

Discussion

Speaking with common ground: from principles to processes in pragmatics: a reply to Polichak and Gerrig

Boaz Keysar*, William S. Horton

*Department of Psychology, University of Chicago, 5848 South University Avenue,
Chicago, IL 60637, USA*

It looks as if speakers design their utterances for specific addressees. According to the standard pragmatic theory, speakers follow a principle of optimal design: 'The speaker designs his utterance in such a way that he has good reason to believe that the addressee can readily and uniquely compute what he meant on the basis of the utterance along with the rest of their common ground.' (Clark et al., 1983, p. 246; see also Clark and Marshall, 1981). As an example, suppose that Linda is looking out the window and says to her husband 'the car is gone.' The principle of optimal design suggests that she tailored this utterance to her husband in a way that takes into account their mutual knowledge or common ground. The mutual beliefs that she presumably used for such a design could include the fact that they have only one car, that it was parked within view of the window, that they are the only ones using it legally, that neither of them could have moved the car since it was parked, that she has the ability to recognize their car, that cars don't typically go in and out of sight without someone driving them, and so on. Additionally, the principle of optimal design is supposed to work both ways. On the one hand speakers design their utterances by taking into account the common ground with the addressees; on the other hand the addressees assume that speakers engage in such optimal design and use their mutual knowledge to interpret the utterance.

Much research has accumulated evidence for this view. The basic idea has been to show that when mutual knowledge varies, utterances vary accordingly. Indeed, people tend to speak differently when they address different audiences (e.g. Isaacs and Clark, 1987; Krauss, 1987; Fussell and Krauss, 1989). Yet the fact that it looks as if speakers are using mutual knowledge to design utterances does not mean that they are doing so (Brown and Dell, 1987). The heart of the problem is paradigmatic: different from common practice in psycholinguistics research, the traditional

* Corresponding author. Tel.: +1 773 7025830; fax: +1 773 8340873; e-mail: boaz@ccp.uchicago.edu

research program in pragmatics focuses on general principles and does not investigate underlying mental processes.

Horton and Keysar (1996) adopted a new approach in pragmatics that goes beyond the description of principles. That paper focused on the question: *how* do speakers use common ground? Specifically, it compared two production models and conducted a critical experiment to test them. Polichak and Gerrig (1998) raise several concerns with respect to the Horton and Keysar paper. Their concerns are quite comprehensive because they cover the major elements involved in evaluating a new research program:

1. *Paradigmatic considerations* – the relationship between the new paradigm and the traditional approach
2. *Theoretical considerations* – the possibility of alternative models
3. *Empirical considerations* – the experimental design and task
4. *Analytic considerations* – the inferential logic from the data to the theory.

In the remainder of this paper, we will attempt to re-evaluate the proposal in Horton and Keysar (1996) in light of these considerations.

1. Paradigmatic considerations: beyond principles

Horton and Keysar (1996) evaluated the mental processes that might underlie the use of mutual knowledge in production. They contrasted two process models: the ‘Initial Design’ and the ‘Monitoring and Adjustment’ model. The Initial Design model assumes that speakers plan their utterances by taking into account their mutual knowledge with the addressee. In contrast, the ‘Monitoring and Adjustment’ model assumes that utterance plans do not involve consideration of mutual knowledge. Instead, plans are monitored for violations of mutual knowledge via a slower and more effortful process. If such a violation is detected, then the plan is revised. The models differ in the role they assign to the mutuality of knowledge – one assumes that it guides the design of the utterance, while the other assumes that it is only used as a correction mechanism.

Before considering the specifics of the models and the details of the data, it is important to attempt to relate the newer proposal to the traditional, accepted approach. In particular, how does the Horton and Keysar paper represent the basic assumptions of the traditional approach? One sin that might be committed by a new paradigm is to misrepresent the traditional paradigm, and indeed, one of Polichak and Gerrig’s central claims is that the Horton and Keysar paper commits just such a sin. Specifically, they argue that the paper misattributes the Initial Design model to Clark and Marshall (1981).

Polichak and Gerrig are correct to assert that Clark and Marshall never proposed the Initial Design model. They are wrong, however, to suggest that Horton and Keysar attribute this model to Clark and Marshall. Indeed, part of the point of the new paradigm is that the traditional pragmatic approach has not specified *any* process model of production. Therefore, it would have been internally inconsistent

for Horton and Keysar to make such an attribution. The Initial Design model is a process model of production and as such cannot be attributed to Clark and Marshall.

So, if the new paradigm does not misrepresent the traditional one, how *does* it represent it? One can see points of departure and connections between the two approaches. With respect to production, the traditional approach suggested only *that* speakers use mutual knowledge. The new approach, however, investigates *how* they do it. In the domain of comprehension, the points of connection are more apparent; the difference between the two approaches is one of focus. For example, Clark and Carlson (1981) make some assumptions about processing during comprehension. Specifically they suggest that when speakers use definite reference, their addressees restrict their memory search to mutually-known entities (p. 319). Keysar et al. (1998a) presented preliminary evidence against this ‘restricted search’ assumption, but future research might discover conditions under which such a restricted search does indeed hold.

