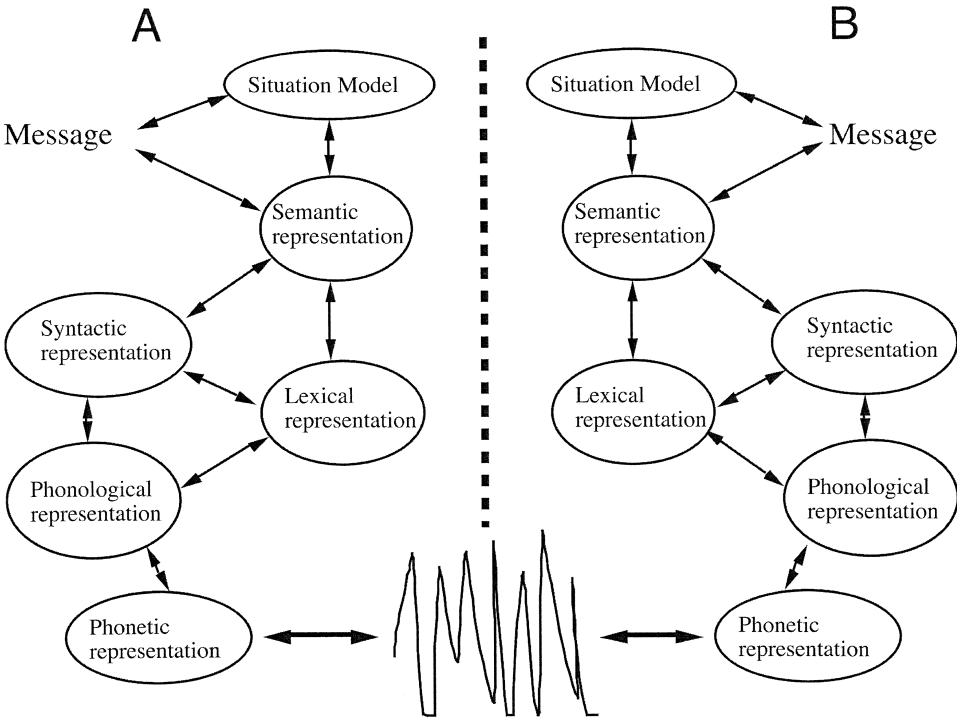

Figure 3. A and B represent two interlocutors in a dialogue in this schematic representation of the stages of comprehension and production processes according to the autonomous transmission account. The details of the various levels of representation and interactions between levels are chosen to illustrate the overall architecture of the system rather than to reflect commitment to a specific model. Links between interlocutors are missing here.

ing 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 bidirectional, 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"). 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 sect. 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 in dialogue is very different 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 example, why we can complete one another's utterances (and get the syntax, semantics, and phonology correct; see sect. 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). Goldinger 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 et al. 2001). We return to this issue in section 9. 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 because 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 interactive alignment 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 that 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 et al. (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 et al. (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, 2003, 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. 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 2003), though they proposed that this occurs at a later monitoring stage, in a process that they called perspective adjustment. More recently, Hanna et al. (2003) 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 (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 that 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, *B* 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 et al. 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. 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 et al. 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 generally 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, although 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):

> 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. (p. 385)

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. 246–50). In our example dialogue in Table 1, 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 sect. 5.2.2) 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 occurs at 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 et al. 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 sect. 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" in a political discussion. 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 "short-circuit" 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 form).

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) suggests that this explanation is correct. Kuiper 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 accessing 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 Kuiper's 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 while 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 example, isolated picture description. In both sentence re-

call 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 activated, and the appropriate syntax is activated. 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 discusses 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).

5.3. 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//æ//t/. Evidence for the sequential nature of activation comes from time-course data (Schriefers et al. 1990; van Turennout et al. 1998), "tip-of-the-tongue" data (Vigliocco et al. 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 (1990) 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 example, 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.

5.4. 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 et al. 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 et al. 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 et al. 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 "short-circuit" 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, formulates an utterance as a series of linguistic representations and then articulates 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 are 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, to make so-called "appropriateness repairs," for example. 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 other's 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 "cat" will be produced. Self-correction involves a repair process that is essentially similar to the straightforward repair process used during interaction (see sect. 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 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.[6] 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 (Morgan 1973; 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 lin-

guistic context. In part, this is because they are very common: According to Fernández and Ginzburg (2002), non-sentential utterances constitute more than 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,*) that 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 other's 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 (and 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 et al. 2001; Phillips 2003). 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 example, 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 (Gazdar et al. 1985; Kaplan & Bresnan 1982; 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] Jackendoff's framework also provides a natural account of idioms and other routines, because the lexicon includes complex expressions (2002, Ch. 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 et al. 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 target article 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, backchanneling), 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 other's 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 they could only hear each other or when they 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 et al. (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 et al. 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 con-

versations (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 et al. 1996; Dijksterhuis & Bargh 2001; Dijksterhuis & Van Knippenberg 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; Rizzolatti & 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.[11] 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 et al. 2001). However, it is important to stress that alignment is unlikely to require 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 article has 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.

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[3.]): the setting up of semi-fixed complex expressions that directly encode specific 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 ac-

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

Autonomous transmission account	Interactive alignment account
1. *Linkage between interlocutors* Via sound alone – no direct links across other levels of representation.	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.
2. *Inference* Internalized in the mind of speaker/listener: Speaker in terms of audience design; Listener in terms of bridging inference process.	2. *Inference* Externalized in the interaction between interlocutors via a basic interactive repair mechanism.
3. *Routines* Special case of language largely associated with idioms.	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.
4. *Self-monitoring* Inner loop monitoring requires a special internal route from production to comprehension.	4. *Self-monitoring* Monitoring occurs at any level of representation that is subject to alignment as a consequence of the account.
5. *Repair mechanisms* Distinct repair mechanisms for self-repair and other-repair in dialogue.	5. *Repair mechanisms* The same basic repair mechanism for self-repair and other repair
6. *Linguistic representations* Only need to account for the structure of isolated and complete sentences.	6. *Linguistic representations* Needed to deal with linked utterances in dialogue, including non-sentential "fragments."

count 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.

ACKNOWLEDGMENTS

Order of authorship is arbitrary. We wish to thank Ellen Bard, Holly Branigan, Nick Chater, Herb Clark, Jonathan Ginzburg, Art Glenberg, Rob Hartsuiker, Bernhard Hommel, Gerard Kempen, Art Markman, Keith Rayner, Tony Sanford, Philippe Schyns, Mark Steedman, Patrick Sturt, and two anonymous reviewers for valuable comments, criticisms, and helpful suggestions on earlier versions of this paper. MP was supported by ESRC Grant No. R000239363.

NOTES

1. In more detail, the procedure is as follows. Two players are confronted with two computer-controlled mazes that 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 to successfully negotiate their mazes. The dialogue shown in Table 1 is taken from a conversation that occurred at the beginning of a game. Garrod and 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.

3. 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.*)

4. 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

nonalignment 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*) 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.

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

6. Most theories accept that a few dialogue phenomena do need to be explained. For example, "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 in a monological account, we regard them as a particularly orderly aspect of dialogue.

7. Roughly, Ginzburg and Sag assume feature structures taken from Head-Driven Phrase Structure Grammar (Pollard & Sag 1994), in which context is incorporated into the representation of the fragments using the critical notion of QUDs ("questions under discussion").

8. Estimates from small group dialogues indicate that as many as 31% of turns are interrupted by the listener (Fay et al. 2000).

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

10. Note that Jackendoff (2002) assumes interface rules between semantic (conceptual) structures and phonological structures (p. 127, Fig. 5.5). If this is correct, it suggests that Figure 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).

11. The tendency might even be stronger for young children than adults, at least when it is the verb that 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).

Open Peer Commentary

Is language processing different in dialogue?

Dale J. Barr[a] and Boaz Keysar[b]

[a]*Department of Psychology, University of California, Riverside, CA 92521;*
[b]*Department of Psychology, University of Chicago, Chicago, IL 60637.*
dale.barr@ucr.edu boaz@uchicago.edu

Abstract: Pickering & Garrod (P&G) claim that the automatic mechanisms that underlie language processing in dialogue are absent in monologue. We disagree with this claim, and argue that dialogue simply provides a different context in which the same basic processes operate.

Pickering & Garrod (P&G) call for closer attention to the mechanisms underlying coordination in dialogue. There are good empirical grounds for accepting many of the basic assumptions of the interactive alignment model. Specifically, the strong egocentrism of speakers and listeners that we have uncovered in our own studies (cf. Barr & Keysar [in press], for a recent review) makes much sense within a context of strong representational overlap, and the interactive alignment model provides an appealing explanation for how such overlap comes about.

A central assumption of P&G is a categorical distinction between language processing in monologue and in dialogue. Clearly, monologue is different from dialogue because in monologue there is no feedback and no opportunity for interactive repair. But P&G go further than this, asserting that in monologue, "the automatic mechanisms of alignment are not present" (sect. 3.2, para. 3) and that "there is no opportunity to call on aligned linguistic representations" (sect. 8.1, para. 5). Under this view, dialogue involves processes that are fundamentally distinct from those present in monologue. We disagree. We suggest that only the strategic mechanisms of feedback and interactive repair are absent in monologue, not the automatic mechanisms of alignment. Just as there is a "dialogic continuum" defining different kinds of interactive activities, we argue that there is a continuum of processing and that alignment will be observed even under non-interactive circumstances.

Differences in processing may simply be a matter of degree, not of kind, especially given that monologue-like episodes are common even in naturalistic conversation. Consider, for example, the first utterance of a conversation, which in many cases is a mini-monologue before a full-fledged interactive exchange develops. The speaker will need to go through the various stages to produce the first utterance (e.g., Levelt 1989), and the comprehender must parse the utterance in order to appreciate its significance. Later on in the exchange, enough of shared but nonmutual information (what P&G unfortunately term "implicit common ground")[1] may have built up to short-circuit many aspects of these processes. But there is no reason to believe that different theories would be required to explain the processing of the first utterance of a dialogue versus the hundred-and-first. Indeed, P&G claim that the alignment processes are automatic and resource-free, and processes of these sorts cannot simply be switched off. Finally, the "dialogic continuum" cited by the authors does not just provide a means for classifying whole conversations but actually represents a domain of activity that can be fully traversed even within the span of a single conversation. A conversationalist can at any point secure turn space in order to engage in an extended monologue (Sacks et al. 1974) – for example, about her trip to India – followed by close coordination with her interlocutor in order to arrange a time for a future meeting. There is no reason to expect alignment processes and bridging inferences to toggle on and off as the interactivity of the discourse changes.

Studies that find differences between the quality of comprehension of interlocutors and of overhearers (e.g., Schober & Clark 1989; Wilkes-Gibbs & Clark 1992) might be taken as evidence to support the idea of radically different cognitive processes in interactive discourse. They need not be. In such studies interlocutors had opportunities for feedback and repair that the overhearers lacked. Because different people will misunderstand different things, those who can ask for clarification will receive feedback that is relevant to them and consequently might understand better. Therefore any difference between such noninteractive and interactive comprehension could be fully attributable to strategic, effortful feedback but not necessarily to automatic alignment. In fact, Barr and Keysar (2002) found that even when such feedback is removed, listeners who believed themselves to be overhearers automatically aligned their semantic representations with the speaker's to the same degree as listeners who believed themselves to be addressees.

In closing, far from qualitatively changing the nature of processing, it is likely that dialogue provides a radically different context in which the same processes operate. The context includes an interlocutor and mechanisms for feedback and interactive repair. For us there is no question that it is important to study conversation in vivo, but it remains to be seen whether this would reveal automatic processes that are truly unique to dialogue.

NOTE

1. In their seminal work on common ground, Clark and Marshall (1981) clearly make the case that common ground is a form of meta-knowledge that is conceptually distinct from shared knowledge. What P&G are referring to by "implicit common ground" is really just shared knowledge, not common ground, because interlocutors need not represent the fact that their representations are shared. Such usage is certain to contribute to the legacy of confusion that has plagued discussions of mutual knowledge and common ground (see Keysar 1997 and Lee 2001 for discussion).

Full alignment of some but not all representations in dialogue

Holly P. Branigan

Department of Psychology, University of Edinburgh, Edinburgh EH8 9JZ, United Kingdom. **holly@cogsci.ed.ac.uk**
http://www.psy.ed.ac.uk/Staff/ra/hbranigan/publications.html

Abstract: I argue that alignment of linguistic representations and situation models in dialogue are qualitatively distinct. By virtue of the isomorphy between interlocutors' linguistic representations, interlocutors align their linguistic representations fully. However, evidence about situation models is indirect and mediated through language, with the result that alignment of situation models is only partial.

Pickering & Garrod (P&G) provide a plausible and very welcome account of language processing in dialogue. Their account assigns a central importance to the notion of alignment. Here I consider the nature of alignment and, in particular, whether alignment of strictly linguistic representations and alignment of situation models are qualitatively distinct.

P&G suggest that successful dialogue arises from the alignment of representations between interlocutors, and particularly alignment of situation models. Alignment is defined as interlocutors having the same representation at a particular level of structure. It is uncontroversial that language is represented in the same way in different speakers. Linguists and psycholinguists assume a common competence grammar in adult speakers (Chomsky 1965), even though they may differ as to whether this grammar is innate (e.g., Pinker 1989) or to some extent constructed through experience (e.g., Tomasello 2000). Crucially, adult speakers' internalised knowledge of the syntax, lexicon, and morpho-phonology of a language is held to be identical, such that there is an isomorphic mapping from any one speaker's internalised representation of the language to any other's. In a dialogue, then, interlocutors necessarily make use of identical representations in producing their utterances.

Under P&G's maximally parsimonious assumption of parity of representations, interlocutors also necessarily draw upon identical representations in both producing and comprehending utterances. Note also that speakers' utterances provide direct linguistic evidence to the listener. So when a listener hears an utterance, he receives direct evidence (except in cases of mishearing or unresolved structural ambiguities) about the syntactic, lexical, and morpho-phonological representations that the speaker has employed. If a speaker produces an utterance like *I am in row two,* for example, the listener has direct evidence that she has used the words *I, am,* and so on (and their relevant inflectional markings), that she has used a pronoun and a verb and so on, and that she has used a noun phrase, a verb phrase, a prepositional phrase, and so on. Taken together, the combination of isomorphy of representations and direct evidence strongly supports P&G's contention (summarised in their Fig. 2) that linguistic representations used by interlocutors in dialogue act directly upon one another, and that, in a very real sense, when we talk about interlocutors having aligned linguistic representations, we mean that those representations are identical. In summary, P&G's arguments for full alignment at linguistic levels of representation seem well founded.

But are situation models aligned in the same way as linguistic representations? It is unclear that this is the case. In P&G's model, interlocutors' situation models act directly upon one another (see the authors' Fig. 2), in the same way as syntactic, lexical, and morpho-phonological representations do, and alignment of situation models is taken as critical for successful communication. But situation models differ qualitatively from strictly linguistic representations. A speaker's utterances do not give direct evidence of the situation model that the speaker holds, only indirect evidence encoded in linguistic representations, from which the listener has to infer the speaker's situation model. So whereas an utterance like *I am in row two* gives direct evidence about the speaker's syntactic, lexical, and morpho-phonological representations, it gives only indirect evidence about the speaker's situation model. The listener must construct a situation model based upon his or her interpretation of the speaker's meaning – which may or may not be correct. Of course, as P&G note, misunderstandings may come to light, and interlocutors may initiate repairs to bring about situation models that are aligned in the relevant aspects. But as they also note, some misunderstandings may not be repaired. In fact, it seems likely that interlocutors quite frequently have situation models that are misaligned in major respects. Communication will be (apparently) successful as long as the misalignment is not apparent to the interlocutors. To take P&G's example of interlocutors using *John* to refer to different people, it is quite possible for them to have a mutually satisfying dialogue concerning this person without ever realising that they are discussing different people; unless one of them says something that is inconsistent with the other's knowledge, they can successfully (for their purposes) complete a dialogue with quite radically different situation models. Equally, a doctor and a patient may have a dialogue concerning the patient's *chronic back problem* that appears to be successful, in that they are both satisfied that they understand each other well; yet their situation models may differ considerably because of unresolved (and unapparent) differences in their interpretation of *chronic.* Situation models need only be aligned sufficiently for the current communicative goal to be (apparently) met.

So it seems that alignment of situation models and alignment of linguistic representations are quite different. With linguistic representations, interlocutors genuinely employ aligned (i.e., identical) representations that act directly upon one another; whereas, because evidence for situation models is only indirect, interlocu-

tors rarely if ever have *identical* models. Instead, they have partially aligned models that may differ in many – sometimes important – respects. And because evidence for situation models is mediated through language, it seems highly unlikely that they can act directly upon one another (contra P&G's Fig. 2).

One interesting result of the distinction between alignment of linguistic and situation models is that alignment of linguistic representations may sometimes lead to misaligned situation models. Garrod and Clark (1993) found that young children had a tendency to use the same words to describe a maze – that is, showed lexical alignment – even when their situation models were quite different. Similarly, in the case of the doctor-and-patient scenario, one speaker's use of the term *chronic* may well reinforce the other's use of the same term, leading to more misunderstanding than if a different term were used. In both examples, full alignment at the linguistic level misleads interlocutors into believing that they also have alignment at the level of situation models.

ACKNOWLEDGMENTS
This research was supported by a British Academy Postdoctoral Fellowship and by ESRC grant R000239363.

Two steps forward, one step back: Partner-specific effects in a psychology of dialogue

Susan E. Brennan and Charles A. Metzing
Department of Psychology, State University of New York, Stony Brook, NY 11794-2500. **susan.brennan@sunysb.edu cmetzing@ic.sunysb.edu http://www.psychology.sunysb.edu/psychology/personnel/Susan.htm**

Abstract: Pickering & Garrod's (P&G's) call to study language processing in dialogue context is an appealing one. Their interactive alignment model is ambitious, aiming to explain the converging behavior of dialogue partners via both intra- and interpersonal priming. However, they ignore the flexible, partner-specific processing demonstrated by some recent dialogue studies. We discuss implications of these data.

In human language processing, the whole is greater than the sum of the parts; therefore, those who study the language processing system in dialogue contexts are poised to make different sorts of discoveries than those who study the parts working alone. Pickering & Garrod (P&G) present a convincing argument that psycholinguists should pay attention to dialogue. In fields such as artificial intelligence and human-computer interaction, where the goal is often to build a fully working dialogue partner, many will find this a worthy enterprise as well. After presenting evidence for phonological, lexical, and syntactic convergence between dialogue partners and for representations shared between comprehension and production, P&G make a strong claim that is far *less* convincing: "normal conversation does not routinely require modeling the interlocutor's mind" (sect. 4.4, para. 4). They support this position with evidence from studies that fail to meet the very standards they seek to advance, while ignoring evidence that complicates matters for their interactive alignment model. Thus, their position on the importance of studying language in dialogue does not go far enough.

This position assumes that interlocutors achieve aligned mental representations without having to track anything specific about each other's knowledge because both have evolved with the same cognitive architecture; what is easiest for speakers is easiest for addressees (Brown & Dell 1987). It further assumes that there is no need to track common ground, as interlocutors each use their own memory of the conversation as a proxy. By this argument, what appears to be partner-specific or "audience design" is actually inflexible and unavoidable, at least in the earliest moments of processing. P&G propose a two-stage model (similar to that of Horton & Keysar 1996), arguing that interlocutors "do not routinely take common ground into account during initial processing . . . full common ground is only used when simpler mechanisms are ineffective" (sect. 4.1, last para.). This (circular) view relegates any aspect of production or interpretation that displays flexibility or sensitivity to an interlocutor's needs (as distinct from one's own) to the status of a relatively late adjustment, managed as a kind of repair or pragmatic garden path.

Granted, it is difficult to design a good experiment on audience design. A good experiment must distinguish one interlocutor's perspective from another's, avoid confounding individual perspectives with common ground (Keysar 1997), and allow interlocutors to interact naturally or contingently (Schober & Brennan 2003). But we are surprised that studies succeeding in all this (and finding partner-specific effects early in processing, e.g., Hanna et al. 2003; Nadig & Sedivy 2002) are dismissed by P&G: "their task was repetitive and involved a small number of items, and listeners were given explicit information about the discrepancies in knowledge" (target article, sect. 4.2, para. 3). Then follows a very broad claim: "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."

Paradoxically, evidence to support this position comes mainly from studies that did not allow any potential for interaction. These include Brown and Dell (1987), Ferreira and Dell (2000), Horton and Keysar (1996), and others in which partners did not interact naturally or provide contingent feedback. Sometimes this matters; for example, Brown and Dell (1987) concluded that speakers did not take addressees' specific needs into account when retelling stories; but their addressees had no needs (they were confederates who knew the stories better than the speakers did). When we ran a similar study using spontaneously interacting speakers and addressees (Lockridge & Brennan 2002), speakers' early syntactic choices indeed showed sensitivity to addressees' needs.

There is additional good evidence of rapid, partner-specific effects from the comprehension side. Hanna and Tanenhaus (2004) asked addressees to follow a (confederate) speaker's directions in a cooking task (e.g., *Hand me the cake mix*); the addressees' eye fixations showed that they restricted candidate referents for ambiguous expressions (e.g., when two cake mixes were present) depending on what the speaker was holding and what she could not reach; they did this from the earliest moments of processing.

And we have demonstrated that addressees interpret the same utterance differently when it is spoken by different speakers with whom the addressees have different dialogue histories (Metzing & Brennan 2001; 2003). In our experiment, addressees were instructed by (confederate) speakers to reposition objects among a relatively large set; they did this several times, evolving shared perspectives and terms for critical objects (e.g., *the shiny cylinder*). Then the speaker left the room and either returned or else a new confederate speaker entered. In the final trial, the new or old speaker used either the familiar term or a new, equally good term (e.g., *the silver pipe*) for the same critical object (amid many other references that did not use different terms). Addressees gazed immediately at the object when either speaker used the old term. However, when the old speaker used a new term (inexplicably breaking a conceptual pact), addressees experienced interference, delaying gazing at the target object. There was no such delay when the new speaker used the new term (in fact, resolving this was just as fast as the old term spoken by the new speaker). This partner-specific interference suggests that the pragmatic force of breaking a conceptual pact has impact immediately, rather than just as a late adjustment or repair.

Such immediate effects provide evidence of impressive agility and potential for partner-specific processing in the language processing system, which the interactive alignment proposal fails to address. Pragmatic and partner-specific knowledge is implemented by basic mechanisms of memory and does not rely on special processes or exhaustive partner models. Audience design – truly partner-specific processing – can occur immediately and effortlessly as well as more slowly and deliberately, depending on

how activated relevant information is. P&G's strong separation between "implicit common ground" (automatic but excluding any partner specificity) and "full common ground" (requiring reprocessing) is unconvincing.

Yes, the potential for interaction matters! But the interactive alignment model has farther to go before it can accommodate the flexible and adaptive processing that these data support.

ACKNOWLEDGMENTS
This material is based upon work supported by the National Science Foundation under grants No. 0082602 and No. 9980013. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Priming and alignment: Mechanism or consequence?

Sarah Brown-Schmidt and Michael K. Tanenhaus
Department of Brain and Cognitive Sciences, University of Rochester, River Campus, Rochester, NY 14627. **sschmidt@bcs.rochester.edu**
mtan@bcs.Rochester.edu
http://www.bcs.rochester.edu/people/sschmidt/
http://www.bcs.rochester.edu/people/mtan/mtan.html

Abstract: We agree with Pickering & Garrod's (P&G's) proposal that dialogue is an important empirical and theoretical test bed for models of language processing. However, we offer two cautionary notes. First, the enterprise will require explicit computational models. Second, such models will need to incorporate both joint and separate speaker and hearer commitments in ways that go beyond priming and alignment.

We applaud and second Pickering & Garrod's (P&G's) call to psycholinguists to include dialogue as an empirical and theoretical test bed for models of language processing. There is much to be gained by combining the tools for studying real-time processing developed within the language-as-product tradition with the more natural interactive situations typically used within the action tradition (Tanenhaus et al. 2004). And we believe that many of the basic processes in language comprehension, including spoken-word recognition, syntactic processing, and reference resolution, can be studied with the same precision in dialogue settings as in more traditional controlled experiments (e.g., Brown-Schmidt et al. 2002; in press). However, we would place a somewhat different emphasis on why it is important to study dialogue.

First, unlike P&G, who suggest that theories of language processing within the product tradition are admirably well-specified, we think that there has been a dearth of explicit mechanistic models of language processing. This is especially true within language comprehension, where, with the exception of some limited models developed within the neural network tradition, most computational models make only tenuous contact with behavioral data, and vice versa (Christiansen & Chater 2001). However, we believe that the emergence of interactive dialogue systems within computational linguistics offers an opportunity to develop explicit computational models in domains that can also be used to study human language processing, thus creating a synergistic feedback loop between modeling and experimentation. Although the field has not yet reached this stage, the opportunity is on the horizon, as computational linguists strive to implement systems that can engage in continuous generation and understanding.

The motivation for continuous understanding and generation in dialogue systems is instructive (Allen et al. 2001). During utterance generation, a speaker needs to have the capacity to monitor feedback from an addressee, both verbal and nonverbal, and plan or adjust the continuation of her utterance accordingly. Consider, for example, an utterance which begins, "Now, take thee,[1] uh, Phillips head screwdriver. . . ." If the addressee nods or says "uh huh," the speaker can continue with, "and tighten the bolt." However, if the addressee says "huh," looks perplexed, or begins to reach for the wrong tool, the speaker is likely to continue with, "the one with the blue handle, the one closest to your wrench" or say "No, *that* one" (while pointing). This example illustrates one of the potential benefits of studying interactive dialogue: It can shed different light on some basic assumptions. For example, most psycholinguists use the fact that comprehension occurs more or less continuously as a reason why they focus on response measures that are closely time-locked to the input. Yet, if pressed to answer the question of why comprehension is so incremental, they would likely appeal to the fact that memory capacity is limited. But, working memory constraints do not explain why processing is as relentlessly continuous as it is. In fact, the early theories of language comprehension, which were largely driven by working-memory assumptions, assumed that comprehension was a catch-up game, with many delays so that listeners could avoid making premature commitments. However, incremental processing is necessary for efficient use of feedback from an interlocutor.

In confronting the challenges involved in integrating comprehension and production, as well as speech recognition, intention recognition, and utterance planning, psycholinguists will be required to go beyond the boxes and arrows and consider more detailed models of what needs to be computed. Here, the study of dialogue provides a useful paradigm by allowing psycholinguists and computational linguists to study and model these processes in relatively constrained domains where it is possible to be explicit about the relevant components of the system and how they interact.

We believe that when P&G consider the problem of dialogue from this perspective, they will need to rethink two assumptions that guide their approach. The first is that interlocutors do not need to take into account differences between speaker and hearer knowledge and perspectives, except for repair strategies. We agree that many aspects of language use can be egocentric, though we think that P&G overstate the case, mistaking evidence that common ground does not fully constrain referential domains, with evidence for the stronger claim that it is ignored in initial processing (cf. Hanna et al. 2003). Moreover, natural language is rife with constructions that depend crucially on differences between assumed speaker and hearer knowledge. To take a simple example, the declarative question "You used the Phillips tool?" can be uttered only when the speaker believes the addressee is committed to the presupposed assertion that is being questioned (Gunlogson 2003). If the addressee had in fact used the tool, she might respond with an explanation; if she had not, she would need to contradict the presupposed commitment by saying something like "No, of course not, I. . . ."

These kinds of phenomena will emerge as psycholinguists consider richer dialogue situations in which the participants have reasons to make choices between alternative forms of utterances, and more complex interactions where it is important for interlocutors to track each other's attention and intentions. Finally, note that both intention and attention, which are crucial components of monitoring an interlocutor, do not necessarily appeal to high-level, resource-demanding processes. Eye gaze, for example, is a powerful source of information about attention. Moreover, the aspect of an interlocutor's knowledge that has to be monitored can often be circumscribed by goal structures. Cast in this light, perspective monitoring can be seen as one of the basic components of communication, rather than a special purpose, resource-intensive, revision process. In many of the examples that P&G focus on, the goal structures are completely defined by the task and the choice of utterances is limited. Under these conditions, it is easy to view alignment and priming as the primary mechanisms of dialogue, rather than as interesting phenomena that can be used to provide insight into the representations and processes underlying interactive conversation.

ACKNOWLEDGMENT
This work was supported by a National Institutes of Health grant, No. HD-27206, to Michael K. Tanenhaus.

NOTE

1. We use the word "thee" to indicate a disfluent pronunciation of the word "the" (and not in the old English usage as a pronoun).

A call for more dialogue and more details

J. Cooper Cutting

Department of Psychology, Illinois State University, Normal, IL 61790.
jccutti@ilstu.edu
http://main.psy.ilstu.edu/faculty/cutting/index.html

Abstract: Pickering & Garrod (P&G) argue that contemporary models of language use are inadequate. This has resulted largely because of an experimental focus on monologue rather than dialogue. I agree with the need for increased experimentation that focuses on the interplay between production and comprehension. However, I have some concerns about the Interactive Alignment model that the authors propose.

