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,

https://doi.org/10.1017/S0140525X0436005X Published online by Cambridge University Press

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-

https://doi.org/10.1017/S0140525X0436005X Published online by Cambridge University Press