Pickering & Garrod (P&G) make an excellent argument for the need to increase the study of the interplay between production and comprehension in mainstream psycholinguistic research. Historically, most research on language has treated the processes of comprehension and production as largely independent systems. This separation may reflect the different starting points of the two processes. Input for comprehension is easy to manipulate and control experimentally. Input for production is thought, something that is much harder to control experimentally (Bock 1996). As a result, language comprehension research has a long history relative to that of language production research. Furthermore, theories of comprehension are rarely informed by research on production, and vice versa; however, the two processes are clearly intertwined. A complete model of language use needs to consider how production and comprehension processes work in conjunction. The Interactive Alignment model is explicitly designed to do just that. Rather than focus on how we process language as a monologue, as most mainstream psycholinguistic research does, the model focuses on the processes involved in dialogue between two (or more) language users.

Although I support the idea that much more research focusing on the interplay of production and comprehension is needed, I believe that the Interactive Alignment model that the authors propose is underspecified in several important respects.

1. What is alignment? The central feature of the model is the *alignment* of multiple levels of linguistic representations between interlocutors. This is achieved via "channels of alignment" through which "the activation of a representation in one interlocutor leads to the activation of the matching representation in the other interlocutor *directly*" (sect. 3.2, para. 2; my emphasis). In P&G's Figure 2, these channels are depicted as bidirectional links between representations within both dialogue participants. Although it is easy to imagine direct links between representations within a person (see the parity assumption below), it is much more difficult to determine what these "direct" links between different individuals correspond to. The authors propose that the mechanism of alignment is a "primitive and resource-free priming mechanism" both within and between levels of representation. Although this may be different from strict serial models of language use, it seems to be a feature of existing interactive models of language use (for recent reviews, see Dell et al. 1999; Pickering et al. 2000). Within these models, the impact of recent use (of representations) in dialogue could be modeled with residual activation of representations from earlier parts of the conversation. However, this priming mechanism does not seem to be very "direct" but is instead indirect.

2. Is all priming alike? The authors suggest that alignment of multiple levels of representation (including phonological, syntactic, lexical, semantic, and situational) is achieved by priming. However, "priming" phenomena in language may not all share the same underlying mechanisms. For example, Bock and Griffin (2000) have suggested that syntactic persistence is a reflection of learning rather than spreading of activation. It is unclear how the interactive alignment model distinguishes between activation-based and learning-based priming mechanisms.

3. Parity of representations. The authors propose that there is parity between the representation used by production and that used by comprehension. In other words, although production and comprehension processes may be different, they share the same set of linguistic representations. However, the issue of shared versus distinct representations needs to be considered for each level of representation (e.g., Balota 1990; Caramazza 1991; Levelt 1989). For example, Cutting (1998) used a word-picture priming technique in which two words were presented in the prime trial. The participant produced one word while the other was ignored. Cutting found that both produced and ignored semantically related prime words interfered with picture naming. However, only produced phonologically related primes influenced picture naming; ignored phonologically related primes had no effect. Cutting interpreted this pattern of results as support for a model in which semantic representations are shared by production and comprehension whereas phonological representations are separate. The authors refer to Hommel et al.'s (2001) Theory of Event Coding (TEC) as a model for language representations. Interestingly, Hommel et al.'s proposal also states that TEC is most appropriate for abstract distal coding. In other words, because phonological representations are the linguistic representations "nearest" to the sensory code and muscular innervation patterns, it should not be surprising that they are different for production and comprehension (cf. Martin et al. 1999; Zwitserlood 1994).

4. Routinized language. The authors propose that interlocutors use *routines* which are developed "on the fly" during dialogue (similar to idioms like "kick the bucket"). These routines result from the processes of alignment in dialogue. As speakers and listeners use particular lexical, semantic, and syntactic representations, some of these representations become bound into routines. The use of these routines may allow a speaker to "bypass" some of the early stages of processing assumed by traditional models of language production (e.g., Bock & Levelt 1994; Levelt 1989; Levelt et al. 1999). However, the process of routinization in the interactive alignment account is underdeveloped. It is unclear how these representations are bound (concurrent activation, learned connections, etc.), how they are represented (in either the short-term or the long-term; Cutting & Bock 1997), or how they actually "bypass" stages of language production (i.e., do routines bypass stages or just grease the wheels a bit?). Without further specification of these issues it is difficult to evaluate the model's predictions.

In conclusion, the interactive alignment model as currently proposed is underspecified with respect to how processes of alignment, priming, and routines work. As currently stated, it is difficult to determine specific testable predictions that would distinguish this model from currently available models of production and comprehension. However, the point of taking a dialogue perspective to these models (and the creation of new ones) is an excellent one. Although there is no shortage of monologue use of language (television, radio, print, lecture, etc.), clearly dialogue should be a central aspect of research. It is time to tear down some of the historically designed barriers between models of language comprehension and production and examine what impact a dialogue perspective has on these models.

Situation alignment and routinization in language acquisition

Peter F. Dominey

Institut des Sciences Cognitives, CNRS UMR 5015, 69675 BRON Cedex, France. **dominey@isc.cnrs.fr** **http://www.isc.cnrs.fr**

Abstract: Pickering & Garrod (P&G) describe a mechanism by which the situation models of dialog participants become progressively aligned via priming at different levels. This commentary attempts to characterize how alignment and routinization can be extended into the language acquisition domain by establishing links between alignment and joint attention, and between routinization and grammatical construction learning.

Pickering & Garrod (P&G) describe a mechanism by which the situation models of dialogue participants become progressively aligned via priming at different levels, including lexical, syntactic, semantic, and situational representations. An essential interest and novelty of this approach is that, instead of requiring a complex and effortful mechanism for explicitly constructing a common ground, it offers a rather straightforward mechanism that operates largely automatically via priming.

It is of potential interest that this type of alignment can be seen to be useful in other communicative contexts besides dialogue. Two such contexts can be considered, both of which extend the situation alignment mechanism into the domain of language acquisition. The first concerns the alignment of situation models in which one of the interlocutors is in a prelingual, acquisition phase. This emphasizes the suggestion that alignment can take place via nonverbal influences. Second, in the current formulation, the process of alignment and the formation of routines takes place on the time scale of single dialogues; however, these mechanisms can also be considered to span time frames that greatly exceed a single dialogue, particularly in the case of familiar repeated situations (feeding, bathing, playing), yielding "virtual dialogues" that can span a time period of several months. In such a situation, we can consider the formation of routines in the context of language acquisition to be analogous to the development of grammatical constructions.

Language acquisition can be functionally defined as the process of establishing the relation between sentences/discourses and their meanings. A significant part of this problem concerns the issue that before these relations can be established, the speaker and listener should be aligned with respect to the target meaning. If the meaning for the target utterance is not established both for the speaker and the listener, then construction of the mapping from utterance to meaning is indeterminate. This suggests the required existence of extra- or prelinguistic alignment mechanisms. Interestingly, there is indeed a significant body of research indicating that by six months of age, human infants achieve prelinguistic situation alignment by exploiting joint attention cues (e.g., gaze direction, postural orientation) in order to identify intended referents (e.g., Morales et al. 2000; Tomasello 2003). This indicates that P&G's Figure 2 could be modified to include nonlinguistic inputs at the semantic and situation model levels. Such a modification will allow both the "alignment bootstrapping" in which initial situation model alignment will play a crucial role in language acquisition as well as the influence of extralinguistic inputs in adult alignment contexts.

In a related extension of the alignment model into the acquisition domain, we can consider the relation between the development of production and comprehension routines in the time frame of a single dialogue and the development of grammatical constructions in the time frame of the first years of language acquisition. As specified by P&G, the creation of routines requires a coherent context in which the routines are applicable, and so, stretching this time frame to the scale of months and years is a non-negligible issue. Interestingly, Tomasello (2003) notes that repetitive events such as feeding, bathing, playing, and so on are relatively similar from episode to episode, and thus provide appropriate contexts that coherently span significant time periods. Given a temporally extended "virtual dialogue" domain, we can consider the development of routines as facilitatory not only within the context of a single dialogue but also in the more fundamental role of the development of communicative conventions that span significant time periods, thus forming the basis for language acquisition. In this context, routines take on the alternative identity of grammatical constructions (see Goldberg 1995), with all of their processing advantages. In particular, as described by P&G, the use of routines significantly eliminates the need for syntactic derivation of the appropriate grammatical structural forms, both for production and comprehension. When this approach is applied at the acquisition time scale, it is remarkably similar to the usage-based developmental approach to language acquisition advocated by Tomasello (2003).

In this framework, relatively fixed grammatical forms are linked to their corresponding meanings in the context of repetitive events (e.g., feeding, playing, etc.). These constructions/routines are then progressively opened to allow generalization within a given construction (e.g., variable replacement) to form new instances, and subsequent generalization to new constructions. Again, in both P&G's dialogue context and Tomasello's development context, highly functional communicative form-meaning constructions/routines are developed without reliance on a heavy initial investment in generative syntactic capabilities.

I have recently performed a series of simulation (Dominey 2000) and robotic (Dominey 2003a; 2003b) experiments to determine the feasibility of this type of approach to language acquisition in a restricted context. The underlying assumptions in the model are (1) that grammatical constructions correspond to the learned mapping between a given sentence type and its corresponding meaning frame (see Goldberg 1995), and (2) that grammatical constructions are uniquely identified by a limited set of cues that include word order and grammatical morphology including free and bound morphemes (Bates & MacWhinney 1987).

The model is provided with ⟨sentence, meaning⟩ pairs as input and should learn the Word-to-Referent and Sentence-to-Meaning mappings. For the current discussion, we assume that a limited set of concrete, open-class elements have been learned and will consider how this knowledge allows the learning of simple grammatical constructions. When a (sentence, meaning) pair is presented, the configuration of closed-class (function) elements is extracted and used as an index to "look up" the corresponding construction (routine) in the construction inventory. The construction corresponds to the learned mapping of open-class element positions in the sentence onto their thematic and event roles in the meaning representation. If there is no entry in the construction inventory (i.e., the current sentence type has never been previously encountered), then the construction is built on the fly by matching the referents for the open-class words with their respective roles in the meaning representation. The construction is then stored for future use. The developmental aspects of this learning are presented in more detail in Dominey (2000).

Thus, similar to P&G's routines, constructions are built by pairing the grammatical form with the aligned meaning (situation) representation. The interesting suggestion is that, at least to a certain degree, P&G's proposed situation alignment and routine construction capabilities provide a mechanism for language acquisition (at least the learning of fixed grammatical constructions that can generalize to new instances of the same constructions) which avoids the enlistment of generative grammar mechanisms. If a situation alignment priming mechanism could be demonstrated to perform in both the dialogue and acquisition time scales, this would be evidence for an ingenious economy of functional mechanisms for language processing in the context of dialogue.

Production-comprehension asymmetries

Fernanda Ferreira

Department of Psychology, Michigan State University, East Lansing, MI 48824-1117. **fernanda@eyelab.msu.edu**
http://eyelab.msu.edu/people/ferreira/fernanda.html

Abstract: Pickering & Garrod's (P&G's) mechanistic theory of dialogue is a major advance for psycholinguistics. But the commitment to representational parity in production and comprehension is problematic. Recent research suggests that speakers frequently produce a structure that listeners find ungrammatical and have trouble understanding. If the grammars of the two systems are different, then the assumption of representational parity must be relaxed.

The field of psycholinguistics has needed precisely what Pickering & Garrod (P&G) provide: a mechanistic theory of dialogue. P&G's core idea is that priming is more than just a tool cognitive psychologists use to learn about the structure or processing characteristics of some cognitive system; in dialogue, priming is itself a mechanism for producing alignment between interlocutors, and from it ultimately emerges the common ground that is critical for successful communication. But their views seem to commit them to the idea that the production and comprehension systems use the same representations, and this is one problematic aspect of their approach.

If the same representations are used for comprehension and production, speakers should not make use of a syntactic construction that comprehenders find unacceptable and hard to understand. Yet recent work from our laboratory demonstrates that they do (Ferreira & Swets 2003). We invented a paradigm for eliciting sentences such as "This is a dog that I don't know what it eats," which is ungrammatical in English. In the experiments, people first see a picture of a cat (for example) combined with a short verbal label such as "eats fish." Then they see a picture of another cat with a minimally different label – "eats chicken," for example. The next picture in the series consists of a cat and a question mark in the same spot where the labels occurred in the other examples. About 70% to 75% of the time, speakers will describe this third picture with a sentence containing a relative clause that violates island constraints (Chomsky 1973) and a resumptive pronoun in place of the illegal trace (Creswell 2002). (Henceforth we refer to these as IRPs: Island Resumptive Pronoun sentences.) This experiment, then, provided evidence that speakers can reliably be induced to produce a structure that is ungrammatical. But perhaps this result is not surprising, as such sentences occur fairly often in natural situations (Prince 1990).

More interesting were the results from a follow-up study in which speakers performed the same task but under time pressure. Some views of the IRP structure assume that these forms result from a failure to plan adequately. The idea is that speakers paint themselves into a syntactic corner and at the last moment try to salvage the utterance by inserting the resumptive pronoun. Yet, we observed that this form was no more likely to occur when speakers were under time pressure, and alternate, more acceptable forms were actually more frequent (e.g., "This is a dog and I don't know what it eats," or "This dog, I don't know what it eats"). It appears, then, that the production system plans this form and considers it part of its expressive repertoire. After all, the meaning that the utterance attempts to convey is not rare, and the IRP form is really the only way to verbalize it succinctly (neither of the paraphrases provided here as examples of alternate forms are as semantically accurate).

Now, the interesting question is: How does the comprehension system feel about these same structures? We knew from the start that they are generally viewed as ungrammatical (McDaniel & Cowart 1999). But, to be certain, we gave our listener-subjects written versions of the sentences that the subjects from the production experiments had uttered. Because this form is generally encountered only in spoken language and therefore is perhaps more easily processed with appropriate prosody, we conducted an auditory version of the same grammaticality judgment task. We obtained the same results. We were not shocked at these findings, though, because our speaker-subjects often giggled nervously after producing IRPs, betraying their awareness that the utterances were off in some way. We then conducted two more comprehension studies, one showing that comprehenders do not answer questions about IRPs as accurately as questions about matched controls (e.g., "This is a dog that doesn't know what it eats"), and the other demonstrating that IRPs elicited more regressive eye movements, launched from the clause containing the resumptive pronoun and aimed at the head noun of the relative clause.

It appears that the production system makes reliable use of a structure that the comprehension system views as ungrammatical and has a hard time understanding. We believe these results suggest a disconnect between the two systems. Now, P&G might respond that this asymmetry could be because of processing differences rather than representational nonparity. There are three problems with this argument. First, P&G invoked parity to explain how interlocutors are able to complete each other's utterances; but certainly, if there are forms that only one system likes, then these predictive feats will not be possible. Second, it is hard to imagine that hearing an IRP sentence could prime a speaker to produce one, given that the interpretive system balks at the form. And third, and most important, the syntactic representations used by the two systems appear to be nonoverlapping: The IRP form is part of the grammar that the production system uses, but it is not available to the comprehension system.

The authors might respond that these studies were obtained in production and comprehension experiments conducted entirely independently. Perhaps if these structures occurred as part of an interactive dialogue, they would either fail to emerge (which is unlikely, given that they are attested in natural conversations) or they would be treated as licit and understood by listeners. This tack is not unreasonable, but it does make me uncomfortable and it should make the authors a bit nervous as well. For if we accept this line of argument, then all of the findings from basic psycholinguistic experiments that have been done in monologue situations are suspect. Moreover, this response would undermine the authors' fundamental goal, which is to provide a mechanistic theory of dialogue – a theory that is informed by decades of systematic research which has isolated the mechanisms of language processing through experiments that could only be done in monologue contexts. As a result, the authors' theory would become less attractive, because one major reason for its appeal is that it brings together two important traditions in linguistics. Better to question the assumption of representational parity than to cast doubt on three decades of significant results, many of which critically informed the authors' approach.

Visual copresence and conversational coordination

Susan R. Fussell and Robert E. Kraut

Human-Computer Interaction Institute, Carnegie Mellon University, Pittsburgh, PA 15213. **susan.fussell@cmu.edu robert.kraut@cmu.edu**
http://www.andrew.cmu.edu/~sfussell/Susan_Fussell.html
http://www-2.cs.cmu.edu/~kraut/>

Abstract: Pickering & Garrod's (P&G's) theory of dialogue production cannot completely explain recent data showing that when interactants in referential communication tasks have different views of a physical space, they accommodate their language to their partner's view rather than mimicking their partner's expressions. Instead, these data are consistent with the hypothesis that interactants are taking the perspective of their conversational partners.

We applaud Pickering & Garrod's (P&G's) attempt to explain one of the most basic features of human language – its dialogue struc-

ture. They provide a thought-provoking theory of dialogue in which coordination in message production occurs when interactants generate their messages from similar situation models and mimic their partner's production at the syntactic, semantic, lexical, phonological, and phonetic levels, based on primitive priming mechanisms. They argue that these alignment processes plus techniques for repairing misalignments are sufficient to explain most cases of what others have considered evidence of a deeper type of perspective-taking, in which speakers take their partners' mental states into account in forming their own speech.

We believe, however, that P&G's theory cannot completely explain recent data about language production. In our own work, for example, we find evidence across several experiments that when interactants in referential communication tasks have different views of a physical space, they accommodate their language to their partner's view rather than mimicking their partner's expressions (e.g., Fussell et al. 2000; 2003; Kraut et al. 2002; 2003). These data are consistent with the hypothesis that interactants are taking the perspective of their conversational partners.

Consider, for example, the case of deictic reference in a bicycle repair task (Kraut et al. 2003). In this task, one person (the "worker") performs a series of repair tasks under the guidance of a second person (the "helper"). The helper is located either beside the worker, where both can see and interact with the work area, or in a separate room connected only by an audio link. In a third condition, they are connected by an audio/video link through which the helper can see what the worker is doing but cannot interact with the work area. The conversations typically consist of helper's instructions followed by worker's actions, questions, or acknowledgments of understanding. Interactants can refer to task objects and locations with either extended linguistic expressions (e.g., "take the long dangling piece and put it in where the two large screws are") or shorter deictic references (e.g., "take *this* piece and put it *there*").

As Figure 1 shows, the ways in which workers refer to parts, tools, and other task objects depend on their partners' ability to see the work area. In the side-by-side condition, both helpers and workers can view one another and task objects, and both use a large number of deictic expressions. In the audio-only condition, the remote helpers cannot see the work space, and neither workers nor helpers use deictic expressions. The interesting case, from an alignment point of view, is the video condition. Here, the helpers can see the workers and work space but cannot point to objects in it. Under these conditions, helpers rarely use deixis. However, workers can point to task objects and they know that helpers can see them do so through the video link. They use deixis instead of matching the helpers' nondeictic expressions. If conversational alignment were driven by primitive priming mechanisms, then the workers should use nondeictic references in the video condition, after hearing helpers' uttering many of these expressions. (Because the helper could not be seen, he or she would have no way of using deictic expressions to match the worker's utterances.) In short, the way workers referred to task objects and locations depended upon what their partners could see, not the language their partners previously used to refer to these same objects and locations.

We believe these results demonstrate that one type of deep common ground – visual copresence – is assessed during message production, at least for lexical selection processes. Indeed, in experiments where the views can change, interactants often explicitly exchange information about what each can see, with phrases such as: "Can you see the table?" (Kraut et al. 2002; 2003).

P&G might argue that video-mediated communication is a nonprototypical dialogue setting and hence may elicit special processes of assessing deep common ground. Note, however, that the audio-only discourse from Garrod and Anderson's (1987) maze study is similarly nonprototypical. Rather than demonstrating that people in face-to-face dialogues use processes of verbal alignment in lieu of deeper considerations of common ground, Garrod and Anderson's results may indicate that people verbally align primarily when the context has been stripped of all other indicators of common ground.

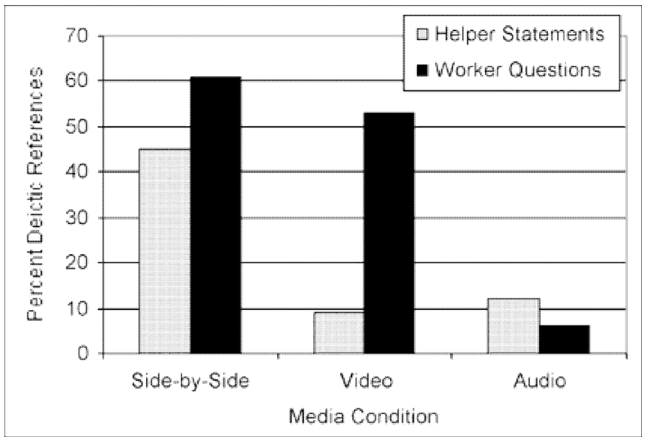

Figure 1 (Fussell & Kraut). Percentage of references to task objects containing verbal deixis, by media condition and participant role (from Kraut et al. 2003).

In their discussion of deep common ground versus automatic alignment, P&G in essence take a straw-man approach to describing the processes involved in conversational grounding. As Clark and Marshall (1981) have discussed, common ground can be determined using heuristics based on community comembership, linguistic copresence, and physical copresence. Some calculations of common ground (e.g., a helper trying to determine whether a worker on the bicycle task knows what a derailleur is) may be difficult; others (e.g., a worker trying to determine whether the helper can see the work space) may be relatively easy. Ruling out deep common ground as a fundamental process in dialogue production would require a series of carefully controlled studies that have not been performed to date. P&G's article is valuable for the detail with which it specifies an alternative model that would need to be included in such experiments.

We conclude by observing that calculations of deep common ground are essential for determining when to speak and what to say. For example, in the bicycle repair studies, in the audio condition, when the helpers cannot see them, workers describe what they are doing. Helpers rely on these verbal reports to determine when to provide new instructions or clarify preceding ones. In the video and side-by-side conditions, workers do not bother describing what they are doing because they know that the helpers are watching their activities. If common ground is available to communicators for these processes of message timing and content, it should not require much additional effort for them to incorporate it into the messages themselves.

Intrinsic misalignment in dialogue: Why there is no unique context in a conversation

Jonathan Ginzburg

Dept of Computer Science, King's College London, London WC2R 2LS, United Kingdom. **ginzburg@dcs.kcl.ac.uk**
http://www.dcs.kcl.ac.uk/staff/ginzburg

Abstract: Pickering & Garrod's (P&G's) claim that conversationalists do not explicitly keep track of their interlocuters' information states is important. Nonetheless, via alignment, they seem to create a virtually symmetrical view of the information states of speaker and addressee – a key component of their accounts of collaborative utterances and of self-monitoring. As I show, there is significant evidence for intrinsic contextual misalignment between conversationalists that can persist across turns.

I am very sympathetic with Pickering & Garrod's (P&G's) central message: namely, that there is a need to develop a detailed, computationally oriented account of the mechanisms underlying language use in dialogue, and moreover, that these – not mechanisms developed for monologue – should be regarded as the primary mechanisms of the language faculty.[1] On a more technical front, I think that a key claim of the authors, that dialogue participants do not explicitly keep track of their interlocuters' information states, but rather, that this is emergent from the dynamic alignment of each other's information states, is an important one.

Nonetheless, in seeking intrinsically to couple dialogue participants via alignment, they seem to create a virtually symmetrical view of the information states of speaker and addressee. Thus, a key component of P&G's accounts of collaborative utterances as well of self-monitoring is the claim that who is speaking at a given point does not, in some sense, make a difference given alignment: The addressee can take over or the speaker can "change voice" and self-correct. As I show in this commentary, this claim is incorrect: There is significant evidence that the contexts available to the conversationalists are *not* identical. Hence, there is actually intrinsic contextual *misalignment* between conversationalists that can persist across turns.

That a common context (cf. the *common ground* prominent in work by Clark & Marshall 1981; Lewis 1979; Stalnaker 1978) *emerges* in dialogue is an important insight. It yields a better picture of, for instance, querying, than classical speech-act views provide (Searle 1969). So in asking a question, to take one example, a speaker puts up a question for discussion and whoever takes over the turn can address it – either the original asker or the original addressee:

1. A: Whom should we invite to the conference?
   A or B: Would Phil be a good idea?

And yet, the actual situation is not as symmetrical as this: The speaker's options for self-repair or, indeed, other follow-up are quite distinct from the addressee's options. This can be illustrated succinctly by a phenomenon I have dubbed the Turn Taking Puzzle (Ginzburg 1997a; 1997b). Questions of the form "Why?" involve radical context dependence – pre-theoretically, the context supplies a propositional referent of some kind (Moore 1995). Interestingly, examples 2a and 2b (below) show that the resolution accorded to the bare "why" changes according to who keeps or takes over the turn. The resolution that can be associated with "Why?" if A keeps the turn is unavailable to B if he or she had taken over, and vice versa:

2a. A: Which members of our team own a parakeet?
    A: Why? (= Why own a parakeet?)

2b. A: Which members of our team own a parakeet?
    B: Why? (= Why are you asking which members of our team own a parakeet?)

2c. Which members of our team own a parakeet? Why am I asking this question?

Example 2c shows that these facts cannot be reduced to coherence or plausibility – the resolution unavailable to A in example 2a yields a coherent follow-up to A's initial query if it is expressed by means of a nonelliptical form. In other words, the context is responsible for these interpretational asymmetries or, rather, they are a consequence of the fact that distinct contexts are associated with the conversationalists.

Similarly, a common strategy for requesting a clarification is by means of a *reprise fragment* – a word or constituent of the previous utterance (see Purver et al. 2002; 2003 for corpus and experimental evidence on clarification requests, particularly reprise fragments). Reprise fragments have two prominent understandings (Ginzburg & Cooper 2004), exemplified in 3a (below). However, it is quite strange for a speaker to follow up her utterance with a reprise fragment. This becomes felicitous only if followed up by an additional correction, such as "Wait, did I say Bo, no I mean Lou," or some such. However, even then the readings that arise in example 3a, whose resolution is radically context-dependent, are not manifested:

3a. A: Did Bo leave?
    B: Bo? (= Either: Are you asking if BO of all people left? Or: Who were you referring to as "Bo"?)

3b. A: Did Bo leave?
    A: #Bo?[2]

It is worth noting that contextual asymmetries of this kind can persist for quite a number of turns – essentially as long as a given discourse topic remains under discussion. Example 4 is an extract from the British National Corpus in which Chris's "Why?" is naturally understood to refer to Norrine's utterance five turns back, an utterance which seems to be viewed as *grounded* (Clark 1996):

4. Norrine (1): When is the barbecue, the twentieth? (pause) Something of June.
   Chris (2): Thirtieth.
   Norrine (3): A Sunday.
   Chris (4): Sunday.
   Norrine (5): Mm.
   Chris (6): Why? (= Why do you ask when is the barbecue?)
   Norrine (7): Because I forgot (pause) That was the day I was thinking of having a proper lunch party but I won't do it if you're going out.

Note that the resolution associated with Chris's "Why?" is simply unavailable to Norrine at all subsequent points, as illustrated in the constructed variant of statement (4) in example 5a. As with previous examples, this cannot be explained on "pragmatic" grounds because the speaker can fairly coherently express the requisite reading in nonelliptical fashion, as in example 5b:

5a. Norrine (1): When is the barbecue, the twentieth? (pause) Something of June.
    Chris (2): Thirtieth.
    Norrine (3): A Sunday.
    Chris (4): Sunday.
    Norrine (5): # Why? (Cannot mean: Why is Norrine asking when is the barbecue?)

5b. Norrine (5′): Mm. Why am I asking? Can you guess?
    Chris (6′): No idea.

Phenomena such as this suggest that the different roles conversationalists play with respect to a given utterance (speaker vs. addressee) are not something that gets neutralized in the utterance's aftermath. The contextual possibilities available for one conversationalist differ from those of the other conversationalist (I am referring to dialogue here; as P&G point out, multilogue is a genre with various distinct properties from two-person dialogue). In other words, a single context is not fully adequate to describe dialogue, even when talking about "public" context, which results from overtly registered conversational actions. Instead, one needs to view dialogue as involving updates by each conversationalist of some type of a publicly accessible domain which is relative to each conversationalist and so is parametrizable by unpublicized factors such as individual goals and intentions (cf. Hamblin's *individual commitment slate,* Hamblin 1970). The framework named KOS spells out this view and develops theoretical accounts as well as computational implementations of illocutionary and metacommunicative acts, including a detailed account of puzzles such as the Turn Taking Puzzles exemplified above (cf. Cooper et al. 2000; Ginzburg 1996; 2002; forthcoming; Larsson 2002).

ACKNOWLEDGMENTS

The research described here is funded by grant number RES-000-23-0065 from the Economic and Social Research Council of the United Kingdom and by grant number GR/R04942/01 from the Engineering and Physical Sciences Research Council of the United Kingdom.

NOTES

1. The authors point out the dearth of work in mechanistic psychology and theoretical linguistics (primarily by syntacticians) on dialogue. Since the late 1990s, however, there has been work by formal and computational

semanticists precisely on developing theories of information states and their dynamics in dialogue – see, for example, work within the EU TRINDI project (Consortium 2000) and the annual SEMDIAL series of conferences on the semantics and pragmatics of dialogue. Links to all previous conferences, starting in 1997, in Munich, are available from: http://cswww.essex.ac.uk/semdial/

2. The # symbol is used here, as standard in linguistics, to mark an infelicitous utterance.

Dialogue: Can two be cheaper than one?

Sam Glucksberg

Psychology Department, Princeton University, Princeton, NJ 08544.
samg@princeton.edu **http://www.princeton.edu**

Abstract: Pickering & Garrod (P&G) argue that language processing in dialogue is in principle easier than in monologue. Although dialogue situations may provide more opportunities for facilitative priming, those priming mechanisms are also available in monologue situations. In both cases, the interactive alignment model calls strict modular accounts of language processing into serious question.

Pickering & Garrod's (P&G's) argument for the primacy of dialogue is a refreshing alternative to the isolated sentence as the unit of analysis in psycholinguistics. Others, such as Krauss, Clark, and their students and colleagues, have made similar arguments (cf. Clark 1996; Fussell & Krauss 1992), but none have made the additional argument that much of dialogue production and comprehension is automatic, requiring minimal cognitive or linguistic resources. The details of the priming processes have yet to be worked out in detail, but P&G provide enough empirical evidence to warrant optimism about an eventual fleshing out of their interactive alignment proposal. Beyond such details, their account has several intriguing implications, not the least of which is the claim that language production and comprehension are easier in interactive dialogue than in monologue. This flies in the face of the consensus that dialogue requires complex inferences about the mental state of one's interlocutor that are not required when one is, say, talking to oneself, or processing utterances under monologue conditions.

Is language processing in dialogues inevitably easier (more fluent, requires less processing, etc.) than in monologues? I do not think this question can be answered unequivocally. I can think of contexts in which dialogue language production would be quite difficult and contexts in which monologue production would be quite easy. An example of the former might be trying to hold a conversation with a sullen teenaged son or daughter on the merits of atonal music. Examples of the latter might include the think-aloud paradigm used in research on reasoning and problem-solving, or young children talking to themselves or to their dolls when no one else is around. So, as usual, the answer is: It depends.

A more useful question addresses the ways in which dialogue might be more likely than monologue to facilitate language processing. Such ways might in principle be available only in dialogue. Alternatively, it might just be a matter of likelihood, rather than a principled difference between monologue and dialogue. Consider the stages of speech production that could be facilitated in monologues and in dialogues. Following Levelt (1989), production could be facilitated at any or all of the following stages: the message (what to talk about); grammatical encoding (syntactic representation and lexical access); phonological encoding; and articulation. One of the more obvious ways to facilitate speech production is to provide something to talk about. In dialogue, the situation (including one's interlocutor) usually provides this, but of course there are (sometimes painful) situations where one does not know what to talk about next. In monologue, including the occasional classroom lecture (or a *BBS* commentary, for that matter), one can be at a loss as to what to say or write, but this is not a principled characteristic of monologues, just an unfortunate occasional one.

Grammatical and phonological encoding, on the other hand, may well be advantaged in dialogue via the mechanisms of priming and the opportunities for imitation. P&G make a convincing argument that such priming not only occurs but may well be automatic. However, such priming could in principle operate in monologue as well, although the opportunities are undoubtedly less likely. In monologue, as in dialogue, opportunities to repeat specific syntactic forms should facilitate successive productions. Similarly, opportunities to repeat specific word senses should facilitate lexical access and sense selection. What's missing from monologue, of course, are the contributions of an interlocutor, which can provide opportunities for imitation, as well as for syntactic, lexical, and phonological priming (but see Ferreira & Griffin 2003 for evidence of potential sources of priming in monologue).

Finally, P&G note that contextual constraints might operate more efficiently in dialogue than in monologue, but, again, this is probably not a principled difference. They write "With respect to lexical ambiguity, we predict that context will have a very strong role, so that effects of meaning frequency can be overridden" (sect. 5.4, para. 3). To my knowledge, this prediction has not yet been tested under dialogue conditions. However, it has been confirmed under monologue conditions, most recently by Sereno et al (2003), who found very early context effects on word recognition that overrode frequency effects.

To the extent that automatic priming can facilitate language comprehension and production, dialogue may well afford more opportunities for such priming than monologue. Nevertheless, the same mechanisms should be available for both. P&G's contribution is thus more profound than the observation that language processing in dialogue might be easier than in monologue. Instead, the contribution is in the insight that there may be mechanisms of language production and comprehension that can circumvent the need for costly cognitive processing in interactive conversation. People may not need to use a theory of mind or make inferences about common ground and mutual knowledge. They may, instead, rely on automatic priming both within and between interlocutors to produce alignment of representations, which, in turn, makes for successful and relatively effortless speech processing. The theoretical implication for theories of language processing is clear: A strictly modular, bottom-up language processing model simply will not do the job.

ACKNOWLEDGMENT
I thank Boaz Keysar for helpful comments on an earlier draft of this commentary.

Resonance within and between linguistic beings

Stephen D. Goldinger[a] and Tamiko Azuma[b]

[a]*Department of Psychology, Arizona State University, Tempe, AZ 85287-1104;* [b]*Department of Speech and Hearing Science, Arizona State University, Tempe, AZ 85287-0102.* **goldinger@asu.edu** **azuma@asu.edu**

Abstract: Pickering & Garrod (P&G) deserve appreciation for their cogent argument that dialogue merits greater scientific consideration. Current models make little contact with behaviors of dialogue, motivating the interactive alignment theory. However, the theory is not truly "mechanistic." A full account requires both representations and *processes* bringing those representations into harmony. We suggest that Grossberg's (1980) adaptive resonance theory may naturally conform to the principles of dialogue.

In their target article, Pickering & Garrod (P&G) present a clear and compelling case: Despite its being the predominate form of linguistic behavior, dialogue has been unduly ignored. Moreover, once dialogue is studied, it displays properties of entrainment at all levels (semantic, syntactic, phonetic, and others), which few psycholinguistic theories can address. To fill this void, the authors propose the *interactive alignment* model, wherein interlocutors share information at all linguistic levels, quickly arriving at states near equilibrium, allowing people to communicate with tremendous ease and efficiency. A by-product of alignment is that people tend to converge in dialogue, using similar expressions, words, and phrase structures. Beyond such linguistic constructs, interlocutors also tend to converge in their manner of speaking and even their physical postures (Capella 1981; Giles et al. 1991; Newtson 1994; Shockley et al. 2003).

The interactive alignment model is advanced to explain these coupling dynamics, using standard priming mechanisms. This approach is representation-driven, wherein one person's utterances prime ideas, words, and syntactic structures in his or her interlocutor. Mutual priming persists as a loosely coupled feedback loop (Van Gelder & Port 1995), so the same forms tend to occur repeatedly in conversation. At a general level, we believe this idea must be correct. However, the proffered model is underspecified, stated without explicit processes that act on the activated representations. Thus, the model nicely characterizes the problem space for a psychology of dialogue, but leaves many open questions. For example, how does the priming of extant representations predict new entities, such as routines, new jargon, or momentary changes in articulatory habit? How does speech perception lead to cascaded priming across linguistic levels? In short, what are the *processes* that explain the emergence of new psychological structures when people engage in dialogue?

In both language perception and production, models based on interactive activation have enjoyed great success (e.g., Dell 1986; Gaskell & Marslen-Wilson 1997; Levelt et al. 1999; McClelland & Elman 1986; McClelland et al. 1989). However, as P&G observe, these are all monologue models. Moreover, all generally assume fixed networks of nodes and connections, with weighting schemes that develop over extensive periods of training. Given their architecture, it is difficult to imagine how such models could communicate at multiple "levels," as suggested in Figure 2 of the target article. Of greater importance is that the core processes of interactive activation are not amenable to rapid transformation of representations or combinatorial rules. If one interlocutor can rapidly change another's internal weights, people would be subject to "catastrophic unlearning" of language.

As an alternative approach, we believe that interactive alignment may naturally "fall out" of processing in Grossberg's adaptive resonance theory (ART; Grossberg 1980; 1999; Grossberg & Myers 2000; Grossberg & Stone 1986; Grossberg et al. 1997; Vitevitch & Luce 1999). In speech perception, a basic hypothesis of ART is that conscious percepts are emergent products of resonant brain states. This is an interdependent system: Perception occurs when bottom-up and top-down knowledge sources bind into stable states. Processing begins when feature input activates *items* (feature clusters) in working memory. Items, in turn, activate *list chunks* in memory. Chunks are products of prior learning (perhaps prototypes) corresponding to feature combinations, such as phonemes, words, or common expressions.

Once items activate list chunks, a feedback cycle begins. Items feed activation upward through synaptic connections, and input-consistent chunks return activation. If items receive sufficient top-down confirmation, they continue sending activation upward. Within limits, this feedback loop (a *resonance*) is self-perpetuating, binding the respective activation patterns into a coherent whole. The bottom-up pattern that initiates interactive activation need not perfectly match its resultant feedback for resonance to occur. Cooperation between "levels" and competition within "levels" smooth out small mismatches, but large mismatches prohibit resonance. This stipulation seems necessary for dialogue, as another person's utterances (even their manner of producing phonemes) may never perfectly match one's stored representations.

Rather than assuming tiered representations, chunks in ART are *attractor states.* Of particular importance is that resonance can momentarily bind attractors in original combinations, allowing the creation of new mental structures (Carpenter & Grossberg 1987). This is the natural basis of communication, wherein novel utterances are constructed from familiar parts. In ART, however, those "parts" are not relegated to modular levels of representation: Although familiar patterns (e.g., common words) will cohere quickly in processing, all structures – both familiar and novel – must self-organize through competitive dynamics. Hence, in a dialogue, different structures, such as nascent jargon or momentary routines, can naturally emerge, as P&G describe.

As we have argued elsewhere (Goldinger & Azuma, in press), the self-organizing nature of ART makes it uniquely suited to model context-sensitive perception, as dialogue seems to require. For example, when ART recognizes a word spoken by Mary, its emergent percept will reflect the unique aspects of her utterance, allowing token-specific encoding (Goldinger 1998). Hence, the model has a natural basis to engage in dialogue, where context-specific cognition is pervasive. However, we are not suggesting that ART can currently explain the wide-ranging behaviors described by P&G. ART has successfully simulated many perceptual and memorial behaviors (among others; see Carpenter & Grossberg 1991). However, full sentence processing, even in monologue, remains a serious challenge.

As befits the study of dialogue, we envision a two-way street with equal benefits to students of dialogue and students of self-organizing networks. The properties of ART may give substance to the metaphoric theory described by P&G. By the same token, the unique challenges of dialogue may force changes in ART: To study dialogue, it would be ideal to have separate, self-contained ART simulations that communicate with each other through one or more channels, working together toward some shared goal (as in the authors' maze task). To our knowledge, no such study has ever been conducted using ART. However, a conceptually similar study by de Boer (2000) suggests that dialogues between autonomous models can evoke self-organization. De Boer created speech-perceiving "agents" equipped with articulatory synthesizers, perceptual systems that calculated the differences between signals, and associative memories for storing signals. Each agent was able to interact with others in an "imitation game." The agents' task was to imitate other agents accurately with a large repertoire of vowels; they could update their vowel repertoires based on their interactions. In many simulations, de Boer found that stable vowel systems consistently self-organized from these simple "dialogues." Moreover, these emergent vowel systems were remarkably similar to those found in human languages. This suggests that self-organizing structures may partially drive the basic forms of language. It also suggests that dialogue is a powerful force that should be studied both empirically and computationally.

ACKNOWLEDGMENTS

Support for this commentary was provided by grants R01-DC04535-03 (SDG) and 1-R03-DC4231-0182 from the National Institute of Deafness and Communicative Disorders (NIH). We thank Paul Luce, Greg Stone, and Guy Van Orden for many helpful discussions on the topics of resonance and speech units.

Dialogue in the degenerate case?

Patrick G. T. Healey

Department of Computer Science, Interaction, Media, and Communication Research Group, Queen Mary, University of London, London E7 0QY, United Kingdom. **ph@dcs.qmul.ac.uk** **http://www.dcs.qmul.ac.uk/~ph**

Abstract: The interactive alignment model treats within- and between-individual co-ordination as essentially equivalent. It is argued that this leads to a conservative account of alignment that presupposes high levels of linguistic and conceptual co-ordination. Data from the maze task are used to argue that this approach is not sufficient to account for important co-ordination phenomena.

Pickering & Garrod (P&G) bring together a large body of empirical evidence that highlights an important phenomenon. Participants in dialogue display higher levels of phonological, lexical, syntactic, semantic, and situational co-ordination than is expected from considering processes of production and comprehension in isolation. The authors propose that the central mechanism driving this level of communicative alignment is priming. This is envisaged as an automatic, bidirectional process operating in parallel on several different levels of representation. Importantly, they suggest that this process operates in essentially the same way both within and between speakers.

The treatment of dialogue participants as effectively interchangeable embodies a strong claim. It presupposes a high degree of equivalence between the representational systems underpinning communicative co-ordination. In places this is qualified as equivalent "in all essential respects" or "basically the same." However, in cases where priming is the mechanism of alignment, the residual differences between individual's representations are irrelevant.

What could underwrite the requisite similarity in mental representation? Priming is itself a conservative mechanism that alters the accessibility but not the form or content of the associated mental representations. For example, the account of co-ordination in the maze game distinguishes a basic repertoire of situational models of the maze (figure scheme, path scheme, line scheme, and matrix scheme) and a variety of referring expressions built up from the lexicon, syntax, and semantics of a fragment of English (Garrod & Anderson 1987). By hypothesis, the same basic repertoire of mental models and fragment of English are available to both participants prior to co-ordination. Priming does not change the repertoire; rather, it promotes alignment by favouring one particular model and its associated referring expressions.

Interactive alignment predicts that when maze game participants co-ordinate on a particular description and situational model, it should become more strongly primed, and they should stick with it. However, several aspects of the experimental data do not fit this prediction. Evolution and development of description schemes are common. As P&G observe, maze game pairs often develop their own idiosyncratic description schemes. Pairs switch description schemes frequently; for example, the isolated pairs in Garrod and Anderson (1987) and the non-community group in Garrod and Doherty (1994). Moreover, pairs do not reliably co-ordinate on the scheme that is most highly primed initially. For example, in Garrod and Doherty's (1994) community group, the line scheme is most frequent initially, but they subsequently co-ordinate on a matrix scheme.

In order to account for phenomena of this kind and to provide for co-ordination where participants are not effectively interchangeable, P&G propose iterative interactive repair as a key additional co-ordination mechanism. If a participant cannot find an interpretation for an utterance, he or she either shifts perspectives until an interpretation can be found or reformulates. This suggestion is not developed in detail, but it appears P&G envisage this basic form of repair as a shift between different co-ordination equilibria in the sense of Lewis (1969).

Data from the maze task suggest that this is inadequate. The results reported by Garrod and Anderson (1987), Garrod and Doherty (1994), and Healey (1997; in preparation) indicate underlying patterns of migration from figure and path schemes to more abstract line and matrix schemes. In addition, when maze game pairs have trouble co-ordinating, there is evidence that they show a reliable preference for shifting to a more basic (figure or path) scheme. The direction of these shifts is not predicted by frequency of prior exposure (Healey 1997; in preparation).

These observations show that the choice between different situational models is not neutral in the way that a co-ordination game or priming mechanism presumes. It is sensitive to the particular properties of the different situational models and, presumably, this relates to their implications for co-ordination. The pattern of preferences for shifts in description scheme and/or reformulations needs to be accounted for by the co-ordination mechanisms.

P&G aim to provide a model of dialogue co-ordination that avoids implausible assumptions about interlocutors constructing elaborate models of each other's context and mental states. This is surely right. Explicit negotiation and repair are relatively rare and do not provide a general account of co-ordination. However, the mechanisms of the interactive alignment model do not seem to do the work required.

Ironically, the idealisation of speaker and hearer as interchangeable reproduces one of the problems with treating dialogue as a form of monologue: the implication that participants are linguistically and conceptually transparent to one another. Interactive alignment focuses on dialogue in the degenerate case: interactions in which people are, in a sense, already co-ordinated. As P&G note, dialogue is important partly because it is the primary context for exposure to and acquisition of language. But this is also the situation in which the assumptions embodied in the interactive alignment model are least likely to be satisfied.

Interactive alignment: Priming or memory retrieval?

Michael Kaschak[a] and Arthur Glenberg[b]

[a]*Department of Psychology, Florida State University, Tallahassee, FL 32310;* [b]*Department of Psychology, University of Wisconsin, Madison, WI 53706.*
kaschak@psy.fsu.edu **glenberg@facstaff.wisc.edu**

Abstract: Pickering & Garrod's (P&G's) interactive alignment model explains the existence of alignment between speakers via an automatic priming mechanism. We propose that it may be preferable to explain alignment through processes of memory retrieval. Our discussion highlights how memory retrieval can produce the same results as the priming mechanism and presents data that favor the memory-based view.

Pickering & Garrod (P&G) claim that dialogue is marked by a great degree of alignment between interlocutors. The alignment is taken to arise from the operation of a largely automatic priming mechanism. We suggest that much of what P&G ascribe to the operation of a priming mechanism can be captured by a memory-based view of language processing.

The memory-based view is rooted in the assumption that language processing is performed against the background of one's experience. The comprehension of an incoming utterance is guided by experience of utterances with similar structures and meanings, just as experience communicating via speech shapes the production of utterances. This assumption forms the basis of many extant psycholinguistic theories: constraint-based approaches to sentence processing (e.g., MacDonald et al. 1994), memory-based models of text processing (Myers & O'Brien 1998), and theories of lexical access (Goldinger 1998). The memory retrieval mechanism is what has been called a "global matching" process (e.g., Hintzman 1986). That is, the memory retrieval process involves a global, parallel search of memory that is influenced by both contextual and temporal factors (e.g., the retrieval process shows a recency effect when recently processed language is treated as a separate context from previously processed language).

The memory-based view explains why processing appears different in dialogue and monologue situations. It also provides an alternative to the priming mechanism around which P&G build their theory. For monologue experiments, which involve processing language in a decontextualized and artificial situation, there is little in the way of specific prior experience that might constrain the memory retrieval process. The structure of the experiments (e.g., randomized trials, random selection of materials) also makes it unlikely that recent experience within the session will exert a systematic influence on task performance. Hence, monologue-based experiments produce memory retrieval conditions that are sensitive to general properties of one's linguistic experience, such as the frequency of particular syntactic structures or word meanings. On the other hand, dialogue situations foster highly constrained memory retrieval. For example, the perceptual similarity between utterances within a dialogue can constrain memory retrieval so that it most strongly resonates with previous experiences with that particular speaker (see Goldinger 1998). All other things being equal, memory retrieval is also weighted toward more recent experience (see Bock & Griffin 2000). Together, these factors can produce the kinds of repetition and alignment that P&G discuss.

Support for the learning- and memory-based view comes from experiments reported in Kaschak (2003). The participants in these experiments were trained on a syntactic construction with which they were not familiar, called the Needs construction. In the first part of the experiments, participants in the Needs training condition were exposed to the novel construction in sentences such as: "*The meal needs cooked* given that dinner is in an hour." Participants in the control training condition had an identical training phase in which they read the standard version of the Needs construction for their dialect ("*The meal needs to be cooked* given that dinner is in an hour"). By the end of the training phase (which included 10 to 12 exposures to the Needs or standard constructions, depending on the experiment), participants in the Needs training condition read the novel construction as quickly as participants in the control training condition read the standard construction. This is consonant with P&G's claim that recent experience has a strong influence on language processing. Indeed, it fits nicely with the claim that recent experience can override global frequency effects in determining the speed of processing for particular kinds of language.

One of the questions addressed in these experiments was how training on the Needs construction would affect the processing of another construction with which the participants were already familiar: the modifier construction ("*The meal needs cooked vegetables* given that the guests are vegetarian"). Training on the Needs construction sets up an ambiguity between the novel structure and the existing modifier construction, such that the sentence is ambiguous at the word "cooked." The priming mechanism favored by P&G (and discussed earlier in Pickering & Branigan 1998) makes a straightforward prediction about the influence of Needs training on the processing of the modifier construction. Recent exposure to Needs sentences will prime the syntactic features associated with the construction, leading readers to initially choose the Needs interpretation of the ambiguous sentences (thereby slowing processing of the modifier construction).

Kaschak's (2003) results showed that the effect of Needs training on the reading of the modifier construction depends on the particular nature of one's experiences during training. If participants reading, "The meal needs cooked . . ." during training initially interpreted the sentence as the modifier construction before figuring out that the sentence was an example of the Needs construction, Needs training facilitated the participants' reading of the modifier construction. If the participants were given instructions such that they did not initially misinterpret the Needs sentences as examples of the modifier construction, the Needs training did not facilitate processing of the modifier construction. Kaschak (2003) discusses how these data can be explained by assuming that the participants in these experiments remembered not only the outcome of processing the initial examples of the Needs construction, but also the processing work that went into comprehending those sentences. This leads to the conclusion that it is the memory that one has for particular linguistic experiences, rather than a priming mechanism, that may better explain how recent experience with language shapes the way subsequent utterances are processed.

We resonate with many of the arguments and claims advanced by P&G, but we think their explanation can be substantially improved by replacing the notion of alignment-by-priming with alignment based on retrieval from memory of recent processing episodes.

Grammars with parsing dynamics: A new perspective on alignment

Ruth Kempson

Philosophy Department, King's College London, London WC2R 2LS, United Kingdom. **ruth.kempson@kcl.ac.uk**
http://www.kcl.ac.uk/kis/schools/hums/philosophy/top.html

Abstract: This commentary argues that dialogue alignment can be explained if parsing-directed grammar formalisms are adopted. With syntax defined as monotonic growth of semantic representations as each word is parsed, alignment between interlocutors is shown to be expected. Hence, grammars can be evaluated according to relative success in characterizing dialogue phenomena.

Although Pickering & Garrod (P&G) suggest grammars should be compatible with alignment patterns in dialogue, commending multilevel constraint-based formalisms, they could have posed a stronger challenge. Given that dialogues constitute primary data, can linguistic theory provide a basis for explaining alignment? Does some grammar-external mechanism determine alignment, or is it a consequence of the architecture of the language system and the way it is put to use in dialogue? The authors opt for the first alternative. However, with grammar formalisms being promoted which reflect the dynamics of a parser (Kempson et al. 2001; Phillips 1996), alignment can be more strongly buttressed by natural-language grammars than they anticipated.

In *Dynamic Syntax* (Kempson et al. 2001), a grammar formalism is defined that provides such underpinnings. A constraint-based architecture is defined, with syntax as the progressive growth of semantic representations, specifying possible parse routines. That syntactic and semantic alignment go hand in hand, is thus ensured. Semantic representations (logical forms) are in tree format, with the propositional formula assigned to a string decorating the top node and the dominated nodes in the tree decorated with sub-terms of this formula (quantification is expressed in the form of names denoting witness sets, so resulting representations resemble the *situational models* of Johnson-Laird 1983). Growth of logical forms, central to this concept of syntax, is goal-directed – first setting out the overall goal to establish a propositional structure, with subgoals introduced for predicate-argument structure. The partial structure induced by such goals is enriched by updates which the words in sequence provide for these skeletal structures; and these, together with construal of pronouns relative to context (context construed as a sequence of semantic representations), determine the building up of a semantic representation which, when complete, meets the goal (a *goal tree*).

Lexical specifications are also defined in terms of tree growth, with each word associated with some input it provides to progressive articulation of semantic representations. That lexical alignment should go along with syntactic/semantic alignment thus follows. The apparent syntax-specific alignment across interlocutors relative to constant semantic content, for example, repeating double-object constructions rather than their full dative equivalent – a putative counter-example to such collapse of syntactic/semantic/

lexical distinctions – follows from this definition of lexical content. A word with both double-object and full-dative variants provides two distinct forms of update onto a single structure, hence a processing ambiguity. Repeating words across interlocutors will retrieve the lexical specification just used, not some alternant. By collapsing what would in other systems be discrete syntactic/semantic/lexical levels, the multilevel nature of alignment is transformed into a single phenomenon.

We can now see perception/production correspondences and self-monitoring of production as essential, not merely functionally convenient. The consequences of taking parsing as basic when designing a production model are twofold. First, in using the same grammar formalism, the production mechanism must be using the same structure-inducing growth process as is articulated for parsing, the major difference being that in production, the goal tree is known. Following general tree development steps defined by the parser, which in the initial steps of the parse define only a skeletal partial tree, the task is to retrieve from the lexicon some appropriate first word that can take the partial tree so far defined and update it to provide a more developed structure that subsumes the goal tree. If such a subsumption relation holds, then utterance of that word is licensed. This task repeats itself to induce a sequence of words until the goal tree is obtained through the parse mechanism – hence the correspondence between comprehension and production and essential self-monitoring in production.[1]

Alignment between interlocutors is now explicable. The major challenge of production, given some conceptual structure to be communicated, is selecting words to express that content, imposing a general lexicon search. The incrementality of the tree growth mechanisms suggests that this search is activated successively, which is a huge task. However, production, like parsing, is context-dependent; and part of the context is that small subset of the lexicon already accessed. Repeating words accordingly avoids a general lexicon search. Pronouns equally, being place-holders substituted by contextually available representations, sidestep full lexicon access. And with elliptical fragments, lexical search is restricted to the fragment. So the high rate of alignment, as displayed by these forms, arises because production, following the parsing dynamics, is context-dependent, which is essential for minimizing the word-retrieval burden.

Finally, shared utterances are expected. Given the use of parsing tools to induce production steps, in successful communication, interlocutors must coincide on constructing some particular sequence of structures. The shift from parser to producer is thus straightforward if the parser, having constructed some partial tree, makes an abduction step to determine what is needed to complete it. The shift from producer to parser is equally natural: It is a shift into the task of processing lexical input to complete the partial tree, which up to that point was constructed as the means of making production choices.

Hence, there is a stronger conclusion than the authors' modest challenge to linguists. Rather, we might adopt methodologies in which linguistic theories are evaluated by their potential for expressing coordination of comprehension and production in dialogue. Grammar formalisms defined in terms of parsing meet this challenge in a particularly direct way.

ACKNOWLEDGMENTS

This commentary is a reflection of the work of a team. Thanks are particularly due to Masayuki Otsuka and Matthew Purver for extensive discussions on generation modeling. However, nobody other than myself can be blamed for the particular dialogue perspective suggested here.

NOTES

1. Modeling generation using the dynamic syntax framework constitutes ongoing research (Purver & Otsuka 2003).

Is alignment always the result of automatic priming?

Robert M. Krauss and Jennifer S. Pardo

Psychology Department, Columbia University, New York, NY 10027.

rmk@psych.columbia.edu jsp2003@columbia.edu

http://www.columbia.edu/cu/psychology/commlab/

http://www.columbia.edu/~jsp2003/

Abstract: Pickering & Garrod's (P&G's) mechanistic theory of dialogue attempts to detail the psychological processes involved in communication that are lacking in Clark's theory. By relying on automatic priming and alignment processes, however, the theory falters when it comes to explaining much of dialogic interaction. We argue for the inclusion of less automatic, though not completely conscious and deliberate, processes to explain such phenomena.

In his influential book *Using Language,* Clark (1996) argued against a conceptualization of communicators as autonomous information processors, contending that language use is intrinsically a joint activity and examining communicators' practices from this perspective. His account provides a compelling description of some of the things talkers accomplish in dialogue, but it is weak on details of the psychological processes on which these accomplishments rest. Pickering & Garrod's (P&G's) mechanistic theory of dialogue is an attempt to provide such an account. In many respects, it is quite successful, providing a glimmer of light at the end of a long psycholinguistic tunnel. However, there are some respects in which the theory falls short of its authors' goal of providing a mechanistic explanation for the phenomena Clark described.

There are two key propositions in the target article's argument: first, that communication entails the alignment of participants' situation models; second, that priming is the principle mechanism by which this is accomplished. We find the first proposition more convincing than the second. By stressing the automaticity of the process, P&G's mechanistic theory appears incapable of accounting for the way interlocutors use information in what appears to be a more reflective fashion. Because of space limitations, we shall confine ourselves to just a few instances where the theory is deficient.

1. The automatic priming mechanism appears to leave no room for addressee accommodation in the absence of a misunderstanding, yet there are many examples of interlocutors taking their partners' informational needs into account that are incompatible with automatic priming. For example, Kingsbury (1968) found that Bostonians who were asked, "I'm from out of town, can you tell me how to get to Jordan Marsh?" gave more detailed directions than those simply asked, "Can you tell me how to get to Jordan Marsh?" When asked the latter question in an exotic (nonlocal) dialect, Bostonians also gave more detailed directions. Fussell and Krauss (1992) found that the number of words used in the initial reference to a photo of a landmark in an interactive coordination task was a function of the landmark's perceived identifiability – the more identifiable the landmark was thought to be, the fewer the words used to refer to it. It is not clear how priming could account for these results or those of a host of similar studies (see Krauss & Fussell 1996 for a review). We believe that such "audience design" effects (Clark & Murphy 1982; Fussell & Krauss 1989) occur prior to referent selection, and not just as an attempt to remedy an emergent misunderstanding.

2. Representational alignment requires that two or more entities be identical in some way. Assertions of identity may work for descriptions of relatively abstract syntactic and lexical levels of representation, but not for representations at the phonological level, because the phonological level is graded and repeated phonetic elements (even within the same talker) are not physically identical. Although different instances of a phoneme may be perceived as members of the same phoneme category, perception preserves some phonetic distinctions. Hence, the very level that is the point of contact between talkers, the acoustic-phonetic level,

cannot support an automatic alignment/imitation-based model because it is impossible to produce a perfect imitation – the monitoring system would be reporting continual error. Although the mechanism allows for degrees of alignment, we lack a rule for determining how much alignment is required.

Interestingly, the strongest evidence cited by P&G for phonetic alignment comes from Goldinger's (1998) study of lexical shadowing. However, Goldinger's procedure assessed perceived *imitation,* which is not equivalent to phonetic similarity. Imitations of the voices of well-known figures by vocal impressionists are caricatures that exaggerate particularly salient features rather than produce acoustically accurate reproductions. The remainder of the published evidence for phonological imitation is mainly of increased similarity in speech rate and pitch (Giles et al. 1991; Natale 1975a; 1975b) and voice onset timing (Sancier & Fowler 1997). In a continuously variable system, what degree of similarity constitutes an imitation?

3. Interlocutors' speech does not always become more similar over the course of their interaction; in some cases, interaction yields divergence rather than convergence. Moreover, the speech of different participants may change to different degrees; convergence can be radically asymmetrical. It would be little more than an annoyance if such departures from symmetrical convergence were random, but frequently they reflect social processes that are fundamental to the interlocutors' interpersonal relationship and the ways in which they define the interaction situation. For example, Bourhis and Giles (1977) found divergence in accentedness when a talker's ethnic identity was devalued. Gregory and Webster (1996) found that the symmetry of pitch convergence between a talk-show host and his guests depended on the guest's status relative to that of the host – not surprisingly, higher-status guests changed less than their lower-status counterparts. Again, it is difficult to reconcile such phenomena with an automatic priming explanation. It seems more plausible to suppose that they derive from a prior assessment that sets up the system to evoke particular kinds of priming.

Although our commentary is directed at what we see as deficiencies in P&G's theory, we applaud their attempt to move beyond participants' goals and intentions and focus on the psychological mechanisms that make dialogue possible. Their thoughtful article is admirable in both its scope and depth, and offers much to contemplate. A complete account, we believe, will require a hybrid model in which alignment or imitation derives from both the kinds of automatic processes they describe and processes that are more directed or reflective. Hybrid models of this sort may be less tidy (although not necessarily less mechanistic) than the one P&G propose, but they do seem necessary to capture the subtlety and richness of dialogic phenomena. We are reminded of an anecdote about French President François Mitterand, who, when asked by an acquaintance if she might address him using the personal *tu* form, responded, "Si vous voulez." Even in cooperative settings without misunderstanding, alignment may be used strategically – language is used in the pursuit of individual goals. An elaboration of how a situation model incorporates key aspects of social and interpersonal dynamics would increase the explanatory power of a mechanistic theory of dialogue.

One alignment mechanism or many?

Arthur B. Markman, Kyungil Kim, Levi B. Larkey, Lisa Narvaez, and C. Hunt Stilwell

Department of Psychology, University of Texas, 1 University Station, A8000, Austin, TX 78712. **markman@psy.utexas.edu**
kyungil@mail.utexas.edu larkey@mail.utexas.edu
grimmlr@mail.utexas.edu stilwell@psy.utexas.edu
http://www.psy.utexas.edu/psy/FACULTY/Markman/index.html

Abstract: Pickering & Garrod (P&G) suggest that communicators synchronize their processing at a number of linguistic levels. Whereas their explanation suggests that representations are being compared across individuals, there must be some representation of all conversation participants in each participant's head. At the level of the situation model, it is important to maintain separate representations for each participant. At other levels, it seems less crucial to have a separate representation for each participant. This analysis suggests that different mechanisms may synchronize representations at different linguistic levels.

Introduction. The core of Pickering & Garrod's (P&G's) article is illustrated in their Figure 2. Participants in a conversation are attempting to achieve alignment between linguistic representations at the phonetic, lexical, semantic, and discourse levels simultaneously. Two key questions arise from this figure. First, what sorts of representations are being aligned; and second, how is this alignment achieved?

What is being aligned? The target article's Figure 2 illustrates that linguistic representations at a variety of levels are aligned. The authors suggest that representations from one participant's head are aligned directly with those from the other participant's head. This notation is a convenient shorthand, but people cannot directly access each other's mental states. To make this model work as it is drawn, participants must keep track separately of their own knowledge at these levels, as well as the knowledge of other participants.

For some levels of discourse, this separation is more critical than others. For the situation model, it is important that speakers know what information is possessed by the listener. This information is crucial for ensuring that given new conventions are followed when generating sentences, for ensuring that new utterances are relevant to the discourse, and for maintaining common ground (Clark 1996; Sperber & Wilson 1986). A key question is the degree of information participants must have about their partner's knowledge (Keysar 1994; Keysar et al. 1998).

At other linguistic levels, the distinction between one's own knowledge and that of a partner may be less crucial. It may not be necessary to distinguish between one's own grammatical constructions and those of a partner. Similarly, representations of phonology and prosody of speech need not be kept separate in order to process a discourse. This distinction in the knowledge required to process and use language for the situation model as opposed to grammatical or phonetic representations suggests that these levels may differ in the degree to which people are aware of the effects of alignment. In particular, people may recognize that they have designed their utterances to convey particular kinds of information that their partner does not have. In contrast, they may be unaware that the grammatical, phonological, or prosodic form of one sentence has been influenced by the form of a sentence spoken previously by a partner.

How are representations aligned? The target article refers to the process that synchronizes participants' linguistic representations as alignment. The discussion of the roles of linguistic representations in the previous section suggests that there may not be a single alignment mechanism at work. In particular, the discourse-level representation (e.g., the situation model) is the only one that really seems to require a separate representation of what is known by each conversational participant. In contrast, the lower levels of representation need not have distinct representations for each participant.

When there are separate representations for each conversa-

tional participant, a structural alignment process like the one posited to be involved in analogical reasoning and similarity can be used to help synchronize representations (Gentner 1983; Gentner & Markman 1997). The structural alignment process takes pairs of relational representations and places them in correspondence. Structural alignment requires that the relations and objects that are matched be seen as identical (or can be decomposed into structures that are partially identical). Thus, this process provides a method for creating semantic parallels between people's situation models and their representations of conversational partners' situation models.

One advantage of thinking about the alignment of situation models as a structural alignment process is that there are several established empirical benchmarks that are hallmarks of analogical reasoning (Markman & Gentner 2000). It is possible to test for the presence of these benchmarks in discourse-level processing. For example, there is substantial interest in discourse on the role of inferences in comprehension (e.g., Graesser et al. 2001; McKoon & Ratcliff 1992). The structural alignment process suggests that inferences are made on the basis of systematicity. Specifically, a piece of information is inferred when there is a correspondence between two representations and there is information in one representation that is connected to the correspondence that can be carried over to the second representation (Clement & Gentner 1991; Markman 1997). Other information is not carried over as an inference. Thus, this view makes a testable prediction about what kinds of information are likely to be inferred by conversational participants. Similarly, experiments could be devised to explore the other published benchmarks.

It is not clear that the structural alignment process should be involved at other linguistic levels. If people do not maintain separate representations for themselves and their partners for grammatical and phonetic processing, then some other process must be involved in synchronizing these representations across individuals. The authors suggest one way this could occur. If the same representations are used for both production and comprehension at the level of grammar and phonology, then these representations will be synchronized by virtue of the nature of the language process. Representations at these levels would be influenced by factors such as priming (as suggested in the target article).

This analysis suggests that factors such as systematicity that should affect the formation of situation models should not have an influence on phonetic and grammatical processing. It would be interesting to carry out studies to explore whether systematicity and other benchmark phenomena of structural alignment affect other levels of linguistic processing. If they did, this would suggest that people are maintaining separate representations of their own processing and that of their partners at a variety of linguistic levels.

Beyond linguistic alignment

Allan Mazur

Center for Environmental Policy and Administration, Maxwell School, Syracuse University, Syracuse, NY 13244. **amazur@syr.edu**

Abstract: Dialogue requires ability beyond the production and comprehension of word strings. The interactive alignment account is good as far as it goes, but it must be embedded in a broader model encompassing alignment of paralinguistic representations.

Dialogue is not only the natural arena for language; it requires abilities that are different from the production and comprehension of sentences and phrases. Children with the variant of autism known as Asperger's Syndrome have good sentence skills but cannot carry on a normal conversation because they have difficulty with eye-to-eye gaze, facial expression, and body gestures that regulate the exchange.

Pickering & Garrod's (P&G's) ingenious account of interactive alignment illustrates gains that can be had in broadening the study of language from monologue to dialogue. I urge upon the authors a further expansion, embedding spoken dialogue within a broader model of human (and ape) turn-taking communication.

Conversations allow interlocutors to share meaning conveyed by body postures, facial cues, and intonation – apart from the particular strings of words that are used. Sometimes the words spoken are irrelevant to the primary representation that is to be aligned. A conversation between two knowledgeable people, ostensibly about Chaucer, may be more importantly a dominance contest, or a seduction, or bonding chatter between friends – though words manifestly about these meanings may never be exchanged.

We often recognize the relationship between conversing foreigners and the general intent of their communication, even though we may not understand their language.

As long as actors can see one another, they can communicate with remarkable efficiency even when distance or a barrier precludes speech. I recall standing with a handsome male colleague on a busy street corner. He whispered aside to me that he was visually flirting with a woman seated on a bus some distance away; he predicted (correctly) that as the bus drove off, she would turn toward him with a broad smile.

Paralinguistic signaling is especially effective in aligning social relationships within a dyad. This enables Ego and Alter to agree on who is the leader and who the follower; to communicate affiliation, affection, or hostility; and to sympathize emotionally. Consider, as an example, how dialogue can establish the relative status of interlocutors, independently of the words that are used.

Spoken or nonverbal "dialogue" is governed by mutually understood rules (Mazur 1985). Some rules are asymmetrical, specifying different actions for a high-status actor and for a low-status one. Two asymmetrical rules are:

1. *The high-status person sets the pace and mood of the conversation, and the low-status person follows.* Pace and mood may be set with smiles, jokes, frowns, exclamations, and volume, rapidity, or intonation of speech. If Ego tells a loud joke, Alter can deferentially comply with a loud laugh or can challenge by substituting an inappropriate response.

2. *The high-status person introduces and terminates major topics of conversation.* This rule, like the previous one, indicates that the high-status person can take control of the conversation, which is the essence of having high status. If both interlocutors attempt to set the conversational agenda, there is a dominance contest. Other rules are symmetrical, applying without regard for the status of the actor. It is the *violation* of symmetrical rules that signals a dominant act, whereas strict conformity to them signals deference or politeness. Important symmetrical rules are:

3. *If one individual is speaking, the other should remain quiet.* If Ego interrupts Alter's speech, Ego has acted dominantly.

4. *A listener who is offered the floor should speak.* A speaker can pass the floor by asking a question of the listener, or by directing his eyes to the listener after concluding a speech. If Ego remains silent after Alter offers the floor, Ego has acted dominantly.

5. *Do not look into another individual's eyes when no one is speaking* (unless in a romantic context). The violation of this rule, silent staring, is a common dominant act among primates, whereas rule-following eye aversion indicates deference.

6. *Look at the speaker's face, especially if the speaker is looking at you.* To look away, suggesting inattention, is hard to do if you respect the person speaking to you. If the speaker is of minor consequence, it is easy to violate the rule, thus showing your dominance. (This rule is inoperative when averted eyes overtly signal submission, as in looking down while being scolded.)

7. *Do not speak loudly, sternly, or angrily.* Shouting matches and arguments are obviously dominance contests.

8. *The speaker should direct the listener's actions by request rather than command, and should avoid a stern or stubborn tone.* To speak in a commanding or inflexible way implies that the listener is of lower status.

These rules operate within a context of linguistic interaction. The words spoken may or may not be an important part of the whole display. When we deferentially compliment someone, we speak in strict accordance with the rules, whereas our verbal insults gain emphasis when violating the rules of dialogue. We accompany our speech with appropriate gestures, perhaps glaring for dominance or smiling for deference. This full array of actions – words, gestures, and rules – constitutes the status display.

A natural model of conversation must go beyond the interactive alignment of word strings. P&G have briskly stepped onto the road of dialogue. I hope they soon take another step forward.

Correspondences between the interactive alignment account and Skinner's in *Verbal Behavior*

Joseph J. Pear

Department of Psychology, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada. **pear@cc.umanitoba.ca**
http://www.umanitoba.ca/cgi-bin/psychology/hpg_main.cgi?data/pear.txt

Abstract: Pickering & Garrod's (P&G's) interactive alignment account corresponds directly with the account Skinner (1957) gave in his book *Verbal Behavior.* This correspondence becomes evident when "properties of verbal stimuli" substitutes for "channels of alignment." Skinner's account appears to have the dual advantages of requiring fewer basic terms and integrating the field of verbal behavior with the whole field of human behavior.

There are strong correspondences between Pickering & Garrod's (P&G's) interactive alignment (IA) account and Skinner's (1957) verbal behavior (VB) account. Similar to P&G's assumption that dialogue is the basic form of language processing, Skinner took the interaction between speaker and listener as fundamental (see Skinner 1957, Figures 1–6, pp. 38, 39, 57, 84, 85). I explain further correspondences with reference to Table 1 below, which is based on Table 2 in the target article.

Regarding row 1, the situation models in the IA account correspond in the VB account with variables controlling each interlocutor's behavior, some of which may be private (Skinner 1957, pp. 130–46; also see Skinner 1953, Ch. 17). The alignment channels in the IA account correspond with properties of verbal stimuli. Because of reinforcement of many different instances of echoic behavior, which is the direct imitation of the properties of verbal stimuli, echoic behavior generalizes widely (Skinner 1957, pp. 55–56), just as the tendency to align generalizes across alignment channels in the IA account.

Regarding row 2, sustained dialogue occurs only if the interlocutors have implicit common ground in the IA account or emit similar verbal behavior in the VB account. In the VB account, effective verbal interaction occurs between the extremes of identical and completely dissimilar verbal behavior (Skinner 1957, pp. 271–72). In the IA account, there is a repair mechanism for preventing dialogue from breaking down or for restoring it if it does. In the VB account there are several reinforced response-strengthening techniques that maintain dialogue, such as prompts and probes (Skinner 1957, Ch. 10, pp. 253–92; see also p. 58 for examples of clarification and expansion requests involving echoic behavior).

What in the IA account is called full common ground – a complex model shared by both interlocutors – is not common in ordinary dialogue according to either account. Both accounts maintain that constructing and using these models are derived (e.g., learned) abilities or behaviors.

Both accounts also agree that we obtain a distorted view of grammar by looking at it only in monologue. In the VB account, grammar is the expression of responses called autoclitics that modify the effect on the listener of responses that Skinner (1957, p. 312) called "the raw material out of which sustained verbal behavior is manufactured." When listeners cannot frequently or immediately prompt or probe the speaker's verbal behavior, there is

Table 1 (Pear). *Correspondences between the interactive alignment account of language processing in dialogue and the verbal behavior account*

Interactive alignment account (IA)	Verbal behavior account (VB)
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.	1. *Linkage between interlocutors* Control across multiple properties of verbal stimuli. Sound comes to bring the listener's behavior under the control of variables controlling the speaker's behavior.
2. *Inference* Externalized in the interaction between interlocutors via a basic interactive repair mechanism.	2. *Inference* Externalized in the interaction between interlocutors via reinforced techniques of strengthening verbal behavior.
3. *Routines* Arise out of the application of the interactive alignment in specific situations. A high proportion of dialogue contains these units, which facilitate both production and comprehension.	3. *Routines* Functional verbal units that are conditioned or strengthened process. A high proportion of dialogue uses routines, which simplify both production and comprehension.
4. *Self-monitoring* Monitoring occurs at any level of representation that is subject to alignment as a consequence of the account.	4. *Self-monitoring* Occurs for all aspects of verbal behavior because speakers typically are also listeners.
5. *Repair mechanisms* The same basic repair mechanism for self-repair and other-repair.	5. *Repair mechanisms* The same basic principles apply for strengthening one's own verbal behavior and strengthening that of others in specific situations.
6. *Linguistic representations* Needed to deal with linked utterances in dialogue, including nonsentential "fragment."	6. *Linguistic representations* Not needed because the contextual stimuli suffice to produce linkages in dialogue.

often pressure on speakers to incorporate a high density of autoclitics according to constraints embodied in formal rules of grammar. When these pressures are weak, as in much of dialogue as well as certain literary styles (Skinner 1957, p. 356), verbal behavior is more fluid. Skinner likely would have endorsed the call in the last sentence of the target article for "a more flexible account of grammar capable of capturing linguistic constraints on linked sentence fragments."

Regarding row 3, a routine in the IA account appears to be, in the VB account, a functional verbal unit (Skinner 1957, pp. 21, 116) that has been conditioned or strengthened in a specific situation. In the IA account, repetition of "the previous speaker's utterance" appears to be important in this process (sect. 5.1 of target article). According to Skinner, "a verbal response of a given form sometimes seems to pass easily from one type of operant to another" (Skinner 1957, p. 188). Hence, a response emitted as an instance of echoic behavior may simultaneously or subsequently appear in other categories of verbal behavior (for examples, see Skinner 1957, pp. 188–89, 360–62).

Like routines in the IA account, functional verbal units in the VB account may be larger than a single word. Similar to the process of routinization in the IA account (sect. 5.2 of the target article), a process called composition in the VB account generates large verbal patterns that can come to function as units. Skinner (1957, Ch. 14, pp. 344–67) proposed that composition consists primarily of adding autoclitics (including ordering) to the raw verbal material mentioned above. "Formal evidence alone will not show whether sentences [or other large segments of verbal behavior] have been composed" (Skinner 1957, p. 346) as opposed to being emitted as units. Once a composed utterance has been reinforced several times, it may begin to function as a unit. Given the right conditions, however, a unit may break into smaller units (Skinner 1957, pp. 116–17). This seems very close to the dynamics of routines as described in the IA account.

Regarding row 4, in the VB account "[a]n important fact about verbal behavior is that speaker and listener may reside within the same skin" (Skinner 1957, p. 163). As in the IA account, in the VB account speakers monitor their own verbal behavior and edit it after, during, or even before its emission (Skinner 1957, Ch. 15, 16, pp. 369–402). Both accounts agree that there is no difference in principle between self- and other-monitoring. Both accounts also agree that when interacting with oneself, the stimuli need not be in the form of an external medium.

Regarding row 5, the techniques for strengthening one's own weak verbal behavior are in principle the same as those for strengthening another's verbal behavior (Skinner 1957, Ch. 17, pp. 403–17). These include manipulating stimuli and changing the level of editing.

Finally, regarding row 6, by focusing on contextual stimuli, the VB account provides a flexible account of grammar.

Given the correspondences between the two accounts, it may be impossible to distinguish them empirically. However, the VB account appears to require fewer terms, "and the terms created are derived from a few prior technical terms common to the whole field of human behavior" (Skinner 1957, p. 456).

Putting the interaction back into dialogue

Emanuel A. Schegloff

Department of Sociology, University of California at Los Angeles, Los Angeles, CA 90095-1551. **scheglof@soc.ucla.edu**
http://www.sscnet.ucla.edu/soc/faculty/schegloff/

Abstract: I share the authors' stance on the dialogic or interactional character of language. The authors, however, have left actual interaction out of their conception of dialogue. I sketch a number of organizations of practices of talking and understanding that supply the basic arena for talk-in-interaction. It is by reference to these that mechanisms for speech production and understanding need to be understood.

I write as a conversation analyst. I have spent nearly 40 years studying the sorts of events which Pickering & Garrod (P&G) now take to be the fundamental premise of language. I am, of course, predisposed to take the same view. Indeed I have already done so in a number of publications (Schegloff 1979; 1989; 1996a, inter alia). The arguments of the target article aside, in the choice between a fundamentally monologic, "individualist" stance and a dialogic, interactional one, there are compelling reasons for preferring the latter. For now, one line will have to do.

For most humans on the planet since the species developed "language," the overwhelmingly most common ecological niche for its use has been (1) the *turn* at talk, (2) as part of a coherent *sequence of turns,* (3) through which a *course or trajectory of action* is jointly pursued by some or all of the participants (not necessarily cooperatively, but jointly), (4) in an *episode of interaction,* (5) between two or more *persons,* (6) organized into two or more *parties,* (7) the *occasion* of interaction being composed of one or more such episodes. If that is where language as a publicly deployed resource and utility resides, it is plausible to expect that it has been designed and fashioned by its users and uses in a manner adapted to the contingencies of its "environment" – that is, by the contingencies of talk-in-interaction (of which the foregoing are but several aspects) and its virtually omnipresent bodily companions – gesture, posture, gaze deployment, facial expression, and so on. Such an expectation is not merely plausible; detailed and repeated examination of recorded episodes of naturally occurring talk-in-interaction shows it to be so – indeed, at a thoroughly *im*plausible (and yet demonstrable) level of detail. The mechanisms of production and comprehension being addressed by P&G need to be understood in this context. It is this context that is missing from P&G's treatment.

By "context" here I do not mean the ordinary characterizations of settings as domestic or public, intimate or formal, and others drawing on different genres of social and cultural diction (important as these may be). I mean the various organizations of practice that deal with the various generic organizational contingencies of interaction without which it cannot proceed in an orderly way: (1) The "turn-taking" problem: Who should talk next and when should they do so? How does this affect the construction and understanding of the turns themselves? (2) The "sequence-organizational" problem: How are successive turns formed up to be "coherent" with the prior turn (or *some* prior turn), and what is the nature of that coherence? (3) The "trouble" problem: How should one deal with trouble in speaking, hearing, and/or understanding the talk such that the interaction does not freeze in place, that intersubjectivity is maintained or restored, and that the turn and sequence and activity can progress to possible completion? (4) The word selection problem: How do the components that get selected as the elements of a turn get selected, and how does that selection inform and shape the understanding achieved by the turn's recipients? (5) The overall structural organization problem: How does the overall structural organization of an occasion of interaction get structured, what are those structures, and how does placement in the overall structure inform the construction and understanding of the talk as turns, as sequences, and so on?

The organizations of practice addressed to these issues – turn organization (Goodwin 1979; Schegloff 1996a), turn-taking organization (Jefferson 1986; Sacks et al. 1974; Schegloff 1987a; 2000a; 2001), sequence organization (Schegloff 1990; 1995; forthcoming), the organization of repair (Drew 1997; Jefferson 1974; 1987; Schegloff 1979; 1987b; 1991; 1992; 1997a; 1997b; 2000b; Schegloff et al. 1977), the organization of word selection (Sacks 1972a; 1972b; 1992; Sacks & Schegloff 1979; Schegloff 1972; 1996b), overall structural organization (Schegloff 1986; Schegloff & Sacks 1973), and others – constitute, in the options that they shape and the practices made available, a spate of interaction recognizable as "conversation," as "interview," as "meeting," as "lecturing," as "giving a speech," as "interrogation," and so on. These are what we call "speech-exchange systems" (Sacks et al. 1974,

pp. 729–31), and can be seen as particular, here-and-now-with-these-participants instances of these.

What makes an interaction is not just the juxtaposition of bodies. What mediates and organizes the conduct of the parties is not a structureless, featureless, transparent medium. The composition of a turn at talk – whether it is made up of one or more component units; whether these are sentences or sub-sentential – its syntactic construction and choice of lexicon are shaped in part by the contingencies of turn production imposed by a turn-taking organization that will have others empowered or required or allowed to talk next, at points in the turn's development not wholly under the speaker's control. Particular courses of action implemented through turns at talk (such as request sequences, complaint sequences, storytelling sequences, news-conveying sequences, etc.) implicate certain ways of understanding what is being said that render meaningful and consequential selection between apparently equivalent expressions, the delay of a turn's start by two-tenths of a second or less, and the like. How one says what one says can depend on who the other is; and, of all the persons and categories which could be used to characterize "the other," depend on which ones have been made relevant at that moment in the talk, or can be made relevant by constructing the same "sayable" in this way or that. And so on.

A very high proportion of the matters discussed by P&G as if they were unrelated to anything but the mechanisms the authors are concerned to develop, are not interactionally random. They are part of the fabric of some organization of practices for talk-in-interaction. Many of them have been given quite detailed and systematic treatment in the literature – things like "routines" (target article, sect. 5.2.1, cf. Schegloff 1986) and "how are you" routines in particular (Jefferson 1980; Sacks 1975); things like "joint constructions" (sect. 7.1, para. 3; cf. Lerner 1991; 1996; Sacks 1992, vol. I, pp. 144–47 et passim); things like "non-sentential turns" (sect. 7.1, para. 6; cf. Sacks et al. 1974; Schegloff 1996a): things like "monitoring during overlapping speech (sect. 6, para. 6; cf. Schegloff 2000a; 2001); and so on and on.

Most striking is P&G's treatment of "repair"; the discussion rests on a terminology ("repair," "other-repair," "self-repair") which they neither explicate nor cite but the latter two of which they treat as discrete sets of things, not an organization of practices. This leads them – incorrectly, in my view – to treat the basic mechanisms of self-repair and other-repair as the same (see Table 2 of the target article) when, interactionally speaking, they are not the same in either execution or interactional import (Schegloff 1979, pp. 267–69; Schegloff et al. 1977, inter alia). I believe the analysis of talk-in-interaction along such lines has much to contribute not only to our understanding of the mechanisms addressed by P&G, but to work in the neurobiology of behavior more generally – precisely the remit of this journal. But that is another matter.

Some notes on priming, alignment, and self-monitoring

Niels O. Schiller[a,b] and Jan Peter de Ruiter[b]

[a]*Department of Cognitive Neuroscience, Faculty of Psychology, Maastricht University, 6200 MD Maastricht, The Netherlands;* [b]*Max Planck Institute for Psycholinguistics, 6500 AH Nijmegen, The Netherlands.*

n.schiller@psychology.unimaas.nl Niels.Schiller@mpi.nl
JanPeter.deRuiter@mpi.nl
http://www.psychology.unimaas.nl/Base/Personal/NielsSchiller.htm
http://www.mpi.nl/world/persons/profession/schiller.html

Abstract: Any complete theory of speaking must take the dialogical function of language use into account. Pickering & Garrod (P&G) make some progress on this point. However, we question whether their interactive alignment model is the optimal approach. In this commentary, we specifically criticize (1) their notion of *alignment* being implemented through *priming,* and (2) their claim that *self-monitoring* can occur at all levels of linguistic representation.

The primary way of language use is *dialogue,* not monologue. We want to acknowledge the authors' effort to stress this important point, which needs to be addressed explicitly in empirical and modeling work in speech production and comprehension research. We believe that these issues are especially relevant for syntactic processing. For instance, one wonders how syntactically incomplete (dialogue) utterances can be syntactically encoded in more traditional models, if there is no overt verb present in the generated utterance. Take, for example, the following extract from the dialogue transcript in section 2 of the target article:

1——**B:** . . . Tell me where you are?
[Utterances 2 and 3 omitted]
4——**A:** Right: {*I am*} **two along from the bottom one up:*** [our addition in curly brackets]

In this example, speaker A does not produce the appropriate verb form of "to be" (i.e., "I am") but nevertheless gives an acceptable and cooperative answer to speaker B's question. This type of ellipsis can only be correctly produced if the syntax generator has access to previously stored discourse information, allowing the speaker to omit "I am," even though the original question containing the verb occurred several utterances earlier in the discourse (see also Levelt 1989, p. 89, for a similar analysis).

Although we agree in principle with the authors' assessment that the dialogical structure of language should receive more attention in accounts of language processing, we are not convinced that adopting the interactive alignment model is the right way to do so. For instance, it is unclear to us exactly how *priming* can account for *alignment,* and, in particular, we fail to see in what way priming is more than "a behavioral effect" (see target article, sect. 2.2). We believe that "priming" *does not explain or implement interactive alignment.* Real interactive alignment necessarily involves *storing selected fragments* from previous utterances. Priming can raise the probability of certain linguistic structures being selected, but this is not sufficient for the strong and explicit type of alignment the authors want to incorporate in models of language processing. Also, syntactic priming effects are weak effects. It is hard to see how an elaborate mechanism such as interactive alignment could be realized by only raising the probability of selecting a certain syntactic construct by roughly 10% to 20% (see, e.g., Pickering & Branigan 1998).

Our second critical note concerns one of the few testable predictions from the interactive alignment model, namely, that self-monitoring by the speaker occurs at all levels of linguistic representation (see sect. 6). While other researchers (e.g., Wheeldon & Levelt 1995) have claimed that internal self-monitoring works on abstract phonological form representations, Pickering & Garrod (P&G) propose that self-monitoring can occur at any level of linguistic representation that can be aligned (i.e., semantic, syntactic, lexical, phonological, and phonetic representations) – and not only at the phonological level.

For example, the authors explicitly claim that speakers can correct gender errors, such as *le tête* instead of *la tête* ("the head") in French or *de been* instead of *het been* ("the leg") in Dutch not only after they have been articulated but even before their overt production. This is an interesting claim that needs to be investigated in the future. However, we are somewhat skeptical about this claim because to our knowledge there is no evidence that self-monitoring of gender features (or any other syntactic features) is possible. For example, Desrochers and his collaborators (Desrochers & Paivio 1990; Desrochers et al. 1989; Muller-Gass et al. 2000) found that selecting a gender label (e.g., feminine or masculine) took about 200 msec longer than selecting the indefinite article in French gender decision. Furthermore, Tucker et al. (1977) provided empirical evidence suggesting that French speakers implicitly construct a noun phrase including the article and the noun to determine a noun's gender. However, if speakers can self-monitor abstract gender information at the level of syntactic representation, as suggested by P&G, why would they go through the trouble of generating the gender-marked article of a noun to determine its gender?

In contrast to these findings about syntactic representations, recent evidence from our own laboratory as well as from other laboratories demonstrated that self-monitoring *does* occur at the level of phonological encoding. We have empirical data about the monitoring of phonological segments (Schiller, in press; Wheeldon & Morgan 2002), word stress (Schiller 2001; Schiller et al., in press), syllable boundaries (Jansma & Schiller 2004), and syllables (Morgan & Wheeldon 2003). However, we also have evidence that participants are unlikely to monitor a phonetic-acoustic representation of the respective utterances. Although gender decision as a task is widely used (Müller & Hagoort 2001; Schiller et al. 2003; Schmitt et al. 2001a; 2001b; Van Turennout et al. 1998), it remains to be shown whether or not this task actually taps syntactic processing, because abstract gender information may not be directly available to the speaker. Rather, gender information may be available only via its phonological realization, for example, an article or gender-marking suffix. Interestingly, effects of *gender congruency* have recently been re-interpreted as *determiner congruency* effects occurring at the phonological level and not at the gender feature level (see overview in Schiller & Caramazza 2003).

To conclude, we believe that internal self-monitoring does not occur at every single level of linguistic representation (as claimed by P&G) or at every processing level in models of speech production (e.g., Levelt et al. 1999). Rather, there is abundant evidence that internal self-monitoring works on phonological representations, which are created during phonological encoding in speaking, for example, when segments are prosodified into phonological words. It is at this level that information about segments, syllables, syllable boundaries, and word stress is available to the speaker. Although the abundance of evidence for phonological-level monitoring does not necessarily exclude other monitoring levels (e.g., at the conceptual level; see Levelt 1989), we are unaware of any evidence for self-monitoring at earlier or later levels of grammatical and form encoding.

ACKNOWLEDGMENTS

Niels O. Schiller is supported by a grant from the Netherlands Organization for Scientific Research (NWO). Jan Peter de Ruiter is supported by the European Union Project Conversational Multi-modal Interaction with Computers (COMIC), grant IST-2001-32311. The authors thank Pim Levelt for his comments on an earlier draft.

Just how aligned are interlocutors' representations?

Michael F. Schober

Department of Psychology, F330, New School for Social Research, New York, NY 10003. **schober@newschool.edu**
http://homepage.newschool.edu/~schober

Abstract: Conversational partners' representations may be less aligned than they appear even when interlocutors believe they have successfully understood each other, as data from a series of experiments on surveys about facts and behaviors suggest. Although the goal of a mechanistic psychology of dialogue is laudable, the ultimate model is likely to require far greater specification of individual and contextual variability.

When conversational partners come to agree that they are talking about the same thing, how aligned are their conceptualizations? The interactive alignment account holds that when dialogue is successful, interlocutors' linguistic representations are aligned at multiple levels. Although Pickering & Garrod (P&G) observe that alignment is sometimes only approximate and that evident misalignments can be interactively repaired, the general thrust of their approach is that successful communication requires representations to be the same.

I contend that interlocutors' using the same words can actually mask a surprising degree of undetected misalignment. Fred Conrad and I, with other colleagues, have carried out a series of laboratory and field studies examining how people interpret ordinary words in ongoing U.S. government surveys about facts and behaviors, words like "job," "bedroom," "smoking," and "cigarettes" (Conrad & Schober 2000; Schober & Conrad 1997; Schober et al. 2004; Suessbrick et al. 2000). Because the agencies that carry out these surveys have thorough definitions for the terms, answers to the survey questions provide evidence about the extent to which respondents' conceptions match the survey designers'. Our finding is that people's representations are frequently wildly misaligned with the survey designers' – and with each other's – without anyone's noticing.

For example, in one study (Suessbrick et al. 2000), survey respondents interpreted terms like "smoking" and "cigarettes" in a question such as "have you smoked at least one hundred cigarettes in your entire life?" differently enough (tobacco, cloves, marijuana? Finished or just a puff? Bought or borrowed?) that 10% of the respondents subsequently presented with a definition changed their answer to the question from yes to no or from no to yes. In a national telephone sample (Conrad & Schober 2000), more than 40% of reported purchases did not fit the survey designers' definitions, even though the questions had been widely pretested. And this is not just because the official definitions failed to match the population consensus about the meaning of terms; respondents' interpretations differed from each other's as much as they differed from the survey designers'.

Across our various studies, respondents are quite surprised at the thought that someone else might interpret the same words differently from the way they do; when given the opportunity to request clarification about the meanings of survey terms, they choose to do it a very small percentage of the time. People seem to follow a "presumption of interpretability" (Clark & Schober 1991): It should be the questioner's responsibility to forestall misinterpretation.

These data suggest a far more Quinian view of successful referring than the P&G account encompasses: Seemingly successful referring can mask conceptual misalignments that reflect deep underlying indeterminacies. The point is that people can believe they have understood each other well enough for current purposes (as proposed in Clark & Wilkes-Gibbs 1986) and yet never actually discover that their conceptions were misaligned.

An important contention in the P&G article is that seemingly complex interactional processes can be modeled largely with simple individual mechanistic processes. The proposal is that conversational partners, following a principle of parsimony, do not ordinarily model each other's mental states or make inferences about common ground, except when there is evidence that not doing so has led to obvious misunderstanding and when cognitive resources allow.

I would argue that none of the current data actually allow us to distinguish this position from an alternative: that the ordinary case is that conversational partners *do* model each other, and that they fail to do so only when they are under heavy cognitive load or when circumstances weigh heavily against doing so. Why should we assume that the ordinary case is one where the interlocutor does not need to be modeled and the speaker is under heavy cognitive load? As far as I can tell, no one knows the level of load encountered in the range of ordinary interactive situations. As Susan Brennan and I have argued (Schober & Brennan 2003), the evidence for egocentric processing is far from conclusive; the experiments purported to show egocentric processing as basic rely on null results and experimental methods that are far removed from ordinary processing situations. When such studies are carried out in more realistic settings, the findings can look rather less egocentric.

Not to overstate the case, but one could argue that modeling one's partner only when it is needed may require a level of situational monitoring that leads to a paradox: How can one know exactly when one needs to model one's partner without already

knowing what the partner model would clarify? (This seems akin to Fodor's frame problem or the notorious conundrums about selective attentional filtering). Arguably, it might be simpler to model the partner as the ordinary case. P&G assume that representing common ground requires unusually complex mental representations; this may be so, but it is also possible that representations of common ground are ordinary memory representations with the right content.

In fact, what is parsimonious may well vary across individuals and conversational contexts. Our notions of what is effortful in dialogue are not well worked out (Schober 1995; 1998a); people's knowledge of each other's perspectives, interest in modeling their partners, and abilities for doing so probably vary substantially more than current theories allow. For example, people who have very poor mental rotation abilities may be unable to conceive of a partner's point of view on a scene, leading to a different sort of partner modeling than that carried out by mental rotation whizzes (Schober 1998b). And different abilities may be relevant in different conversational contexts; mental rotation ability may have little to do with the consideration of lexical alternatives in dialogues that are not about physical scenes, whereas working memory capacity differences or different conversational agendas (Russell & Schober 1999) might. P&G are to be applauded for furthering the attempts to bridge the "language as product" and "language as action" traditions, but they have also opened a can of worms; the ultimate mechanistic account of dialogue is likely to require far greater specification of individual and contextual variability.

Dialogue processing: Automatic alignment or controlled understanding?

Hadas Shintel and Howard C. Nusbaum

Department of Psychology, The University of Chicago, Chicago, IL 60637.
hadas@uchicago.edu hcn@speech.spc.uchicago.edu

Abstract: Pickering & Garrod's (P&G's) mechanistic account of dialogue assumes that linguistic alignment between interlocutors takes place automatically, without using cognitive resources. However, even the most basic processes of speech perception depend on resource use. The lack of invariant mapping between input patterns and interpretations in dialogue, as in speech perception, may require controlled, rather than automatic, processing.

In their target article, Pickering & Garrod (P&G) challenge current psycholinguistic theory by proposing the study of dialogue as the appropriate paradigm for understanding language processing. They argue that the main problem facing interlocutors is the alignment of their respective situation models, and they propose a mechanistic account of alignment in dialogue, based on an interactive and resource-free priming mechanism. Linguistic structures, produced on the surface by one interlocutor, putatively prime corresponding structures in the other interlocutor. While we agree that conversation, rather than decontextualized, isolated sentences, should be taken as basic form of language use and as the foundation upon which language has evolved, it seems less plausible that much of the cognitive work in conversation takes place automatically.

P&G's model of automatic alignment assumes that a mapping can be established directly between the representations uttered by one interlocutor and those activated in the head of the other interlocutor. But as they themselves note, the linguistic information is encoded in sound. Hence, for the listener the spoken utterance must be recognized antecedent to any alignment occurring. However, even findings regarding the basic processes of speech perception argue against the assumption of automatic processing. Automaticity implies a passive process in which the input is processed in an invariant, inflexible manner, regardless of the beliefs and expectations of the listener (Nusbaum & Schwab 1986). For a process to be automatized, there must be a consistent mapping between input patterns and responses, the benefit of which is a process that does not impose demands on cognitive and attentional resources (Shiffrin & Schneider 1977).

Of course the hallmark of speech is the lack of consistent, invariant mapping between acoustic patterns and linguistic categories across phonetic contexts (Liberman et al. 1967) and across talkers (Peterson & Barney 1952). As a result of variable mapping, speech perception depends on controlled, active processing requiring attention and working memory (Nusbaum & Magnuson 1997). For example, variation between talkers (as in the circumstance of a conversation among three interlocutors) requires talker normalization, the process by which listeners compensate for acoustic-phonetic variability in vocal characteristics (see, e.g., Nearey 1989). Nusbaum and Morin (1992) found that talker variability slowed speech recognition (see also Mullennix & Pisoni 1990), and this slowing was because of increased demand on working memory.

Furthermore, talker normalization is contingent on listeners' expectations regarding the interpretation of acoustic patterns – which should not happen for an automatic direct-mapping process. Magnuson and Nusbaum (1994) demonstrated that when listeners expected a pitch difference to signal a talker difference, they showed talker normalization, but if the same pitch difference was expected as a way one talker accented speech, no normalization occurred. Similarly, the expectation that a talker was male or female significantly changed the interpretation of vowel tokens (Johnson et al. 1999). The acoustic patterns of speech are processed differently depending on listeners' expectations, arguing against invariant automatic processing. Expectation effects suggest that alignment may not be a result of a direct, automatic, causal link between the activation of a representation in one interlocutor and the activation of an objectively matching representation in the other interlocutor. Rather, the process is mediated by listeners' expectations. For example, alignment at the level of articulation, where interlocutors converge on each other's speaking style, may be mediated by interlocutors' stereotypic expectations about the other interlocutor's accent and speech rate, resulting in subjective but not objective alignment (Thakerar et al. 1982).

Listeners may need to use controlled active processing because the one-to-many mapping (one pattern may have multiple interpretations) in speech represents a nondeterministic computational problem that cannot be solved, in principle, by a deterministic system such as an automatic process (cf. Nusbaum & Magnuson 1997). This problem of one-to-many relationships between linguistic patterns and interpretations occurs across all levels of linguistic analysis (Nusbaum & Henly 1989). Indeed, research has shown that the same spoken sentence can be differently interpreted in different visual referential contexts (Tanenhaus et al. 1995) and that the same indirect requests can be processed differently as a result of speaker status (Holtgraves 1994). Although P&G focused on utterance-level matches that can occur within stretches of dialogue, real conversation is less predictable and routinized – otherwise communication would be unnecessary.

This raises several deep questions facing dialogue research: How do interlocutors cope with variability in linguistic and perceptual context? How are diverse sources of information integrated to constrain production and comprehension? And how do interlocutors flexibly adapt to different conversational circumstances? Although in real conversations these questions are often subjectively resolved quickly and without apparent effort, this is seldom an accurate barometer of the demands on cognitive resources such as attention and working memory.

P&G propose that the process of automatic alignment bypasses the need for modeling the interlocutor's mental states and for distinct, conscious decision stages. We agree that a constant monitoring of common ground would be unnecessary and costly, but the dichotomy between automatic processes and conscious decision processes involving complex inferences about the interlocutor's mental state does not represent the full range of possible pro-

cessing alternatives. In clarifying this point, it is worth considering what is meant by automatic, as distinct from controlled, processing. Automatic processes have been defined by different criteria such as being unintentional, occurring outside of conscious awareness, not requiring cognitive resources, and being autonomous; but these criteria do not necessarily hold simultaneously (Bargh 1989). Speech perception involves controlled active processes that are not resource-free or autonomous, yet they do occur largely outside of awareness, do not require a conscious intention on the part of the listener, and are subjectively experienced as effortless. Likewise, the subjective ease and speed that subjectively characterize language use in dialogue may not reflect the complexity of the underlying processing. The variability and flexibility shown in the processes of language comprehension and production in dialogue call for dynamic adaptation rather than a passive automatic mechanism.

Top-down influences in the interactive alignment model: The power of the situation model

Tessa Warren[a] and Keith Rayner[b]

[a]*607 LRDC (Learning, Research, and Development Center) and Department of Psychology, University of Pittsburgh, Pittsburgh, PA 15260;* [b]*Department of Psychology, University of Massachusetts, Amherst, Amherst, MA 01003.*
tessa@pitt.edu rayner@psych.umass.edu

Abstract: Pickering & Garrod's (P&G's) model is an innovative and important step in the study of naturalistic language. However, the simplicity of its mechanisms for dialogue coordination may be overstated and the hypothesized direct priming channel between interlocutors' situation models is questionable. A complete specification of the model will require more investigation of the role of top-down inhibition among representations.

Pickering & Garrod's (P&G's) new model of linguistic interaction in dialogue is an important contribution to the study of psycholinguistics. This model breaks new ground between two traditionally disparate areas of language research and combines the mechanistic detail characteristic of sentence processing research with the emphasis on language as a cooperative process characteristic of dialogue research.

The central mechanism of P&G's interactive alignment model is a process they term "alignment," whereby dialogue participants' linguistic and discourse representations become more similar over the course of an interaction. Alignment is an automatic process that results from priming between linguistic representations. Priming can occur among different levels of representation within a single individual, as when increased lexical or semantic overlap between a prime and a target causes increased syntactic priming, or between the representations of different individuals, as when a speaker who has just comprehended a particular syntactic form produces the same form in a subsequent utterance. The resource-free characterization of the system relies on the assumption that alignment at one level of representation increases alignment at other levels of representation, and therefore essentially complete alignment can be obtained through simple and automatic processes.

Though the strength of this model is in its elegance and simplicity, it is not clear whether these characteristics will survive a more explicit specification of the system. According to the characterization of the model as a network of linked representations where priming and alignment in one representation causes increased alignment in all others, alignment at the level of the situation model is simply an epiphenomenon of alignment at lower levels. Yet P&G acknowledge that in some cases priming at one level will decrease alignment at other levels and specifically note that alignment of the situation model takes priority over alignment at other levels of representation. This complicates the system, as lower-level priming is constrained by the very alignment it drives. In fact, lower-level priming, supposedly a driving force in alignment, can be inhibited by extremely subtle nuances of local context. For example, P&G discuss Garrod and Anderson's (1987) findings that participants used different words for horizontal groupings of boxes depending on whether the grouping was modified with an ordinal adjective or not. So participants who used the phrase "the second row" later spoke of "the bottom line," even though "the bottom row" was an acceptable, unambiguous, and more lexically primed alternative. From examples like this it is clear that priming and alignment at the level of the situation model are very powerful mechanisms in the system, and that they have significant inhibitory power over the automatic, lower-level priming that is hypothesized to drive alignment at all levels of the system. It remains to be seen whether bottom-up mechanisms are robust enough to drive alignment, as P&G claim, or whether top-down inhibition directs the system.

The interactive alignment model relies heavily on the assumption that the situation model of one individual can directly prime the situation model of another individual through the same automatic mechanisms that are responsible for phonological, lexical, semantic, and syntactic priming between individuals. This assumption seems hasty. Phonological, lexical, and syntactic priming are similar in that the representation that will eventually be primed is an inalienable part of the structure of the message that causes the priming. The words, sounds, and structural patterns that make up an input string necessarily activate representations for exactly those words, sounds, and patterns. Semantic priming is not as direct, as semantic representations cannot be directly read off an input string. However, semantic representations are generally similar across individuals. For example, it is safe to assume that the word "dog" will activate the concept "cat" more strongly than the concept "book" for the wide majority of individuals. This similarity of representation may allow for what P&G represent as a direct priming link between the semantic representations of different individuals. If an individual produces the word "goat," the concepts that will be primed in her own semantic representation are likely similar to the concepts that will be primed in her interlocutor's representation, because the representations are structured in a similar way. In this indirect way, it could be said that the semantic representation of one individual can prime the representation of another.

The same thing cannot be said of situation models. Like semantic priming and unlike phonological, lexical, and syntactic priming, there is no direct priming channel between individuals through physical aspects of a message such as sounds, words, or word patterns. Again, priming must be indirect, through the activation of words or phrases that suggest a particular state or property of a situation model. But unlike semantic representations, individuals do not necessarily begin dialogue with similar situation representations. Therefore the priming link between individuals' situation models in the interactive alignment model must be of a different sort from the priming channels between other representations.

The interactive alignment model opens a new and exciting area of inquiry into language processing. However, more research into the details of situation model priming and the complex interplay of priming and inhibition between different levels of representation will be necessary in order to fully specify the operation of the model and to evaluate its ascribed simplicity. Carrying out this research will not be easy, as it will be difficult to maintain the necessary experimental control in the sorts of experiments that will be required. But as P&G optimistically point out, "Well-controlled studies . . . may require some ingenuity, but such experimental ingenuity has always been a strength of psychology" (target article, sect. 1, para. 4).

Authors' Response

The interactive-alignment model: Developments and refinements

Martin J. Pickering[a] and Simon Garrod[b]

[a]*Department of Psychology, University of Edinburgh, Edinburgh EH8 9JZ, United Kingdom;* [b]*Department of Psychology, University of Glasgow, Glasgow G12 8QT, United Kingdom.* **Martin.Pickering@ed.ac.uk simon@psy.gla.ac.uk**

Abstract: The interactive-alignment model of dialogue provides an account of dialogue at the level of explanation normally associated with cognitive psychology. We develop our claim that interlocutors align their mental models via priming at many levels of linguistic representation, explicate our notion of automaticity, defend the minimal role of "other modeling," and discuss the relationship between monologue and dialogue. The account can be applied to social and developmental psychology, and would benefit from computational modeling.

The target article set out to show how it would be possible to develop a theory of interactive language processing at the level of explanation normally associated with cognitive psychology. In our theory, successful communication involves the alignment of interlocutors' representations. We proposed that each level of representation becomes aligned via an automatic process that we treat as a form of priming, and that alignment at one level automatically strengthens alignment at other levels. The role of conscious or deliberate strategies involving explicit reasoning about the mental states of one's interlocutor is comparatively small in our account.

Our commentators have raised a number of insightful points that have caused us to refine our proposals. Many commentators have focused on the nature of the alignment process. At a basic level, they consider whether alignment is the primary mechanism leading to conversational success, to what extent it is automatic, and whether it can be explained by a single mechanism at all levels and in all contexts. Commentators have also questioned our downplaying of "other modeling" in ordinary conversation and our claims about the nature of the difference between monologue and dialogue. In responding to these and other comments, we have divided our reply into eleven sections whose order roughly follows that of the topics raised in our target article.

R1. To what extent do interlocutors align?

Perhaps the most basic issue about our model is whether interlocutors actually align their situation models, or, less dramatically, whether they align to the extent that we claim they do. **Schober** proposes that interlocutors may be much less aligned than they appear even when they believe that they have understood each other. Of course, this would not matter if it solely concerned rare cases of genuine misunderstanding (e.g., when two interlocutors refer to different people called *John*); see also **Branigan**, who points out that communication may be "successful" in some sense even when there is some misunderstanding. But Schober argues that misalignment is endemic to dialogue. His comments relate particularly to the interpretation of referring expressions with respect to the discourse model. He draws on examples from surveys where respondents interpret terms in ways that are very different from those intended by the survey compositors. Our response is that such surveys do not constitute dialogue: The compositors construct the survey, and the respondents then respond. There is no feedback, no possibility for repair, and hence no *interactive* alignment. Schober also raises the important point that people need not necessarily fully interpret expressions (Clark & Wilkes-Gibbs 1986). In fact, full interpretation probably does not always occur in the comprehension of monologue (Barton & Sanford 1993; Frazier & Rayner 1990; Frisson & Pickering 1999; 2001; cf. Sanford & Sturt 2002), with people often not determining the precise sense of referring expressions (e.g., *newspaper* meaning an object vs. a day's edition), and there is no reason to assume that dialogue is any different. We suspect that both producers and comprehenders determine meaning to the extent necessary for current purposes, and that one way in which interlocutors align is by each processing referring expressions to equal depth.

R2. What precisely are they aligning?

Several commentators appear concerned with the question of what exactly is being aligned within our model. At the "lower levels" of phonology, syntax, the lexicon, and so on, interlocutors presumably align the representational content of each of those levels (phonemes, syntactic structures, lexical items, etc.), but it is perhaps less clear what they align at the level of the situation model. In the target article, our intention was to argue for alignment of *structural* aspects of the situation model, as exemplified by our example of reference frames. Some of our commentators assume that we are referring to the *content* of the situation model. The questions about alignment of content are much more difficult, and we shall try to explain the issues below.

In our account, interlocutors align on representations relevant to the dialogue. These include lexical, semantic, and syntactic representations, but also the situation model. So if, at a given point in a conversation, one interlocutor has a situation model containing two individuals, Mary and John, with Mary in focus, with each at different locations, and so on, then the conversation will be successful to the extent that the other interlocutor constructs the same situation model. Of course, one interlocutor can now introduce another character (or a new relation between the existing characters) – indeed, introducing new information is central to any conversation that is not entirely repetitive. To do this, the speaker draws upon his knowledge (typically using long-term memory) and adds information to his situation model. The effect of the alignment is that the listener updates his model so that it remains similar to that of the speaker. For example, the listener will interpret ambiguous words and utterances in the way that the speaker has employed them.

A much bolder claim is that the choice of new topics is affected by alignment. We did not make this claim in the target article, although we believe that it is true to some extent. For example, if one interlocutor refers to *the couch*, then the other is more likely to refer to *the couch* as well (Brennan & Clark 1996; Garrod & Anderson 1987). As a result of this, the use of *couch* presumably activates knowledge about couches, and hence makes it more likely that the interlocutor will talk about couches rather than some

other topic. To this extent, alignment is surely unsurprising (and simply amounts to the claim that interlocutors will persist with particular topics).

It may also be that interlocutors align on particular styles of reasoning or accessing of knowledge. For instance, if one interlocutor is engaged in a careful search of long-term memory, then the other will tend to behave similarly (e.g., if you play a general-knowledge game seriously, then I am likely to do so too). Alignment on style of reasoning is relevant to the construction of the situation model (cf. Gentner & Markman 1997), but takes us beyond the scope of the target article, just as nonlinguistic imitation more generally does (e.g., Chartrand & Bargh 1999). For now, our goals are limited to understanding linguistic factors that assist in the alignment of situation models.

Cutting questions our characterization of interactive alignment as reflecting a direct link between interlocutors. He suggests that it only has an indirect effect on the language processes themselves. To answer this comment we need to clarify how interactive alignment relates to language processing. As we have said, our contention is that interactive alignment (and in particular the automatic alignment channels) affects the structures used in production and interpretation rather than directly determining the *content* of production and interpretation. In other words, we assume that alignment provides an explanation of the manner in which interlocutors produce and interpret contributions. So we propose that alignment channels only produce a direct link between the structures that the interlocutors use in language processing. Hence, the alignment process is automatic and direct, even though it does not determine exactly what the speaker produces (as this depends on his long-term knowledge) or how the addressee interprets what he hears "beyond" the level of the situation model.

Other commentators also query whether we have specified the appropriate mechanism for alignment of situation models. They raise this concern in relation to two more specific topics: whether there is one alignment mechanism or several (**Branigan**, **Glucksberg**, and **Markman**, **Kim**, **Larkey**, **Narvaez**, **& Stilwell** [**Markman et al**.]), and whether (or in what sense) alignment is automatic (**Krauss & Pardo**). All of these commentaries, in some sense, are concerned with the issue of how alignment could affect the content of situation models. We address these in the following two sections.

R3. The mechanisms of alignment

Several commentators question the details of the interactive alignment mechanism itself and point out that we have not fully specified a mechanistic account. **Brown-Schmidt & Tanenhaus** make a general plea for modeling, which we fully agree with (see sect. R11). **Goldinger & Azuma** argue that we do not give a detailed characterization of the process by which alignment comes about. We have no commitment to interactive-activation models and are open to the suggestion that Grossberg's (1980) adaptive-resonance theory may provide an appropriate framework for the interactive-alignment account.[1]

Beyond this, two somewhat different issues are raised. Some commentators argue that we assume alignment is based on transient activation, and they propose instead that it is based on facilitated memory retrieval or implicit learning. Others claim that we are wrong to assume a unified account for all levels of alignment.

Kaschak & Glenberg argue that alignment is not due to priming but rather to a facilitated memory retrieval mechanism. In response, we note that the interactive-alignment model is specified at a functional level and makes no commitment to specific mechanisms, and that we use the term "priming" to refer to both transient activation and facilitation in memory-based accounts. Our model attempts to capture the way in which representations used for both production and comprehension automatically become aligned as a consequence of the process of interaction. These representations may be subject to transient activation or, instead, there may be enhancement of the mechanisms underlying their retrieval from memory (as envisaged by Kashak & Glenberg).

Perhaps more likely, there may be two separate mechanisms involved in alignment. For example, some recent accounts of syntactic priming are based on implicit learning (Bock & Griffin 2000; Chang et al. 2000), whereas some are based on activation of grammatical nodes (Hartsuiker et al. 2004; Pickering & Branigan 1998). Some experimental research finds clear evidence for long-term priming that is largely unaffected by intervening material (Bock & Griffin 2000; Hartsuiker & Westenberg 2000), whereas others shows rapid decay (Branigan et al. 1999; Levelt & Kelter 1982; Wheeldon & Smith 2003). Most likely, different tasks and sentence types lead to very different time-courses of priming. Although most of this work does not involve dialogue (except Levelt & Kelter 1982), under our account we would expect similar patterns of results to occur in dialogue. We therefore suggest that transient activation explains some aspects of alignment, and memory-based mechanisms explain other aspects of alignment. In section R9 below, we suggest that alignment due to routinization is likely to involve the establishment of memory traces for semi-fixed expressions.

Schiller & de Ruiter argue that interactive alignment involves storing and re-using selected fragments from previous utterances (see sect. R9); this constitutes a specific version of a memory-based account. However, their argument is based on the claim that priming is insufficient to account for interactive alignment because syntactic priming effects are too weak. In fact, the 10–20% effects that they refer to, occur in monologue. In dialogue, our studies have shown 55% priming effects when the verb is repeated (Branigan et al. 2000) and up to 47% with a rare structure when the noun is repeated (Cleland & Pickering 2003). Likewise, lexical entrainment almost always occurs for ambiguous words (Brennan & Clark 1996; Garrod & Anderson 1987). In our model, percolation effects between levels also increase the degree of alignment, and extended dialogue iteratively reinforces alignment.

A number of commentators question whether alignment operates in the same way at all levels in our model. **Markman et al**. argue that there are different requirements on alignment at the different levels. In particular, they separate the situation model from lower levels of linguistic representation. We agree that the structural alignment process they identify may well be appropriate at the level of the situation model, because models reflect complex higher order relations between elements (see sect 2.2 of the target article). However, we disagree with their argument that, unlike

lower level representations, situation models have to be partially misaligned either to ensure that given-new conventions are followed, or for the maintenance of common ground. We propose that these requirements can be fulfilled through the implicit common ground which does not differentiate between the speaker's and listener's situation models.

Branigan also separates the situation model from other levels, but for reasons that differ from those of **Markman et al**. In our terms, she accepts channels of alignment at syntactic, lexical, and morpho-phonological levels but not at the level of the situation model, because she believes that utterances do not provide direct evidence about the situation model. She claims that *I am in row two* provides direct evidence about lower levels, whereas the listener has to interpret the utterance (presumably, by using background knowledge) in order to construct the situation model. We disagree with this, because all levels of analysis require a combination of top-down and bottom-up information. For example, resolving phonemes, ambiguous words, or syntactically ambiguous utterances requires the use of context. It is therefore wrong to assume that only the level of the situation model is "abstract." We therefore see no reason to assume that channels of alignment are used only at lower levels, nor do we see any reason to alter our assumption that alignment at lower levels leads to alignment at the level of the situation model.

Warren & Rayner argue that the priming link between individuals' situation models must be different from that for lower levels. This is because interlocutors do not necessarily begin dialogues with similar situation representations and so alignment has to be built up over a period of interaction. Again, we see no fundamental difference between situation models and lower levels in this respect – alignment at all levels is built up, though the rate of alignment may differ at different levels. Additionally, Warren & Rayner question how conflicts in alignment at different levels are resolved (e.g., when aligning on the same name, "John" might produce a semantic misalignment in contexts where there are two Johns present). In fact, the issue was briefly discussed in the target article where we argued that alignment at the level of the situation model would override alignment at lower levels (target article, endnote 4). Adopting a particular situation model will influence the way a speaker frames almost everything he says, whereas adopting a particular word or syntactic structure will only affect the subsequent choice of that word in preference to another or influence the subsequent use of that particular syntactic structure. Because the situation model is so pervasive, it will be constantly reinforced in implicit common ground, and misalignment at this level will be more likely to trigger interactive repair. This suggests that the timecourse of priming at the level of the situation model may be long-lasting, whereas priming at low levels, such as phonology, may be much more short-lived. Priming at the syntactic level might be intermediate in duration, or depend more on its precise context (as suggested above). It therefore might be the case that priming of the situation model depends primarily on memory representations, whereas priming at low levels might depend primarily on transient activation. All of this, however, requires detailed modeling.

We also believe that routinization plays an important role in reinforcing the links between lower and higher levels of representation. We take this up in section R9.

R4. What it means for interactive alignment to be an automatic process

The commentators raise two important issues about automaticity that require discussion. **Krauss & Pardo** argue against the idea that alignment can be accounted for in terms of automatic priming between interlocutors. **Shintel & Nusbaum** argue that speech comprehension processes may be far from automatic in dialogue. To answer these concerns we need to first explicate our notion of automaticity, and, second, indicate what we assume to be automatic.

Our notion of automaticity is derived from the perspective of perception-action relationships (e.g., Hommel et al. 2001) and, more particularly, social cognition and social cognitive neuroscience (e.g., Dijksterhuis & Bargh 2001; Hurley & Chater, in press). Just as Dijksterhuis and Bargh argue for an automatic perception-behavior expressway, we propose that the alignment channels are automatic (see sect. 3.2) – they operate without any intermediary decision process. Hence, the alignment process is automatic. To be more explicit, we propose that the automaticity of alignment may take place at what Bargh (1989) calls the postconscious level, whereby automaticity requires awareness of the stimulus when it originally occurred. This means that interlocutors have to attend to what the other is saying in order for automatic alignment to occur. Dijksterhuis and Bargh (2001, p. 29) also argue that automatic social influences can be inhibited when they conflict with current goals and purposes. We suggest that the same is true for interactive alignment (see Garrod & Pickering 2004). For example, if a maze game player wants to try a new description scheme because he has failed to understand the last description from his interlocutor (see sect. 2.1 of the target article), then this high level goal of introducing a new scheme may inhibit low level alignment arising from what his interlocutor has just said. However, in a similar vein to Dijksterhuis and Bargh, we predict that overriding alignment is going to be more difficult (or effortful) than adopting alignment. Additionally, this postconscious notion of automaticity can explain why alignment is affected by partner-specific factors (e.g., Branigan et al., submitted; Metzing & Brennan 2003), without invoking additional mechanisms such as "other modeling." It is also presumably relevant to many of the factors that affect the extent of speech accommodation (Giles et al. 1992). In general, we expect that rate of alignment may be affected by social factors even when the interlocutors are unaware that they are aligning. There is evidence for such alignment outside language (Epley & Gilovich 1999; Lakin & Chartrand 2003), and we expect it also to occur in language.

Krauss & Pardo agree with our claim that communication entails the alignment of situation models, but suggest that it does not principally take place via automatic priming. For example, they point to evidence that speakers accommodate to their listeners. This presents no problem according to the above conception of automaticity, which allows inhibition or facilitation by social factors. **Glucksberg** raises an interesting case, involving a difficult dialogue with a non-cooperative teenage son, in which degree of alignment may be reduced.

Shintel & Nusbaum argue that speech comprehension processes may be far from automatic in dialogue. We are quite happy to accept this general point but see no prob-

lems for our proposal. In our account, the process of aligning the structures used in comprehension (and production) is automatic, but other aspects of comprehension (and production) are presumably not automatic. Additionally, their conception of automaticity is that it "implies a passive process in which the input is processed in an invariant, inflexible manner, regardless of the beliefs and expectations of the listener." This is not the notion of automaticity that we intend, and we hope that the above discussion of Bargh (1989) helps to make this clear.

Our conception of automaticity also differs from a Skinnerian one, as suggested by **Pear**. Crucially, we assume that alignment is not due to reinforcement, just as Dijksterhuis and Bargh (2001) assume for the perception-behavior expressway. Instead, alignment follows from a primitive tendency to imitate that does not appear to be learned (e.g., Metzoff & Decety 2003). However, our account does share certain features with Skinner's (1957) account, in particular that alignment implicates low-level learning mechanisms.

R5. Parity

One concern is whether there is true representational parity between production and comprehension. **Ferreira** describes experiments in which participants plan to produce utterances that they know to be ungrammatical (i.e., participants do not simply make errors). She assumes that interlocutors use and understand such utterances during dialogue (which is almost certainly correct) and suggests that comprehenders in dialogue would regard them as illicit. Whereas it is possible that there are differences between monologue and dialogue with respect to judgments of grammaticality, we accept that such differences are unlikely. In her experiments, we suggest that speakers realize they are producing something ungrammatical, but do so anyway because they cannot think of any other way of saying what they want to say. As long as this realization takes place within the production system (i.e., does not purely occur during self-monitoring), there does not appear to be a problem for the parity assumption. Compare sports commentators who sometimes cannot identify a player at the point when they need to produce the utterance ("About to kick the ball, Smith"), which listeners might well regard as illicit. This account seems more likely than a real disconnect between grammars in comprehension and production. However, if there is a disconnect for some highly specific constructions, it merely leads to a very slight weakening of the parity assumption, not its abandonment.

Ginzburg argues that the interpretation of the same sequence of words can change according to whether it represents a single contribution from one speaker or two contributions from different speakers:

1. **A:** Which members of our team own a parakeet? **A:** Why? (= Why own a parakeet?)
2. **A:** Which members of our team own a parakeet? **B:** Why? (= Why are you asking which members of our team own a parakeet?)

He suggests that our interactive alignment mechanism cannot account for the fact that *Why?* has a different interpretation in interactions (1) and (2). This is an interesting observation, but the difference in interpretation between (1) and (2) hinges on the dialogue move (e.g., questioning, answering, checking, informing) being performed at that point. Because dialogue moves are generally associated with particular speakers, it is obviously crucial that interlocutors monitor the source of an utterance when interpreting it (as also follows from results like those of Metzing & Brennan 2003). For example, the speaker treats a question from his interlocutor differently from the way he would treat a question from himself. We accept that interlocutors can monitor the source of a contribution (i.e., they can differentiate between what they are saying and what their partner is saying) and can take this into account in their interpretation at the level of the dialogue move.

Cutting proposes parity for semantic but not phonological representations on the basis of picture-word interference experiments. From his brief description, we suggest that participants process the words that they actually produce both semantically and phonologically, but that they process the words that they are told to ignore semantically but not phonologically (or at least not to a sufficient depth to affect priming). **Krauss & Pardo** also question evidence for phonological alignment (and by implication for phonological parity). Although we accept that Goldinger (1998) does not directly demonstrate phonological alignment, recent evidence does support parity between production and comprehension at this level (Fowler et al. 2003).

Kempson defends a more radical proposal that parity comes from the symmetry between production and parsing processes. In her Dynamic Syntax account of parsing, syntactic information is combined with lexical information, which define semantic interpretations that are built up word-by-word. Production is assumed to work in essentially the same way. Hence, she sees interactive alignment as operating at the level of the production and parsing processes themselves. This is a challenging linguistic proposal, but it would need explicit modeling before it could be incorporated into a mechanistic account of language processing in dialogue.

R6. Is it only misunderstanding that drives interactive repair?

One concern is whether interactive repair is driven primarily by comprehension failure, as we proposed in section 4.3 of the target article. **Healey** points out that even in the context of Garrod and Anderson's (1987) maze-game dialogues, interlocutors change their description scheme in a systematic fashion (e.g., shifting from a *path* or *figural* scheme to a *line* or *matrix* scheme). He argues that it is unlikely that this systematic shift can be accounted for only in terms of an interactive repair mechanism based on comprehension failure. Of course we recognize (see sect. 4.4 of the target article) that alignment does not depend only upon this process. There are many things that determine what people choose to say and even how they do so which go beyond the simple automatic mechanisms discussed in the target article. For example, the shift in description scheme that Healey mentions probably reflects two opposing pressures. Whereas the abstract *line* and *matrix* descriptions are more efficient over a period of time than *figural* or *path* descriptions (e.g., a *line* or *matrix* description involves few words and is not influenced by whether the position is near a salient point in the maze or lies in a salient pattern), they are more difficult to align (e.g., *matrix* descriptions depend upon alignment of the origin and of the counting conven-

tions used). So they can often not be used securely until there is a richer implicit common ground (e.g., repeated use of path descriptions which begin at one corner of the maze can lead to this corner being adopted as the origin for matrix description). We suspect that once the implicit common ground has become sufficiently rich to support the more abstract description, a speaker is more likely to adopt that scheme when he encounters a position that is particularly awkward to describe even when it requires a violation of alignment.

Note that this shift occurs without the speaker having to take account of the listener's knowledge. **Healey** therefore brings up an important general point, that interlocutors can go beyond interactive alignment and repair in ways that do not require other-modeling or the establishment of a full common ground. For example, a speaker can decide that a representational scheme is unnecessarily complex or a referring expression is unnecessarily long even if the interlocutors have aligned on that scheme or expression. Similarly, in preparing lectures, I might change how I am speaking on the basis of my knowledge of the audience (full common ground), but I might also do it on the basis that "Hang on, I'm not doing this efficiently, given my own resources – I am trying to remember too much and can't manage it." This might be argued to involve access to a second model of one's own mental state, which is therefore costly, but less costly than keeping track of full common ground. Such decisions require there to be some inhibition of the basic alignment process in light of a conflicting goal (see sect. R4). In conclusion, Healey's point reflects something that is additional to our account rather than in conflict with it (cf. **Krauss & Pardo**, who point out that not only misunderstanding drives accommodation).

R7. Other modeling

Although interlocutors undeniably do pay attention to each others' mental states on occasion, our contention is that such other-modeling is resource-intensive, essentially because it involves storing two representations: a representation of one's own state of knowledge, and a separate representation of one's partner's state of knowledge. We therefore believe that most of the process of alignment occurs via the interactive-alignment mechanism where other-modeling is not required. But we stress that other-modeling is not purely used for "difficult" cases of interactive repair when automatic processes fail to work. When a boy decides to tell his mother what happened at school today, he presumably realizes that his mother does not know about the event in question, and therefore that he knows something that his mother does not know. This explicit modeling of the difference between knowledge states leads to him running to tell his mother about the event, and does not follow from the failure of interactive alignment and interactive repair. Similarly, a bilingual decides which language to speak on the basis of his assumptions about which language his listener knows. However, the undeniable use of such "broad-brush" other-modeling does not mean that other-modeling is employed in a fine-grained way to explain detailed decisions about one's individual contributions to an ongoing dialogue.

In this context, **Krauss & Pardo** point to evidence that speakers sometimes modulate their utterances to take into account their knowledge of the listener: They produce more informative contributions when they perceive their addressees to be less knowledgeable about the relevant topic (see also Isaacs & Clark 1987). The evidence from Kingsbury (1968) shows that speakers do not simply pay attention to what they believe about their specific interlocutor but make inferences about how much such a person is likely to know on the basis of the evidence at hand, which is, in this case, made quite deliberately apparent to the speaker (e.g., the questioner frames a question to stress his ignorance of the city). In general, we suspect that speakers make a one-off decision based on such issues as the perceived expertise of their addressees about how to frame their contributions (e.g., the decision not to make any assumptions about local geographic knowledge). A teacher can be much less explicit in the common room than in the classroom, and a mother does not speak motherese to her friends. Such decisions need not remain fixed for the whole conversation (e.g., they might change when the speaker guesses that his addressee is not a local but then realizes he is mistaken). But such a change is very different from a continuous, dynamic process of utterance accommodation based on full common-ground inference, which we argue to be implausible for reasons of resource limitations (see sect. 4.1 of the target article). We are therefore grateful to Krauss & Pardo for stressing that explicit modeling does not only occur when automatic processes fail to produce alignment, but we see no concern for our assumption that automatic mechanisms underlie alignment.

Fussell & Kraut argue that speakers with different views of a spatial scene take into account the listener's perspective, in effect modeling the listener's mental state. They describe a collaborative bicycle repair situation in which an expert helper guides a novice repairer. They note that when the repairer knows that he can be seen even when he cannot see his remote helper, he will use deictic expressions to describe the things in front of him (e.g., *See this piece,* while pointing at a cycle component), whereas the remote helper will not (e.g., *See the derailleur*). They argue that this is inconsistent with alignment and provides further evidence of other modeling. We are not convinced. We suspect that speakers in this situation prefer to use deictic expressions because they are shorter, do not require word finding, and so on. But deixis is not an option for the remote helper because he cannot point to anything. Instead, he has to fall back on more complex nondeictic descriptions. (One remote helper is quoted as saying in frustration, "If I could point to it, it's right there"; Kraut et al. 2003, p. 36.) So the circumstances may force the speaker to use a more complex nonaligned utterance. It is of course reasonable that alignment is broken under such circumstances, because it simply would not work. One important point this raises is that the tendency toward alignment is likely to be stronger under conditions where two interlocutors are placed in comparable environments. Presumably this reflects nonlinguistic contributions to linguistic alignment (see also the discussion of **Dominey** in sect. R11).

Nevertheless, we certainly agree with the general point that when communicators share a physical situation they take situational awareness into account in formulating utterances. But is this evidence for listener modeling? In the "side-by-side" situation described by Kraut et al. (2003, communicators use direction of gaze to establish joint attention, but the effect of one partner's point of gaze on the

other partner's focus of attention reflects low-level mechanisms which do not depend on inferences about the partner's mental state (see Schuller & Rossion 2001). And, because in this situation what is accessible to the speaker will usually be equally accessible to the partner (see sect. 4.1), an essentially egocentric approach will generally support successful communication without requiring speakers to model their listeners.

Schober suggests that current evidence cannot distinguish two possibilities: that interlocutors only model each other's mental states under exceptional circumstances, and that interlocutors normally model each other's mental states and only fail to do so when under great cognitive load or when circumstances weigh heavily against doing so. We accept that current evidence does not distinguish between these two positions. However, our account assumes the use of fewer resources and is parsimonious (obviously, an account containing two mental models is harder to falsify than an account limited to one, just as a parallel account is harder to falsify than a serial account). Moreover, Schober's proposal cannot hold for multiparty dialogues containing more than a small number of people, because it must become impossible to retain and regularly update a different mental model for each person. In general, cognitive psychology teaches us that constructing mental models is hard and holding onto different models at the same time is especially hard (e.g., Johnson-Laird 1983). We suggest that the paradox of how one can know when to model one's partner is easy to accommodate: Contributions to the dialogue will make it clear that alignment is breaking down, and if interactive repair does not solve the problem, the interlocutor is forced to assume that what his partner knows is likely to be different from what he knows. Even in such cases, it may be that interlocutors only model those differences between themselves and their partners that need to be modeled in order to allow the recovery of alignment.

Brennan & Metzing also criticize our assumption that interlocutors do not routinely employ full common ground. A fast-growing body of literature suggests that interlocutors sometimes do pay attention to each others' knowledge in comprehension and production (e.g., Hanna et al. 2003; Lockridge & Brennan 2002; Nadig & Sedivy, 2002) and sometimes do not (e.g., Brown & Dell 1987; Ferreira & Dell 2000; Keysar et al. 2003). It is too early to say precisely when such knowledge can affect processes of production and comprehension, but current evidence suggests both that interlocutors can immediately draw upon knowledge about differences between their own knowledge and their beliefs about their partner's knowledge, and that interlocutors can make egocentric decisions about production and comprehension. Most of these studies involve a fairly artificial situation in which the experimental subject is informed that his interlocutor may have knowledge about the situation that differs from his in quite specific ways. Experiments like that of Hanna et al. (2003) show that it can be straightforward to add *one fact* about your interlocutor – namely, that he does not have access to a particular piece of information that you have. Even in such cases, some egocentric behavior remains, as Hanna et al. acknowledge and Keysar et al. (2003) demonstrate. But adding one fact about your interlocutor's knowledge is quite different from maintaining a full representation of the interlocutor's situation model, and performing reasoning based on that model. Available resources do not normally allow interlocutors to constantly update models of each others' mental states. However, this does not lead to communicative breakdown because aligned interlocutors develop the same situation models.

In response to **Brennan & Metzing**, we stress that it was not our intention to commit to a two-stage account (e.g., Horton & Keysar 1996), in which other modeling occurs during revision but not during initial processing (whether production or comprehension). We note that **Krauss & Pardo** and **Brown-Schmidt & Tanenhaus** also interpret us as making this proposal, and accept that we did not make this very clear. Rather, we claim that "performing inferences about common ground is an optional strategy that interlocutors employ only when resources allow" (target article, sect. 4.2, para. 4).

We do not regard Metzing and Brennan's (2003) demonstration of partner-specific effects as problematic, and assume they can be explained in similar ways to Branigan et al.'s (2003) demonstration that syntactic alignment is sensitive to participant status (see sect. 2.3 of the target article). As we have pointed out in section R4 of this response, we assume that alignment is automatic at a postconscious level (Bargh 1989) and, hence, can be affected by a range of social factors from stereotype activation to participant status. A particular speaker is associated with a particular form, and breaking that association causes disruption. There is no need for other modeling to occur in this process of partner-specific lexical entrainment. The term *conceptual pact* appears to suggest that other modeling is used in lexical entrainment. If so, we would question whether it is generally appropriate.

R8. Routines

Schiller & de Ruiter propose that interactive alignment necessarily involves selecting stored fragments from previous utterances. This corresponds to our notion of routinization (see sect. 5 of the target article). We suspect that routinization comes about as a result of a longer lasting alignment mechanism based on memory retrieval rather than transient activation. This is because routines reflect multiple links between different levels of representation (e.g., they fix the relation between a word and its meaning, its syntactic form, and even its interpretation within a situation model) and it is difficult to imagine how this could be captured and routinized through purely transient activation. Rather than assume that routinization is the sole explanation of alignment, we suggest that it is a consequence of implicit learning but that transient activation also promotes alignment (see sect. R3). It may of course be that routines emerge from a resonance process, as **Goldinger & Azuma** suggest. In addition, because routinization works by linking levels of representation, it may explain how alignment percolates up from lower to higher levels (cf. issues raised by **Warren & Rayner** and **Branigan**, as discussed in sect. R3).

Within the interactive-alignment account, we regard routines as an extreme case of alignment, involving a fixed form and interpretation. It may be best to think of routinization as falling on a continuum, with expressions that contain some fixed elements (as in many of Kuiper's 1996 examples) being more or less "semi-routinized." Assuming that it is correct to regard alignment as a mixture of tran-

sient activation and implicit learning, we propose that the more routinized an expression is, the more it is best explained in terms of implicit learning – for the purposes of the conversation at least, the expression and its interpretation are stored and retrieved. Of course, if an expression becomes sufficiently entrenched, it may survive that conversation. Although other frameworks are no doubt possible, we regard Jackendoff's (2002) account of fixed and semi-fixed expressions as an appropriate representational scheme for semi- and completely routinized expressions (see Pickering & Garrod, in press, for discussion).

R9. Self-monitoring

Schiller & de Ruiter question our claim that self-monitoring can occur at any level of linguistic representation that can be aligned. We did not claim that there is conclusive evidence for this hypothesis and we believe that careful empirical work is needed to distinguish our proposal from the proposal that monitoring works externally on sound and internally on phonological representations alone. However, we would query whether the reported evidence provides strong support for this alternative proposal. First, the comparative slowness of selecting a gender label in comparison to selecting the indefinite article in French may have many explanations, perhaps most likely that selecting between genders is a more abstract and difficult task than selecting between (very common) words. Second, the strong evidence for monitoring of various aspects of phonological representations is completely compatible with monitoring of other linguistic representations. Although some or all gender-congruency effects in picture-word interference tasks may really be determiner congruency effects (Schiller & Caramazza 2003), there is also considerable evidence that grammatical gender can be accessed when phonological form is not available (Badecker et al. 1995; Vigliocco et al. 1997). Therefore, it is at least plausible that people can directly monitor for errors of grammatical gender and indeed for other aspects of syntactic representations. If an utterance is ill-formed at different levels of representation simultaneously, we suspect that there may be a race between monitoring processes at these different levels, in which case it might not always be possible to detect monitoring that takes place at the "slower" level.

R10. On the difference between dialogue and monologue

A number of commentators argue that language processing in dialogue is not fundamentally different from that in monologue. For example, both **Barr & Keysar** and **Glucksberg** point out that the same basic language processes operate in monologue and dialogue so there is no principled difference between the two. We agree in the sense that the actual production and comprehension mechanisms are the same (at what we might term a "microlevel"). However, the radically different contexts in which they operate lead to very different results. For example, a speaker's utterances are dramatically affected by the presence of the interlocutor – the speaker aligns with the interlocutor's utterances via the mechanisms we have described. In this respect we argue that the language processing system is designed for dialogue rather than monologue. As a result, speakers have to learn special strategies to deal with monologue which are not required during dialogue processing.

We agree with **Glucksberg** that dialogue is not necessarily easier than monologue, and accept that contextual effects can be very strong in monologue. We propose that the priming mechanisms are ideally suited for dialogue. Presumably they have developed from imitation (Arbib, in press) and it may be that the organization of dialogue (e.g., time between turns) is optimal for the mechanisms of priming. Therefore, dialogue does not need to rely on nonautomatic inference. In contrast, monologue cannot use priming between interlocutors (by definition) and therefore has to rely on inference, other-modeling, and so on. Priming is of course present in monologue, but we contend that it is far less useful than in dialogue (e.g., repetition is much rarer in monologue than in dialogue; see sect. 5.1 of the target article). So we concur that there is not a principled distinction between dialogue and monologue, but at the same time maintain that dialogue will usually but not always be easier than monologue.

Barr & Keysar appear to disagree with us more than we think they actually do. They are mistaken in assuming that we propose a categorical distinction between monologue and dialogue. In section 8 of the target article, we refer to a "dialogic continuum" with monologue at one end, and fully interactive dialogue at the other. We assume that the same mechanisms are present in dialogue and monologue (i.e., people do not set some processing "switch"). In true monologue, the speaker has no interlocutor to align with. He can of course align with himself and certainly does so (e.g., re-using the same word with the same meaning). We completely agree that dialogues go through various stages, with some involving rapid turn-taking (e.g., question answering) and some involving much more limited feedback (e.g., during narratives). Boden (1994) distinguishes between conversational phases and presentational phases in group discussion. These presentational phases are *not* monologues, as even minimal feedback affects them considerably (Bavelas et al. 2000).

Hence, we stress that monologue and dialogue lie on a continuum, and we predict that the degree of alignment will be affected by the position on the continuum. One important area for research is to consider the effects of dialogue genre on alignment (in which context we can regard monologue as particular genre). For example, **Schegloff** points to the importance of different speech-exchange systems (conversation, interview, giving a speech, etc.) in affecting the characteristics of the dialogue (e.g., turn-taking behavior, routinization). We predict that the rate and characteristics of alignment are not constant for all forms of dialogue, but will depend on the speech-exchange system. For example, forms of interaction that do not allow unconstrained feedback and where turn-taking is externally managed (e.g., interviews) will fail to employ the interactive repair mechanism to the extent that is possible in casual conversation.

R11. Extensions and discussion

Schegloff argues that our mechanistic account fails to consider the richness of the interaction afforded by dialogue. Although Schegloff's sociological starting point (i.e., in terms of organizational practice and interaction contingen-

cies) is somewhat different from ours, we certainly agree that there are additional specific details of dialogue organization that must enter into any complete mechanistic account. We also recognize the considerable contribution that Schegloff and colleagues have made in mapping out the details of these organizational practices and the contingencies they afford. However, our mechanistic aspiration goes beyond mapping out such practices and contingencies. Like **Brown-Schmidt & Tanenhaus** we believe that a mechanistic account should make it possible to formulate a computational model of the processes involved in the comprehension and production of dialogue and how these take advantage of the interactional nature of dialogue. We also recognize that any complete model will have to take account of both self and other commitments in dialogue processing (see our response to **Ginzburg** in sect. R5). We stress that our paper is entitled "*Toward* a mechanistic psychology of dialogue"!

Two commentators argue for a broadening of the interactive alignment account to include other interactive situations. **Mazur** proposes that interactive alignment needs to be embedded in a broader theory of communication that pays attention to a range of social conventions. We agree that a full theory of interactive alignment will make reference to nonlinguistic as well as linguistic information, and believe that our suggestions about the relations between our account and implicit social cognition is a step in this direction.

Dominey draws interesting parallels between the interactive alignment process in adult dialogue and certain features of language acquisition. Language learning depends upon extralinguistic or prelinguistic alignment mechanisms (e.g., establishing joint attention on intended referents through gaze direction or postural orientation). Also, there is evidence that routinization of utterances associated with repeated action scenarios (feeding, bathing, etc.) may play an important role in the acquisition of syntax (Tomasello 2003). These suggestions help reinforce the claim that nonlinguistic alignment may lead to linguistic alignment, just as linguistic alignment at one level leads to linguistic alignment at other levels (see our discussion of **Fussell & Kraut** in sect. R7). In fact, Dominey suggests that such linguistic/nonlinguistic links are necessary to explain the process of language acquisition, where one partner (the infant) does not initially have linguistic abilities. A full theory of how interactive alignment might explain acquisition would be fascinating. In particular, we are intrigued by the suggestion that learning by alignment might avoid the enlistment of generative grammar mechanisms, perhaps in a way similar to that envisaged by Tomasello.

Language acquisition is a good example of how it may be possible to extend our account into new domains. Other areas that we have highlighted at various points in the target article and this response include social psychology and human-computer interaction. A recurring theme is that it may be sensible to include nonlinguistic alignment into developments of our model; interlocutors who are aligned in nonlinguistic (e.g., body posture) or paralinguistic (e.g., tone of voice) ways may be more likely to align linguistically.

We emphasize that our use of the term "priming" is at a fairly abstract functional level, as our notion of automaticity makes clear (sect. R4). It allows nonconscious mediation by factors that may originate in distinctions that interlocutors are aware of (e.g., participant status, social status, cooperativeness). We also note that "priming" may employ transient activation or implicit learning or both. To be more speculative, we suspect that interactive alignment may work by two distinct mechanisms: a brief activation-based process that may not be affected by intentional distinctions, and a longer-lasting memory-based process that is intentionally mediated. The effects of these two processes will depend on precise timing, and will therefore be differentially affected by aspects of the conversation that affect timing. For example, a high-engagement face-to-face dialogue between intimate friends may result in timing that is precisely attuned to increasing alignment, whereas a dialogue between strangers that depend on external factors such as rules of engagement (e.g., in an interview) or technology (e.g., walkie-talkies) may not. We suspect that the longer-lasting process will not be affected but the activation process might be impaired in low-involvement dialogue. These speculative comments could inform an extensive program of empirical research concerned with the conditions that lead to alignment in dialogue (e.g., its time course).

The other obvious area for development is explicit computational modeling, as highlighted by **Brown-Schmidt & Tanenhaus** in particular. To perform such modeling, it would of course be necessary to explicate many assumptions of our account that are currently vague or implicit, for instance by developing interactive alignment, interactive repair, and other-modeling components. It would be necessary to model the process whereby alignment at one level leads to alignment at other levels, and to understand how conflicts of alignment are resolved (see **Warren & Rayner**). We need to know whether transient activation and implicit learning should be distinguished, and if so, how they interact. Finally, any such account should explain the process of routinization and describe its effects on alignment.

NOTE

1. Note that the uses of "interactive" in *interactive alignment* and *interactive activation* are unrelated.

References

Letters "a" and "r" appearing before authors' initials refer to target article and response respectively.

Ades, A. & Steedman, M. J. (1982) On the order of words. *Linguistics and Philosophy* 4:517–58. [aMJP]

Aijmer, K. (1996) *Conversational routines in English: Convention and creativity.* Longman. [aMJP]

Allen, J. F., Byron, D. K., Dzikovska, M., Ferguson, G., Galescu, L. & Stent, A. (2001) Towards conversational human-computer interaction. *AI Magazine* 22:27–35. [SB-S]

Altenberg, B. (1990) Speech as linear composition. In: *Proceedings from the Fourth Nordic Conference for English Studies,* ed. G. Caie, K. Haastrup, A. L. Jakobsen, J. E. Nielsen, J. Sevaldsen, H. Specht & A. Zettersten, pp. 133–43. University of Copenhagen. [aMJP]

Altmann, G. T. & Steedman, M. J. (1988) Interaction with context during human sentence processing. *Cognition* 30:191–238. [aMJP]

Amis, K. (1997) *The King's English: A guide to modern English usage.* Harper Collins. [aMJP]

Anderson, A. H. & Boyle, E. (1994) Forms of introduction in dialogues: Their discourse contexts and communicative consequences. *Language and Cognitive Processes* 9:101–22. [aMJP]

Arbib, M. A. (in press) From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences.* [rMJP]

Badecker, W., Miozzo, M. & Zanuttini, R. (1995) The dissociation of form-based lexical retrieval from access to word-specific aspects of grammar: Evidence for two-stage lexical processing. *Cognition* 57:193–216. [rMJP]

Balota, D. A. (1990) The role of meaning in word recognition. In: *Comprehension processes in reading,* ed. D. A. Balota, G. B. Flores d'Arcais & K. Rayner. Erlbaum. [JCC]

Balota, D. A., Paul, S. T. & Spieler, D. H. (1999) Attentional control of lexical processing pathways during word recognition and reading. In: *Language processing,* ed. S. Garrod & M. Pickering. Psychology Press. [aMJP]

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G. & Newlands, A. (2000) Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language* 42:1–22. [aMJP]

Bargh, J. A. (1989) Conditional automaticity: Varieties of automatic influence in social perception and cognition. In: *Unintended thoughts,* ed. J. S. Uleman & J. A. Bargh, pp. 3–51. Guilford Press. [rMJP, HS]

Bargh, J. A. & Chartrand, T. L. (1999) The unbearable automaticity of being. *American Psychologist* 54:462–79. [aMJP]

Bargh, J. A., Chen, M. & Burrows, L. (1996) Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology* 71:230–44. [aMJP]

Baron-Cohen, S., Tager-Flusberg, H. & Cohen, D. (2000) *Understanding other minds: Perspectives from developmental neuroscience.* Oxford University Press. [aMJP]

Barr, D. J. & Keysar, B. (2002) Anchoring comprehension in linguistic precedents. *Journal of Memory and Language* 46:391–418. [DJB]

(in press) Making sense of how we make sense: The paradox of egocentrism in language use. In: *Figurative language comprehension: Social and cultural influences,* ed. H. L. Colston & A. N. Katz. Erlbaum. [DJB]

Barton, S. & Sanford, A. J. (1993) A case-study of pragmatic anomaly-detection: Relevance-driven cohesion patterns. *Memory and Cognition* 21:477–87. [rMJP]

Barwise, J. (1989) Three views of common knowledge. In: *The situation in logic,* ed. J. Barwise. CSLI. [aMJP]

Bates, E. & MacWhinney, B. (1987) Competition, variation and language learning. In: *Mechanisms of language acquisition,* ed. B. MacWhinney & E. Bates. Erlbaum. [PFD]

Bavelas, J. B., Coates, L. & Johnson, T. (2000) Listeners as co-narrators. *Journal of Personality and Social Psychology* 79:941–52. [arMJP]

Binder, K. S. & Rayner, K. (1998) Context strength does not modulate the subordinate bias effect: Evidence from eye fixations and self-paced reading. *Psychonomic Bulletin and Review* 5:271–76. [aMJP]

Bock, J. K. (1986a) Meaning, sound, and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 12:575–86. [aMJP]

(1986b) Syntactic persistence in language production. *Cognitive Psychology* 18:355–87. [aMJP]

(1989) Closed class immanence in sentence production. *Cognition* 31:163–86. [aMJP]

(1996) Language production: Methods and methodologies. *Psychonomic Bulletin and Review* 3:395–421. [JCC, aMJP]

Bock, J. K. & Griffin, Z. M. (2000) The persistent of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General* 129:177–92. [JCC, MK, rMJP]

Bock, J. K. & Huitema, J. (1999) Language production. In: *Language processing,* ed. S. Garrod & M. Pickering. Psychology Press. [aMJP]

Bock, J. K. & Levelt, W. J. M. (1994) Language production: Grammatical encoding. In: *Handbook of psycholinguistics,* ed. M. A. Gernsbacher, pp. 945–84. Elsevier/Academic Press. [JCC, aMJP]

Bock, J. K. & Loebell, H. (1990) Framing sentences. *Cognition* 35:1–39. [aMJP]

Bock, J. K., Loebell, H. & Morey, R. (1992) From conceptual roles to structural relations: Bridging the syntactic cleft. *Psychological Review* 99:150–71. [aMJP]

Boden, D. (1994) *The business of talk.* Polity Press. [rMJP]

Boroditsky, L. (2000) Metaphorical structuring: Understanding time through spatial metaphors. *Cognition* 75:1–28. [aMJP]

Bourhis, R. Y. & Giles, H. (1977) The language of intergroup distinctiveness. In: *Language, ethnicity and intergroup relations,* ed. H. Giles, pp. 119–35. Academic Press. [RMK]

Branigan, H. P., Pickering, M. J. & Cleland, A. A. (1999) Syntactic priming in written production: Evidence for rapid decay. *Psychonomic Bulletin and Review* 6:635–40. [rMJP]

(2000) Syntactic coordination in dialogue. *Cognition* 75:B13–B25. [arMJP]

(submitted) Syntactic alignment and participant status in dialogue. [arMJP]

Brennan, S. E. & Clark, H. H. (1996) Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22:1482–93. [arMJP]

Brennan, S. E. & Schober, M. F. (2001) How listeners compensate for dysfluencies in spontaneous speech. *Journal of Memory and Language* 44:274–96. [aMJP]

Brooks, P. & Tomasello, M. (1999) Young children learn to produce passives with nonce verbs. *Developmental Psychology* 35:29–44. [aMJP]

Brown, P. M. & Dell, G. S. (1987) Adapting production to comprehension: The explicit mention of instruments. *Cognitive Psychology* 19:441–72. [SEB, arMJP]

Brown-Schmidt, S., Campana, E. & Tanenhaus, M. K. (2002) Reference resolution in the wild: How addresses circumscribe referential domains in a natural, interactive problem-solving task. Paper presented at the Annual Meeting of the Cognitive Science Society, Fairfax, VA, August 2002. [SB-S]

(in press) Real-time reference resolution by naïve participants during a task-based unscripted conversation. In: *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions,* ed. J. C. Trueswell & M. K. Tanenhaus. MIT Press. [SB-S, aMJP]

Burnard, L. (2000) *Reference guide for the British National Corpus (World Edition).* Oxford University Computing Services. [aMJP]

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C. R., McGuire, P. K., Woodruff, P. W. R., Iversen, S. D. & David, A. S. (1997) Activation of auditory cortex during silent lipreading. *Science* 276:593–96. [aMJP]

Capella, J. (1981) Mutual influence in expressive behavior: Adult-adult and infant-adult dyadic interaction. *Psychological Bulletin* 89:101–32. [SDG]

Caramazza, A. (1991) *Issues in reading, writing, and speaking. A neuropsychological perspective.* Kluwer Academic. [JCC]

(1997) How many levels of processing are there in lexical access? *Cognitive Neuropsychology* 14:177–208. [aMJP]

Carlson, K. (2001) The effects of parallelism and prosody in the processing of gapping structures. *Language and Speech* 44:1–26. [aMJP]

Carlson-Radvansky, L. A. & Jiang, Y. (1998) Inhibition accompanies reference frame selection. *Psychological Science* 9:386–91. [aMJP]

Carpenter, G. & Grossberg, S. (1987) A massively parallel architecture for a self-organizing neural recognition machine. *Computer Vision, Graphics, and Image Processing* 37:54–115. [SDG]

Carpenter, G. & Grossberg, S., eds. (1991) *Pattern recognition by self-organizing neural networks.* MIT Press. [SDG]

Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H. & Carlson, G. N. (2002) Circumscribing referential domains in real-time sentence comprehension. *Journal of Memory and Language* 47:30–49. [aMJP]

Chang, F., Dell, G. S., Bock, J. K. & Griffin, Z. M. (2000) Structural priming as implicit learning: A comparison of models of sentence production. *Journal of Psycholinguistic Research* 29:217–29. [rMJP]

Charniak, E. (1993) *Statistical language learning.* MIT Press. [aMJP]

Chartrand, T. L. & Bargh, J. A. (1999) The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology* 76:893–910. [arMJP]

Cherry, E. C. (1956) *On human communication.* MIT Press. [aMJP]

Chomsky, N. (1965) *Aspects of the theory of syntax.* MIT Press. [HPB, aMJP]

(1973) Conditions on transformations. In: *A Festschrift for Mossis Halle,* ed. S. R. Anderson & R. Kiparsky, pp. 232–86. Holt, Rinehart and Winston. [FF]

(1981) *Lectures on government and binding.* Foris. [aMJP]

(1995) *The minimalist program.* MIT Press. [aMJP]

Christiansen, M. H. & Chater, N. (2001) Connectionist psycholinguistics: Capturing the empirical data. *Trends in Cognitive Sciences* 5:82–8. [SB-S]

Clark, E. V. (1993) *The lexicon in acquisition.* Cambridge University Press. [aMJP]

Clark, H. H. (1979) Responding to indirect speech acts. *Cognitive Psychology* 11:430–77. [aMJP]

(1985) Language and language users. In: *The handbook of social psychology, 3rd edition,* ed. G. Lindzey & E. Aronson. Harper Row. [aMJP]

(1992) *Arenas of language use.* University of Chicago Press. [aMJP]

(1996) *Using language.* Cambridge University Press. [JG, SG, RMK, ABM, aMJP]

(1998) Communal lexicons. In: *Context in language learning and language understanding,* ed. K. Malmkjoer & J. Williams. Cambridge University Press. [aMJP]

Clark, H. H. & Marshall, C. R. (1981) Definite reference and mutual knowledge. In: *Elements of discourse understanding,* ed. A. K. Joshi, I. A. Sag & B. L. Webber. Cambridge University Press. [DJB, SRF, JG, aMJP]

Clark, H. H. & Murphy, G. L. (1982) Audience design in meaning and reference. In: *Language and comprehension,* ed. J. F. Le Ny & W. Kintsch. North-Holland. [RMK, aMJP]

Clark, H. H. & Schaefer, E. F. (1987) Concealing one's meaning from overhearers. *Journal of Memory and Language* 26:209–25. [aMJP]

Clark, H. H. & Schober, M. F. (1991) Asking questions and influencing answers. In: *Questions about questions: Inquiries into the cognitive bases of surveys,* ed. J. M. Tanur, pp. 15–48. Russell Sage Foundation. [MFS]

Clark, H. H. & Wilkes-Gibbs, D. (1986) Referring as a collaborative process. *Cognition* 22:1–39. [arMJP, MFS]

Cleland, A. A. & Pickering, M. J. (2003) The use of lexical and syntactic information in language production: Evidence from the priming of noun-phrase structure. *Journal of Memory and Language* 49:214–30. [arMJP]

Clement, C. A. & Gentner, D. (1991) Systematicity as a selection constraint in analogical mapping. *Cognitive Science* 15:89–132. [ABM]

Coates, J. (1990) Modal meaning: The semantic-pragmatic interface. *Journal of Semantics* 7:53–64. [aMJP]

Conrad, F. G. & Schober, M. F. (2000) Clarifying question meaning in a household telephone survey. *Public Opinion Quarterly* 64:1–28. [MFS]

Consortium, T. T. (2000) *The TRINDI Book.* University of Gothenburg. Available at: http://www.ling.gu.se/research/projects/trindi. [JG]

Cooper, R., Larsson, S., Hieronymus, J., Ericsson, S., Engdahl, E. & Ljunglof, P. (2000) GODIS and questions under discussion. In: *The TRINDI Book.* University of Gothenburg. Available at: http://www.ling.gu.se/research/projects/trindi. [JG]

Creswell, C. (2002) Resumptive pronouns, wh-island violations, and sentence production. In: *Proceedings of the Sixth International Workshop on Tree Adjoining Grammar and Related Frameworks (TAG+6)*, pp. 101–109. Universita di Venezia. [FF]

Cutting, J. C. (1998) The production and comprehension lexicons: What is shared and what is not. Doctoral dissertation, University of Illinois, 1997. *Dissertation Abstracts International: Section B: The Sciences and Engineering* 58:10–B. [JCC]

Cutting, J. C. & Bock, J. K. (1997) That's the way the cookie bounces: Syntactic and semantic components of experimentally elicited idiom blends. *Memory and Cognition* 25:57–71. [JCC, aMJP]

de Boer, B. (2000) Self-organization in vowel systems. *Journal of Phonetics* 28:441–65. [SDG]

Dell, G. S. (1986) A spreading-activation theory of retrieval in sentence production. *Psychological Review* 93:283–321. [SDG, aMJP]

Dell, G. S, Chang, F. & Griffin, Z. M. (1999) Connectionist models of language production: Lexical access and grammatical encoding. *Cognitive Science* 23:517–42. [JCC]

Desrochers, A. & Paivio, A. (1990) Le phonem initial des noms inanimes et son effet sur l'identification du genre grammatical [The initial phoneme of inanimate nouns and its effect for the identification of grammatical gender]. *Canadian Journal of Psychology* 44:44–57. [NOS]

Desrochers, A., Paivio, A. & Desrochers, S. (1989) L'effet de la frequence d'usage des noms inanimes et de la valeur predictive de leur terminaison sur l'identification du genre grammatical [The effect of frequency of usage of inanimate nouns and of the predictive value of their endings on the identification of grammatical gender]. *Canadian Journal of Psychology* 43:62–73. [NOS]

Dijksterhuis, A. & Bargh, J. A. (2001) The perception-behavior expressway: Automatic effects of social perception on social behavior. In: *Advances in experimental social psychology, vol. 33,* ed. M. P. Zanna, pp. 1–40. Academic Press. [arMJP]

Dijksterhuis, A. & van Knippenberg, A. (1998) The relation between perception and behavior or how to win a game of Trivial Pursuit. *Journal of Personality and Social Psychology* 74:865–77. [aMJP]

Doherty-Sneddon, G., Anderson, A. H., O'Malley, C., Langton, S., Garrod, S. & Bruce, V. (1997) Face-to-face and video-mediated communication: A comparison of dialogue structure and task performance. *Journal of Experimental Psychology: Applied* 3:105–25. [aMJP]

Dominey, P. F. (2000) Conceptual grounding in simulation studies of language acquisition. *Evolution of Communication* 4(1):57–85. [PFD]

(2003a) Learning grammatical constructions in a miniature language from narrated video events. In: *Proceedings of the 25th Annual Meeting of the Cognitive Science Society.* (in press). Available at: http://www.isc.cnrs.fr/wp/DomineyISC-WP0.pdf. [PFD]

(2003b) Learning grammatical constructions from narrated video events for human-robot interaction. In: *Proceedings of the 2003 IEEE Humanoid Robot Conference, Karlsruhe, Germany.* [PFD]

Drew, P. (1997) "Open" class repair initiators in response to sequential sources of troubles in conversation. *Journal of Pragmatics* 28:69–101. [EAS]

Duffy, S. A., Morris, R. K. & Rayner, K. (1988) Lexical ambiguity and fixation times in reading. *Journal of Memory and Language* 27:429–46. [aMJP]

Epley, N. & Gilovich, T. (1999) Just going along: Nonconscious priming and conformity to social pressure. *Journal of Experimental Social Psychology* 35:578–89. [rMJP]

Fay, N., Garrod, S. & Carletta, J. (2000) Group discussion as interactive dialogue or as serial monologue: The influence of group size. *Psychological Science* 11:481–86. [aMJP]

Fernández, R. & Ginzburg, J. (2002) Non-sentential utterances: Grammar and dialogue dynamics in corpus annotation. In: *Proceedings of the 19th International Conference on Computational Linguistics (CoLing),* pp. 253–59. Morgan Kaufman. [aMJP]

Ferreira, F. & Swets, B. (2003) The production and comprehension of resumptive pronouns in relative clause "island" contexts. Paper presented at the Max Planck Institute's Four Corners Workshop Series. Workshop 2: The relationship between language comprehension and language production. Nijmegen, the Netherlands, April 5, 2003. [FF]

Ferreira, V. S. & Dell, G. S. (2000) Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology* 40:296–340. [SEB, arMJP]

Ferreira, V. S. & Griffin, Z. M. (2003) Phonological influences on lexical (mis)selection. *Psychological Science* 14:86–90. [SG]

Fodor, J. A., Bever, T. G. & Garrett, M. F. (1974) *The psychology of language.* McGraw Hill. [aMJP]

Fowler, C. A., Brown, J. M., Sabadini, L. & Weihing, J. (2003) Rapid access to speech gestures in perception: Evidence from choice and simple response-time tasks. *Journal of Memory and Language* 49:396–413. [rMJP]

Fowler, C. & Housum, J. (1987) Talkers' signaling "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language* 26:489–504. [aMJP]

Frazier, L. & Rayner, K. (1990) Taking on semantic commitments: Processing multiple meanings vs. multiple senses. *Journal of Memory and Language* 29:181–200. [rMJP]

Frisson, S. & Pickering, M. J. (1999) The processing of metonymy: Evidence from eye-movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 25:1366–83. [rMJP]

(2001) Obtaining a figurative interpretation of a word: Support for underspecification. *Metaphor and Symbol* 16:149–71. [rMJP]

Fussell, S. R. & Krauss, R. M. (1989) Understanding friends and strangers: The effects of audience design on message comprehension. *Journal of Experimental Social Psychology* 19:509–26. [RMK]

(1992) Coordination of knowledge in communication: Effects of speakers' assumptions about what others know. *Journal of Personality and Social Psychology* 62:378–91. [SG, RMK, aMJP]

Fussell, S. R., Kraut, R. E. & Siegel, J. (2000) Coordination of communication: Effects of shared visual context on collaborative work. In: *Proceedings of the CSCW 2000 Conference on Computer-Supported Cooperative Work.* ACM Press. [SRF]

Fussell, S. R., Setlock, L. D. & Kraut, R. E. (2003) Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks. In: *Proceedings of the CHI 2003 Conference on Human Factors in Computing Systems.* ACM Press [SRF]

Gagné, C. L. (2001) Relation and lexical priming during the interpretation of noun-noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27:236–54. [aMJP]

Gagné, C. L. & Shoben, E. J. (2002) Priming relations in ambiguous noun-noun combinations. *Memory and Cognition* 30:637–46. [aMJP]

Garrett, M. (1980) Levels of processing in speech production. In: *Language production, vol. 1,* ed. B. Butterworth. Academic Press. [aMJP]

Garrod, S. (1999) The challenge of dialogue for theories of language processing. In: *Language processing,* ed. S. Garrod & M. Pickering. Psychology Press. [aMJP]

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

Garrod, S. & Clark, A. (1993) The development of dialogue co-ordination skills in schoolchildren. *Language and Cognitive Processes* 8:101–26. [HPB, aMJP]

Garrod, S. & Doherty, G. (1994) Conversation, co-ordination and convention: An empirical investigation of how groups establish linguistic conventions. *Cognition* 53:181–215. [PGTH, aMJP]

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

Garrod, S. & Sanford, A. J. (1977) Interpreting anaphoric relations: The integration of semantic information while reading. *Journal of Verbal Learning and Verbal Behavior* 16:77–90. [aMJP]

Gaskell, M. G. & Marslen-Wilson, W. D. (1997) Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes* 12:613–56. [SDG]

Gazdar, G., Klein, E., Pullum, G. & Sag, I. A. (1985) *Generalized phrase structure grammar.* Blackwell. [aMJP]

Gentner, D. (1983) Structure-mapping: A theoretical framework for analogy. *Cognitive Science* 7:155–70. [ABM]

Gentner, D. & Markman, A. B. (1997) Structural alignment in analogy and similarity. *American Psychologist* 52(1):45–56. [ABM,arMJP]

Gerrig, R. J. & Bortfeld, H. (1999) Sense creation in and out of discourse contexts. *Journal of Memory and Language* 41:457–68. [aMJP]

Giles, H., Coupland, N. & Coupland, J. (1991) Accommodation theory: Communication, context, and consequence. In: *Contexts of accommodation:*

Developments in applied sociolinguistics, ed. H. Giles, J. Coupland, & N. Coupland, pp. 1–68. Cambridge University Press. [SDG, RMK, arMJP]
Giles, H. & Powesland, P. F. (1975) *Speech styles and social evaluation.* Academic Press. [aMJP]
Ginzburg, J. (1996) Interrogatives: Questions, facts, and dialogue. In: *Handbook of contemporary semantic theory,* ed. S. Lappin. Blackwell. [JG]
(1997a) On some semantic consequences of turn taking. In: *Proceedings of the 11th Amsterdam Colloquium on Formal Semantics and Logic,* ed. P. Dekker, M. Stokhof & Y. Venema, pp. 145–50. Institute for Logic, Language, and Computation (ILLC). Available at: http://www.dcs.kcl.ac.uk/staff/ginzburg. [JG]
(1997b) Structural mismatch in dialogue. In: *Proceedings of MunDial 97* (Technical Report 97–106), ed. G. Jaeger & A. Benz, pp. 59–80. Universität München Centrum für Informations- und Sprachverarbeitung. Available at: http://www.dcs.kcl.ac.uk/staff/ginzburg. [JG]
(1999) Ellipsis resolution with syntactic presuppositions. In: *Computing meaning 1: Current issues in computational semantics,* ed. H. Bunt & R. Muskens. Kluwer. [aMJP]
(2001) Fragmenting meaning: Clarification ellipsis and nominal anaphora. In: *Computing meaning 2: Current issues in computational semantics,* ed. H. Bunt. Kluwer. [aMJP]
(2003) Disentangling public from private meaning. In: *Current and new directions in discourse and dialogue,* ed. J. van Kuppevelt & R. Smith, pp. 183–212. Kluwer. Available at: http://www.dcs.kcl.ac.uk/staff/ginzburg. [JG]
(forthcoming) *Semantics and interaction in dialogue.* CSLI Publications and University of Chicago Press. Draft chapters available at: http://www.dcs.kcl.ac.uk/staff/ginzburg. [JG]
Ginzburg, J. & Cooper, R. (2004) Clarification, ellipsis, and the nature of contextual updates. *Linguistics and Philosophy* 27(3):297–365. Available at: http://www.dcs.kcl.ac.uk/staff/ginzburg. [JG]
Ginzburg, J. & Sag, I. A. (2001) *Interrogative investigations.* CSLI. [aMJP]
Goldberg, A. (1995) *Constructions.* University of Chicago Press. [PFD]
Goldinger, S. D. (1998) Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105:251–79. [SDG, MK, RMK, arMJP]
Goldinger, S. D. & Azuma, T. (in press) Puzzle-solving science: The quixotic quest for units in speech perception. *Journal of Phonetics.* [SDG]
Goodwin, C. (1979) The interactive construction of a sentence in natural conversation. In: *Everyday language: Studies in ethnomethodology,* ed. G. Psathas, pp. 97–121. Irvington. [EAS]
Graesser, A. C., Weimer-Hastings, P. & Weimer-Hastings, K. (2001) Constructing inferences and relations during text comprehension. In: *Text representation: Linguistic and psycholinguistic aspects,* ed. T. Sanders, J. Schilperoord & W. Spooren, pp. 249–72. Benjamins. [ABM]
Gregory, S. W & Webster, S. (1996) A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status. *Journal of Personality and Social Psychology* 70:1231–40. [RMK]
Grossberg, S. (1980) How does a brain build a cognitive code? *Psychological Review* 87:1–51. [SDG, rMJP]
(1999) The link between brain learning, attention, and consciousness. *Consciousness and Cognition* 8:1–44. [SDG]
Grossberg, S., Boardman, I. & Cohen, M. (1997) Neural dynamics of variable-rate speech categorization. *Journal of Experimental Psychology: Human Perception and Performance* 23:483–503. [SDG]
Grossberg, S. & Myers, C. W. (2000) The resonant dynamics of speech perception: Interword integration and duration-dependent backward effects. *Psychological Review* 107:735–67. [SDG]
Grossberg, S. & Stone, G. O. (1986) Neural dynamics of word recognition and recall: Priming, learning, and resonance. *Psychological Review* 93:46–74. [SDG]
Gunlogson, C. (2003) *True to form: Rising and falling declaratives as questions in English.* Routledge. [SB-S]
Halpern, Y. & Moses, Y. (1990) Knowledge and common knowledge in a distributed environment. *Journal of the ACM* 37:549–87. [aMJP]
Hamblin, C. L. (1970) *Fallacies.* Methuen. [JG]
Hanna, J. E. & Tanenhaus, M. K. (2004) Pragmatic effects on reference resolution in a collaborative task: Evidence from eye movements. *Cognitive Science* 28:105–15. [SEB]
Hanna, J. E., Tanenhaus, M. K. & Trueswell, J. C. (2003) The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language* 49(1):43–61. [SEB, arMJP, SB-S]
Hartsuiker, R. J. & Kolk, H. H. J. (2001) Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology* 42:113–57. [aMJP]
Hartsuiker, R. J. & Westenberg, C. (2000) Persistence of word order in written and spoken sentence production. *Cognition* 75:B27–39. [arMJP]
Hartsuiker, R. J., Pickering, M. J. & Veltkamp, E. (2004) Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish/English bilinguals. *Psychological Science* 15:409–14. [rMJP]
Haviland, S. E. & Clark, H. H. (1974) What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior* 13:512–21. [aMJP]
Healey, P. G. T. (1997) Expertise or expertese?: The emergence of task-oriented sub-languages. In: *Proceedings of the 19th Annual Conference of the Cognitive Science Society, 7th–10th August, Stanford University, California,* ed. M. G. Shafto & P. Langley, pp. 301–306. Erlbaum. [PGTH]
(in preparation) Semantic co-ordination in dialogue: Communication as a special case of misunderstanding. Unpublished manuscript. [PGTH]
Heyes, C. M. (2001) Causes and consequences of imitation. *Trends in Cognitive Sciences* 5:253–61. [aMJP]
Hintzman, D. L. (1986) "Schema-abstraction" in a multiple trace model. *Psychological Review* 93:411–28. [MK]
Holtgraves, T. (1994) Communication in context: Effects of speaker status on the comprehension of indirect requests. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20:1205–18. [HS]
Hommel, B., Müsseler, J., Aschersleben, G. & Prinz, W. (2001) The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences* 24:849–937. [JCC, arMJP]
Horton, W. S. & Keysar, B. (1996) When do speakers take into account common ground? *Cognition* 59:91–117. [SEB, arMJP]
Hurley, S. & Chater, N., eds. (in press) *Perspectives on imitation: From mirror neurons to memes, vols. 1 & 2.* MIT Press. [arMJP]
Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C. & Rizzolatti, G. (1999) Cortical mechanisms of human imitation. *Science* 286:2526–28. [aMJP]
Isaacs, E. A. & Clark, H. H. (1987) References in conversations between experts and novices. *Journal of Experimental Psychology: General* 116:26–37. [rMJP]
Jackendoff, R. (1997) *The architecture of the language faculty.* MIT Press. [aMJP]
(1999) Parallel constraint-based generative theories of language. *Trends in Cognitive Sciences* 3:393–400. [aMJP]
(2002) *Foundations of language.* Oxford University Press. [arMJP]
Jansma, B. M. & Schiller, N. O. (2004) Monitoring syllable boundaries during speech production. *Brain and Language* 90:311–17. [NOS]
Jefferson, G. (1974) Error correction as an interactional resource. *Language in Society* 2:181–99. [EAS]
(1980) On "trouble-premonitory" response to inquiry. *Sociological Inquiry* 50(3–4):153–85. [EAS]
(1986) Notes on "latency" in overlap onset. *Human Studies* 9:153–83. [EAS]
(1987) On exposed and embedded corrections in conversation. In: *Talk and social organization,* ed. G. Button & J. R. E. Lee. Multilingual Matters. [aMJP, EAS]
Johnson, K., Strand, E. A. & D'Imperio, M. (1999) Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics* 27(4):359–84. [HS]
Johnson-Laird, P. N. (1983) *Mental models: Toward a cognitive science of language, inference and consciousness.* Harvard University Press. [RK, arMJP]
Just, M. A. & Carpenter, P. A. (1980) A theory of reading: From eye fixations to comprehension. *Psychological Review* 87:329–54. [aMJP]
Kaplan, R. & Bresnan, J. (1982) Lexical-functional grammar: A formal system for grammatical representation. In: *The mental representation of grammatical relations,* ed. J. Bresnan. MIT Press. [aMJP]
Kaschak, M. P. (2003) This syntax needs learned: Adult acquisition of novel syntactic constructions. Unpublished doctoral dissertation, University of Wisconsin–Madison. [MK]
Kellas, G. & Vu, H. (1999) Strength of context does modulate the subordinate bias effect: A reply to Binder and Rayner. *Psychonomic Bulletin and Review* 6:511–17. [aMJP]
Kempen, G. & Huijbers, P. (1983) The lexicalization process in sentence production and naming: Indirect election of words. *Cognition* 14:824–43. [aMJP]
Kempson, R., Meyer-Viol, W. & Gabbay, D. (2001) *Dynamic syntax.* Blackwell. [RK, aMJP]
Keysar, B. (1994) The illusory transparency of intention: Linguistic perspective taking in text. *Cognitive Psychology* 26:165–208. [ABM]
(1997) Unconfounding common ground. *Discourse Processes* 24:253–70. [DJB, SEB]
Keysar, B., Barr, D. J. & Balin, J. A. (1998) Definite reference and mutual knowledge: Process models of common ground in comprehension. *Journal of Memory and Language* 39:1–20. [aMJP, ABM]
Keysar, B., Barr, D. J., Balin, J. A. & Brauner, J. S. (2000) Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science* 11:32–38. [aMJP]
Keysar, B., Lin, S. H. & Barr, D. J. (2003) Limits on theory of mind use in adults. *Cognition* 89:25–41. [rMJP]
Kingsbury, D. (1968) Manipulating the amount of information obtained from a person giving directions. Unpublished honors thesis, Department of Social Relations, Harvard University. [RMK, rMJP]

Krauss, R. M. & Fussell, S. R. (1996) Social psychological models of interpersonal communication. In: *Social psychology: A handbook of basic principles,* ed. E. T. Higgins & A. Kruglanski, pp. 655–701. Guilford. [RMK]

Kraut, R. E., Fussell, S. R. & Siegel, J. (2003) Visual information as a conversational resource in collaborative physical tasks. *Human-computer interaction* 18:13–49. [SRF, rMJP]

Kraut, R. E., Gergle, D. & Fussell, S. R. (2002) The use of visual information in shared visual spaces: Informing the development of virtual co-presence. In: *Proceedings of the CSCW 2002 Conference on Computer-Supported Cooperative Work.* ACM Press. [SRF]

Kuiper, K. (1996) *Smooth talkers: The linguistic performance of auctioneers and sportscasters.* Erlbaum. [arMJP]

Lakin, J. L. & Chartrand, T. L. (2003) Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychological Science* 14:334–39. [rMJP]

Larsson, S. (2002) *Issue based dialogue management.* Doctoral dissertation, Gothenburg University. [JG]

Lee, B. P. H. (2001) Mutual knowledge, background knowledge and shared beliefs: Their roles in establishing common ground. *Journal of Pragmatics* 33:21–44. [DJB]

Lerner, G. H. (1991) On the syntax of sentences-in-progress. *Language in Society* 20:441–58. [EAS]

(1996) On the "semi-permeable" character of grammatical units in conversation: Conditional entry into the turn space of another speaker. In: *Interaction and grammar,* ed. E. Ochs, E. A. Schegloff & S. A. Thompson, pp. 238–76. Cambridge University Press. [EAS]

Levelt, W. J. M. (1983) Monitoring and self-repair in speech. *Cognition* 14:41–104. [aMJP]

(1989) *Speaking: From intention to articulation.* MIT Press. [DJB,JCC,SG,aMJP,NOS]

Levelt, W. J. M. & Kelter, S. (1982) Surface form and memory in question answering. *Cognitive Psychology* 14:78–106. [arMJP]

Levelt, W. J. M. & Maassen, B. (1981) Lexical search and order of mention in sentence production. In: *Crossing the boundaries in linguistics: Studies presented to Manfred Bierwisch,* ed. W. Klein & W. J. M. Levelt. Riedel. [aMJP]

Levelt, W. J. M., Roelofs, A. & Meyer, A. S. (1999) A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22:1–75. [JCC,SDG,aMJP,NOS]

Lewis, D. K. (1969) *Convention: A philosophical study.* Basil Blackwell/Harvard University Press. [PGTH,aMJP]

(1979) Score keeping in a language game. In: *Semantics from different points of view,* ed. R. Bauerle, pp. 172–87. Springer. [JG]

Liberman, A. M., Cooper, F. S., Shankweiler, D. P. & Studdert-Kennedy, M. (1967) Perception of the speech code. *Psychological Review* 74:431–61. [HS]

Liberman, A. M. & Whalen, D. H. (2000) On the relation of speech to language. *Trends in Cognitive Sciences* 4:187–96. [aMJP]

Lindblom, B. (1990) Explaining variation: A sketch of the H and H theory. In: *Speech production and speech modeling,* ed. W. Hardcastle & A. Marchal. Kluwer. [aMJP]

Linnell, P. (1998) *Approaching dialogue: Talk, interaction, and contexts in a dialogical perspective.* Benjamins. [aMJP]

Lockridge, C. B. & Brennan, S. E. (2002) Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin and Review* 9:550–57. [SEB,arMJP]

Lombardi, L. & Potter, M. C. (1992) The regeneration of syntax in short term memory. *Journal of Memory and Language* 31:713–33. [aMJP]

MacDonald, M. C., Pearlmutter, N. J. & Seidenberg, M. S. (1994) Lexical nature of syntactic ambiguity resolution. *Psychological Review* 101:676–703. [MK]

MacKay, D. (1987) *The organization of perception and action.* Springer. [aMJP]

Magnuson, J. & Nusbaum, H. C. (1994) Some acoustic and nonacoustic conditions that produce talker normalization. In: *Proceedings of the Spring Meeting of the Acoustical Society of Japan, Tokyo, March 1994,* pp. 637–38. Acoustical Society of Japan. [HS]

Markman, A. B. (1997) Constraints on analogical inference. *Cognitive Science* 21(4):373–418. [ABM]

Markman, A. B. & Gentner, D. (1993) Structural alignment during similarity comparisons. *Cognitive Psychology* 25:431–67. [aMJP]

(2000) Structure mapping in the comparison process. *American Journal of Psychology* 113(4):501–38. [ABM]

Markman, A. B. & Makin, V. S. (1998) Referential communication and category acquisition. *Journal of Experimental Psychology: General* 127:331–54. [aMJP]

Marslen-Wilson, W. D. (1973) Linguistic structure and speech shadowing at very short latencies. *Nature* 244:522–23. [aMJP]

Martin, R. C., Lesch, M. F. & Bartha, M. C. (1999) Independence of input and output phonology in word processing and short-term memory. *Journal of Memory and Language* 41:3–29. [JCC]

Mattingly, I. G. & Liberman, A. M. (1988) Specialized perceiving systems for speech and other biologically significant sounds. In: *Auditory function,* ed. G. M. Edelman. Wiley. [aMJP]

Mazur, A. (1985) A biosocial model of status in face-to-face primate groups. *Social Forces* 64:377–402. [AM]

McCarthy, J. (1990) Formalization of two puzzles involving knowledge. In: *Formalizing common sense: Papers by John McCarthy,* ed. V. Lifschitz. Ablex. [aMJP]

McClelland, J. L. & Elman, J. L. (1986) The TRACE model of speech perception. *Cognitive Psychology* 18:1–86. [SDG]

McClelland, J. L., St. John, M. & Taraban, R. (1989) Sentence comprehension: A parallel distributed processing approach. *Language and Cognitive Processes* 4:287–336. [SDG]

McDaniel, D. & Cowart, W. (1999) Experimental evidence for a minimalist account of English resumptive pronouns. *Cognition* 70:B15–B24. [FF]

McKoon, G. & Ratcliff, R. (1992) Inference during reading. *Psychological Review* 99:440–66. [ABM]

Meltzoff, A. N. & Decety, J. (2003) What imitation tells us about social cognition: A rapprochement between developmental psychology and cognitive neuroscience. *Philosophical Transactions of the Royal Society of London* 358:491–500. [rMJP]

Metzing, C. & Brennan, S. E. (2001) When conceptual pacts are broken: Partner effects in the comprehension of referring expressions. In: *Abstracts of the Psychonomic Society, 42nd Annual Meeting, Orlando, FL, November 2001,* p. 29. Psychonomic Society. [SEB]

(2003) When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language* 49:201–13. [SEB, rMJP]

Meyer, A. S. (1996) Lexical access in phrase and sentence production: Results from picture-word interference experiments. *Journal of Memory and Language* 35:477–96. [aMJP]

Moore, J. (1995) *Participating in explanatory dialogues.* Bradford Books/MIT Press. [JG]

Morales, M., Mundy, P., Delgado, C. E. F., Yale, M., Messinger, D., Neal, R. & Schwartz, H. K. (2000) Responding to joint attention across the 6- through 24-month age period and early language acquisition. *Journal of Applied Developmental Psychology* 21(3):283–98. [PFD]

Morgan, J. L. (1973) Sentence fragments and the notion "sentence". In: *Issues in linguistics: Papers in honor of Henry and Renée Kahane,* ed. B. B. Kachru, R. B. Lees, Y. Malkiel, A. Pietrangeli & S. Saporta. University of Illinois Press. [aMJP]

Morgan, J. L. & Wheeldon, L. R. (2003) Syllable monitoring in internally and externally generated English words. *Journal of Psycholinguistic Research* 32:269–96. [NOS]

Morton, J. (1969) Interaction of information in word recognition. *Psychological Review* 76:165–78. [aMJP]

Moss, H. E. & Gaskell, G. M. (1999) Lexical semantic processing during speech. In: *Language processing,* ed. S. Garrod & M. Pickering. Psychology Press. [aMJP]

Mullennix, J. W. & Pisoni, D. B. (1990) Stimulus variability and processing dependencies in speech perception. *Perception and Psychophysics* 47(4):379–90. [HS]

Müller, O. & Hagoort, P. (2001) Semantic and syntactic properties of a word – which are first in reading? In: *Abstracts of the Annual Meeting of the Cognitive Neuroscience Society,* p. 124. Cognitive Neuroscience Society. [NOS]

Muller-Gass, A., Gonthier, I., Desrochers, A. & Campbell, K. B. (2000) Multiple P3 evidence of a two-stage process in word gender decision. *NeuroReport* 16:3527–31. [NOS]

Murphy, G. L. (1988) Comprehending complex concepts. *Cognitive Science* 12:529–62. [aMJP]

Myers, J. L. & O'Brien, E. J. (1998) Accessing the discourse representation during reading. *Discourse Processes* 26:131–57. [MK]

Nadig, J. S. & Sedivy, J. C. (2002) Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science* 13:329–36. [SEB, arMJP]

Natale, M. (1975a) Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology* 32:790–804. [RMK]

(1975b) Social desirability as related to convergence of temporal speech patterns. *Perceptual and Motor Skills* 40:827–39. [RMK]

Nearey, T. M. (1989) Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America* 85(5):2088–113. [HS]

Newtson, D. (1994) The perception and coupling of behavior waves. In: *Dynamical systems in social psychology,* ed. R. Vallacher & A. Nowak, pp. 139–67. Academic Press. [SDG]

Niedzielski, N. (1999) The effect of social information on the perception of

sociolinguistic variables. *Journal of Language and Social Psychology* 18(1):62–85. [HS]
Nunberg, G., Sag, I. A. & Wasow, T. (1994) Idioms. *Language* 70:491–538. [aMJP]
Nusbaum, H. C. & Henly, A. S. (1989) Understanding speech from the perspective of cognitive psychology. Paper presented at the Workshop on Spoken Language Understanding, Department of Psychology, State University of New York at Buffalo, 1989. [HS]
Nusbaum, H. C. & Magnuson, J. (1997) Talker normalization. In: *Talker variability in speech processing,* ed. K. Johnson & J. W. Mullennix. Academic Press. [HS]
Nusbaum, H. C. & Morin, T. M. (1992) Paying attention to differences among talkers. In: *Speech production, perception, and linguistic structure,* ed. Y. Tohkura, Y. Sagisaka & E. Vatikiotis-Bateson, pp. 113–34. Ohmasha Publishing. [HS]
Nusbaum, H. C. & Schwab, E. C. (1986) The role of attention and active processing in speech perception. In: *Pattern recognition by humans and machines: Speech perception, vol. 1,* ed. H. C. Nusbaum & E. C. Schwab, pp. 113–57. Academic Press. [HS]
Oomen, C. C. E. & Postma, A. (2002) Limitations in processing resources and speech monitoring. *Language and Cognitive Processes* 17:163–84. [aMJP]
Peterson, G. & Barney, H. (1952) Control methods used in the study of vowels. *Journal of the Acoustical Society of America* 24:175–84. [HS]
Phillips, C. (1996) Order and structure. Unpublished doctoral dissertation, MIT. [RK]
(2003) Linear order and constituency. *Linguistic Inquiry* 34:37–90. [aMJP]
Pickering, M. & Barry, G. (1991) Sentence processing without empty categories. *Language and Cognitive Processes* 6:229–59. [aMJP]
(1993) Dependency categorial grammar and coordination. *Linguistics* 31:855–902. [aMJP]
Pickering, M. J. & Branigan, H. P. (1998) The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language* 39:633–51. [MK, arMJP, NOS]
(1999) Syntactic priming in language production. *Trends in Cognitive Sciences* 3:136–41. [aMJP]
Pickering, M., Clifton, C. & Crocker, M. W. (2000) Architectures and mechanisms in sentence comprehension. In: *Architectures and mechanisms for language processing,* ed. M. W. Crocker & M. Pickering. Cambridge University Press. [JCC]
Pickering, M. J. & Garrod, S. (in press) Routinization in the interactive-alignment model of dialogue. In: *Twenty-first century psycholinguistics: Four cornerstones,* ed. A. Cutler. Erlbaum. [rMJP]
Pinker, S. (1989) *Learnability and cognition: The acquisition of argument structure.* MIT Press/Bradford Books. [HPB]
Pinker, S. & Birdsong, D. (1979) Speakers' sensitivity to rules of frozen word order. *Journal of Verbal Learning and Verbal Behavior* 18:497–508. [aMJP]
Poesio, M. & Traum, D. R. (1997) Conversational actions and discourse situations. *Computational Intelligence* 13:309–47. [aMJP]
Pollard, C. & Sag, I. A. (1994) *Head-driven phrase structure grammar.* University of Chicago Press and CSLI. [aMJP]
Postma, A. (2000) Detection of errors during speech production: A review of speech monitoring models. *Cognition* 77:97–131. [aMJP]
Potter, M. C. & Lombardi, L. (1990) Regeneration in the short-term recall of sentences. *Journal of Memory and Language* 29:633–54. [aMJP]
(1998) Syntactic priming in immediate recall of sentences. *Journal of Memory and Language* 38:265–82. [aMJP]
Prat-Sala, M. & Branigan, H. P. (2000) Discourse constraints on syntactic processing in language production: A cross-linguistic study in English and Spanish. *Journal of Memory and Language* 42:168–82. [aMJP]
Prince, E. F. (1990) Syntax and discourse: A look at resumptive pronouns. In: *Proceedings of the Sixteenth Annual Meeting of the Berkeley Linguistics Society,* ed. University of California at Berkeley Press, pp. 482–97. [FF]
Purver, M., Ginzburg, J. & Healey, P. (2002) On the means for clarification in dialogue. In: *Advances in discourse and dialogue,* ed. J. van Kuppevelt & R. Smith. Kluwer. Available at: http://www.dcs.kcl.ac.uk/pg/purver/publications.html. [JG]
Purver, M., Healey, P. G. T., King, J., Ginzburg, J. & Mills, G. J. (2003) Answering clarification questions. In: *Proceedings of the 4th SIGdial Workshop on Discourse and Dialogue.* Sapporo. Available at: http://www.dcs.kcl.ac.uk/pg/purver/publications.html. [JG]
Purver, M. & Otsuka, M. (2003) Incremental generation by incremental parsing: Tactical generation in Dynamic Syntax. In: *Proceedings of 9th EACL (European Association for Computational Linguistics) Workshop in Natural Language Generation, Budapest, April 2003,* ed. E. Reiter, H. Horecek & K. van Deemter, pp. 79–86. Association for Computational Linguistics. [RK]
Rayner, K., Pacht, J. M. & Duffy, S. A. (1994) Effects of prior encounter and discourse bias on the processing of lexically ambiguous words. *Journal of Memory and Language* 33:527–44. [aMJP]
Reeves, B. & Nass, C. (1996) *The media equation: How people treat computers, television, and new media like real people and places.* Cambridge University Press. [aMJP]
Rizzolatti, G. & Arbib, M. A. (1998) Language within our grasp. *Trends in Neurosciences* 21:188–94. [aMJP]
Ross, J. R. (1969) Guess who? In: *Papers from the Fifth Regional Meeting of the Chicago Linguistics Society,* ed. R. I. Binnick, A. Davison, G. M. Green & J. L. Morgan. University of Chicago. [aMJP]
Russell, A. W. & Schober, M. F. (1999) How beliefs about a partner's goals affect referring in goal-discrepant conversations. *Discourse Processes* 27(1):1–33. [MFS]
Sacks, H. (1972a) An initial investigation of the usability of conversational data for doing sociology. In: *Studies in social interaction,* ed. D. N. Sudnow, pp. 31–74. Free Press. [EAS]
(1972b) On the analyzability of stories by children. In: *Directions in sociolinguistics: The ethnography of communication,* ed. J. J. Gumperz & D. Hymes, pp. 325–55. Holt, Rinehart and Winston. [EAS]
(1975) Everyone has to lie. In: *Sociocultural dimensions of language use,* ed. M. Sanches & B. G. Blount, pp. 57–80. Academic Press. [EAS]
(1987) On the preferences for agreement and contiguity in sequences in conversation. In: *Talk and social organization,* ed. G. Button & J. R. E. Lee. Multilingual Matters. [aMJP]
(1992) *Lectures on conversation,* ed. G. Jefferson, introduction by E. A. Schegloff (2 vols.). Blackwell. [EAS]
Sacks, H. & Schegloff, E. A. (1979) Two preferences in the organization of reference to persons and their interaction. In: *Everyday language: Studies in ethnomethodology,* ed. G. Psathas, pp. 15–21. Irvington. [EAS]
Sacks, H., Schegloff, A. E. & Jefferson, G. (1974) A simplest systematics for the organization of turn-taking for conversation. *Language* 50:696–735. [DJB, aMJP, EAS]
Sancier, M. & Fowler, C. A. (1997) Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics* 25:421–36. [RMK]
Sanford, A. J. & Garrod, S. C. (1981) *Understanding written language.* Wiley. [aMJP]
Sanford, A. J. & Sturt, P. (2002) Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Sciences* 6:382–86. [rMJP]
Schegloff, E. A. (1972) Notes on a conversational practice: Formulating place. In: *Studies in social interaction,* ed. D. N. Sudnow, pp. 75–119. Free Press. [EAS]
(1979) The relevance of repair for syntax-for-conversation. In: *Syntax and semantics 12: Discourse and syntax,* ed. T. Givon, pp. 261–88. Academic Press. [EAS]
(1986) The routine as achievement. *Human Studies* 9:111–51. [EAS]
(1987a) Recycled turn beginnings: A precise repair mechanism in conversation's turn-taking organisation. In: *Talk and social organisation,* ed. G. Button & J. R. E. Lee, pp. 70–85. Multilingual Matters. [EAS]
(1987b) Some sources of misunderstanding in talk-in-interaction. *Linguistics* 25:201–18. [EAS]
(1989) Reflections on language, development, and the interactional character of talk-in-interaction. In: *Interaction in human development,* ed. M. Bornstein & J. S. Bruner, pp. 139–53. Erlbaum. [EAS]
(1990) On the organization of sequences as a source of "coherence" in talk-in-interaction. In: *Conversational organization and its development,* ed. B. Dorval, pp. 51–77. Ablex. [EAS]
(1991) Conversation analysis and socially shared cognition. In: *Perspectives on socially shared cognition,* ed. L. Resnick, J. Levine & S. Teasley, pp. 151–71. American Psychological Association. [EAS]
(1992) Repair after next turn: The last structurally provided place for the defence of intersubjectivity in conversation. *American Journal of Sociology* 95(5):1295–345. [EAS]
(1995) Sequence organization. Unpublished manuscript, Department of Sociology, University of California at Los Angeles. [EAS]
(1996a) Turn organization: One intersection of grammar and interaction. In: *Interaction and grammar,* ed. E. Ochs, E. A. Schegloff & S. A. Thompson, pp. 52–133. Cambridge University Press. [EAS]
(1996b) Some practices for referring to persons in talk-in-interaction: A partial sketch of a systematics. In: *Studies in anaphora,* ed. B. A. Fox, pp. 437–85. John Benjamins. [EAS]
(1997a) Practices and actions: Boundary cases of other-initiated repair. *Discourse Processes* 23:499–545. [EAS]
(1997b) Third turn repair. In: *Towards a social science of language: Papers in honor of William Labov, vol. 2: Social interaction and discourse structures,* ed. G. R. Guy, C. Feagin, D. Schiffrin & J. Baugh, pp. 31–40. John Benjamins. [EAS]
(2000a) Overlapping talk and the organization of turn-taking for conversation. *Language in Society* 29(1):1–63. [EAS]
(2000b) When "others" initiate repair. *Applied Linguistics* 21(2):205–43. [EAS]
(2001) Accounts of conduct in interaction: Interruption, overlap and turn-taking.

In: *Handbook of sociological theory,* ed. J. H. Turner, pp. 287–321. Plenum Press. [EAS]
(forthcoming) *A primer in conversation analysis: Sequence organization.* Cambridge University Press. [EAS]
Schegloff, E. A., Jefferson, G. & Sacks, H. (1977) The preference for self-correction in the organization of repair in conversation. *Language* 53(2):361–82. [EAS]
Schegloff, E. A. & Sacks, H. (1973) Opening up closings. *Semiotica* 8:289–327. [aMJP, EAS]
Schenkein, J. (1980) A taxonomy of repeating action sequences in natural conversation. In: *Language production, vol. 1,* ed. B. Butterworth, pp. 21–47. Academic Press. [aMJP]
Schiffer, S. R. (1972) *Meaning.* Oxford University Press. [aMJP]
Schiller, N. O. (2001) Metrical encoding during speech production. *Abstracts of the Psychonomic Society* 6:29. [NOS]
(in press) Verbal self-monitoring. In: *Twenty-first century psycholinguistics: Four cornerstones,* ed. A. Cutler. Erlbaum. [NOS]
Schiller, N. O. & Caramazza, A. (2003) Grammatical feature selection in noun phrase production: Evidence from German and Dutch. *Journal of Memory and Language* 48:169–94. [rMJP, NOS]
Schiller, N. O., Jansma, B. M., Peters, J. & Levelt, W. J. M. (in press) Monitoring metrical stress in polysyllabic words. *Language and Cognitive Processes*. [NOS]
Schiller, N. O., Münte, T., Horemans, I. & Jansma, B. M. (2003) Semantic and phonological factors in syntactic decisions. *Psychophysiology* 40:869–77. [NOS]
Schmitt, B. M., Rodriguez-Fornells, A., Kutas, M. & Münte, T. F. (2001a) Electrophysiological estimates of semantic and syntactic information access during tacit picture naming and listening to words. *Neuroscience Research* 41:293–98. [NOS]
Schmitt, B. M., Schiltz, K., Zaake, W., Kutas, M. & Münte, T. F. (2001b) An electrophysiological analysis of the time course of conceptual and syntactic encoding during tacit picture naming. *Journal of Cognitive Neuroscience* 13:510–22. [NOS]
Schober, M. F. (1993) Spatial perspective-taking in conversation. *Cognition* 47:1–24. [aMJP]
(1995) Speakers, addressees, and frames of reference: Whose effort is minimized in conversations about location? *Discourse Processes* 20(2):219–47. [MFS]
(1998a) Different kinds of conversational perspective-taking. In: *Social and cognitive psychological approaches to interpersonal communication,* ed. S. R. Fussell & R. J. Kreuz, pp. 145–74. Erlbaum. [MFS]
(1998b) How partners with high and low spatial ability choose perspectives in conversation. *Abstracts of the 39th Annual Meeting of the Psychonomic Society,* No. 39. [MFS]
Schober, M. F. & Brennan, S. E. (2003) Processes of interactive spoken discourse: The role of the partner. In: *Handbook of discourse processes,* ed. A. C. Graesser, M. A. Gernsbacher & S. R. Goldman, pp. 123–64. Erlbaum. [SEB, aMJP, MFS]
Schober, M. F. & Clark, H. H. (1989) Understanding by addressees and overhearers. *Cognitive Psychology* 21:211–32. [DJB, aMJP]
Schober, M. F. & Conrad, F. G. (1997) Does conversational interviewing reduce survey measurement error? *Public Opinion Quarterly* 61:576–602. Reprinted in: *Interviewing, vol. 1,* ed. N. G. Fielding. SAGE Benchmarks in Social Science Research Series, 2003. Sage Publications. [MFS]
Schober, M. F., Conrad, F. G. & Fricker, S. S. (2004) Misunderstanding standardized language in research interviews. *Applied Cognitive Psychology* 18:169–88. [MFS]
Schriefers, H., Meyer, A. S. & Levelt, W. J. M. (1990) Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language* 29:86–102. [aMJP]
Schuller, A. M. & Rossion, B. (2001) Spatial attention triggered by eye gaze increases and speeds up early visual acuity. *NeuroReport* 12:2381–86. [rMJP]
Searle, J. (1969) *Speech acts.* Cambridge University Press. [JG]
Sereno, S. C., Brewer, C. & O'Donnell, P. J. (2003) Context effects in word recognition: Evidence for early interactive processing. *Psychological Science* 14:328–33. [SG]
Sheldon, A. (1974) The role of parallel function in the acquisition of relative clauses in English. *Journal of Verbal Learning and Verbal Behavior* 13:272–81. [aMJP]
Shiffrin, R. M. & Schneider, W. (1977) Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review* 84(2):127–90. [HS]
Shockley, K., Santana, M-V. & Fowler, C. A. (2003) Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance* 29:326–32. [SDG]
Skinner, B. F. (1953) *Science and human behavior.* Macmillan. [JJP]
(1957) *Verbal behavior.* Appleton-Century-Crofts. [JJP, rMJP]
Smith, M. & Wheeldon, L. (2001) Syntactic priming in spoken sentence production – an online study. *Cognition* 78:123–64. [aMJP]
Smyth, R. (1994) Grammatical determinants of ambiguous pronoun resolution. *Journal of Psycholinguistic Research* 23:197–229. [aMJP]
Sperber, D. & Wilson, D. (1986) *Relevance.* Blackwell. [ABM]
Stalnaker, R. C. (1978) Assertion. In: *Syntax and semantics, vol. 9: Pragmatics,* ed. P. Cole, pp. 315–32. Academic Press. [JG, aMJP]
Steedman, M. (2000) *The syntactic process.* MIT Press. [aMJP]
Suessbrick, A. L., Schober, M. F. & Conrad, F. G. (2000) Different respondents interpret ordinary questions quite differently. In: *Proceedings of the American Statistical Association* (Section on *Survey Research Methods*). American Statistical Association. [MFS]
Swinney, D. A. (1979) Lexical access during sentence comprehension. *Journal of Verbal Learning and Verbal Behavior* 18:645–59. [aMJP]
Tanenhaus, M. K., Chambers, C. C. & Hanna, J. E. (2004) Referential domains in spoken language comprehension: Using eye movements to bridge the product and action traditions. In: *The interface of language, vision, and action: Eye movements and the visual world,* ed. J. M. Henderson & F. Ferreira. Psychology Press. [SB-S]
Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M. & Sedivy, J. C. (1995) Integration of visual and linguistic information in spoken language comprehension. *Science* 268(5217):632–34. [aMJP, HS]
Tannen, D. (1989) *Talking voices: Repetition, dialogue, and imagery in conversational discourse.* Cambridge University Press. [aMJP]
Thakerar, J. N., Giles, H. & Chesire, J. (1982) Psychological and linguistic parameters of speech accommodation theory. In: *Advances in the social psychology of language,* ed. C. Fraser & K. R. Scherer, pp. 205–255. Cambridge University Press. [HS]
Tomasello, M. (2000) Do young children have adult syntactic competence? *Cognition* 74:209–53. [HPB, aMJP]
(2003) *Constructing a language: A usage-based theory of language acquisition.* Harvard University Press. [PFD, rMJP]
Traxler, M. J. & Gernsbacher, M. A. (1992) Improving written communication through minimal feedback. *Language and Cognitive Processes* 7:1–22. [aMJP]
(1993) Improving written communication through perspective-taking. *Language and Cognitive Processes* 8:311–36. [aMJP]
Tucker, G. R., Lambert, W. E. & Rigault, A. A. (1977) *The French speaker's skill with grammatical gender: An example of rule-governed behavior.* Mouton. [NOS]
van Dijk, T. A. & Kintsch, W. (1983) *Strategies in discourse comprehension.* Academic Press. [aMJP]
van Gelder, T. & Port, R. F. (1995) It's about time: An overview of the dynamical approach to cognition. In: *Mind as motion,* ed. R. F. Port & T. van Gelder, pp. 1–44. MIT Press. [SDG]
van Turennout, M., Hagoort, P. & Brown, C. M. (1998) Brain activity during speaking: From syntax to phonology in 40 milliseconds. *Science* 280:572–74. [aMJP, NOS]
Vigliocco, G., Antonini, T. & Garrett, M. F. (1997) Grammatical gender is on the tip of Italian tongues. *Psychological Science* 8:314–17. [arMJP]
Vitevitch, M. S. & Luce, P. A. (1999) Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language* 40:374–408. [SDG]
Wheeldon, L. & Levelt, W. J. M. (1995) Monitoring the time course of phonological encoding. *Journal of Memory and Language* 34:311–34. [aMJP, NOS]
Wheeldon, L. & Morgan, J. L. (2002) Phoneme monitoring in internal and external speech. *Language and Cognitive Processes* 17:503–35. [NOS]
Wheeldon, L. R. & Smith, M. C. (2003) Phrase structure priming: A short-lived effect. *Language and Cognitive Processes* 18:431–42. [rMJP]
Wilkes-Gibbs, D. & Clark, H. H. (1992) Coordinating beliefs in conversation. *Journal of Memory and Language* 31:183–94. [DJB, aMJP]
Williams, J. H. G., Whiten, A., Suddendorf, T. & Perrett, D. I. (2001) Imitation, mirror neurons, and autism. *Neuroscience and Biobehavioral Reviews* 25:287–95. [aMJP]
Wisniewski, E. L. (1996) Construal and similarity in conceptual combination. *Journal of Memory and Language* 35:434–53. [aMJP]
Wray, A. & Perkins, M. R. (2001) The functions of formulaic language: An integrated model. *Language and Communication* 20:1–28. [aMJP]
Zwaan, R. A. & Radvansky, G. A. (1998) Situation models in language comprehension and memory. *Psychological Bulletin* 123:162–85. [aMJP]
Zwitserlood, P. (1994) Access to phonological-form representations in language comprehension and production. In: *Perspectives on sentence processing,* ed. C. Clifton, L. Frazier & K. Rayner, pp. 83–106. Erlbaum. [JCC]