In general, the thrust of the processing-centered approach in pragmatics is to spell out and test models in increasing detail. This is not to say that the description of general principles is not useful, but given that any particular principle can be realized by vastly different underlying processes, it seems crucial to design the types of experiment that will allow us to discover these processes.

2. Theoretic considerations: the generation and evaluation of process models

Horton and Keysar evaluated two process models of the use of mutual knowledge in production. The next step could be to look for alternative, perhaps better models. Polichak and Gerrig mention a third model, which is a variation of the Initial Design model. According to this model, speakers take mutual knowledge into account as part of the initial planning, but such a consideration can be very effortful. Polichak and Gerrig make the very important point that it is not always obvious what is mutually known, and that sometimes interlocutors must figure out what is mutually known, consider what is potentially known, retrieve information about common ground from long term memory, and so on. These situations are likely to induce an effortful, resource-demanding search for mutual knowledge. This is an interesting model and it would be important to develop appropriate tasks in the future to test it.

There is one major difference between the Initial Design model modified by Polichak and Gerrig and the Monitoring and Adjustment model. The modified Initial Design model assumes that the difficulty in using common ground in planning comes from the effort involved in solving a particular problem: figuring out what is common and what is not. In contrast, the Monitoring and Adjustment model assumes that even when speakers no longer have the problem – even when they know exactly what is mutual – they still do not use this knowledge to plan utterances. The effort, instead, is exerted in the slower monitoring process. In other words, even when *finding* the knowledge is not effortful, using it *is* because of the nature of monitoring.

This is precisely what the experiment in Horton and Keysar (1996) showed. Even when participants knew precisely what was mutually-known and what was not – when they did not have to ‘figure it out’ – they still did not *use* this knowledge when they were under time pressure. To demonstrate this, the experiment used a trivially easy task that is similar to what people do on a daily basis. Yet Polichak and Gerrig challenge the assumption that the task instantiates a ‘natural’ activity, which raises the possibility that participants did need to expend effort in figuring out what is mutually known. Evaluating this possibility requires us to more closely consider the details of the experiment.

3. Empirical considerations: beyond intuition

In order to compare the two models and to investigate the underlying process, the original experiment manipulated not only the mutuality of knowledge but also the response deadline. Participants described to addressees objects that appeared above context objects. The context objects were either shared with the addressee or privileged to the speaker. The context object was always below the target object. For example, participants described a circle which appeared in the context of a larger circle. The question of interest was, to what extent would speakers rely on the context object in their description. In this example, if they say ‘a small’ circle, it could suggest the use of context¹. To the extent that speakers design their utterances for their addressee, they might rely on shared context more than on privileged context. If they rely on shared and privileged context to the same degree, then one could conclude that they are not sensitive to the mutuality of knowledge.

The experiment demonstrated that speakers rely more on shared context than on privileged context, suggesting sensitivity to the knowledge of the addressee. But when speakers were hurried by being required to start talking within a 1.5-s deadline, the difference disappeared. Under speeded conditions speakers in the shared and privileged context conditions relied on context to the same degree.

How can the models explain these results? The Initial Design model as described in the Horton and Keysar paper does not predict a systematic change in performance as a result of time pressure, while the Monitoring and Adjustment model predicts just that: under time pressure, speakers do not have sufficient time or resources to monitor for violations of mutual knowledge and to adjust their utterances accord-

¹In order to understand the experiment, it is very important to understand the limitations of the design, especially if such a design is to be used in future research. One might think that the experiment is designed to see to what extent people use privileged knowledge. This would be a mistake because the design does not allow such a conclusion. The only thing that the experiment can reveal is whether or not speakers are at all sensitive to the difference between shared and privileged context. The extent to which speakers seem to use privileged context is by itself meaningless. For example, if speakers refer to the ‘small circle’ in the context of a larger privileged circle, it does not mean that they rely on the context object; they might have perceived the circle as small independent of the context. Therefore, one can only draw conclusions from the relative performance in the privileged and shared context conditions. When one finds a difference one can conclude that speakers are somewhat sensitive to the mutuality of the context, but one cannot even tell the *degree* to which they are sensitive to mutual knowledge from such a result.

ingly². Consequently, they fall back on the initial plan which is insensitive to considerations of mutual knowledge.

The logic of the design rests on the assumption that what is mutually known is apparent to the speaker participant. This is true only if the task requires participants to perform operations that they routinely do. If, however, the task presents a situation that people rarely encounter, then perhaps they cannot easily access the information about the mutuality of the knowledge. Polichak and Gerrig suggest that the task poses precisely this difficulty because it forces participants to do unnatural things. They suggest that ‘speakers have little or no experience dissecting perceptual displays into shared and unshared information, with the dissection achieved only by virtue of community membership.’ Their intuition is that such activity is ‘virtually non-existent.’ To bolster their intuition, they provide a very complicated example of a situation where the world is partitioned in this way. Indeed, both the analysis of the task and the example make the experiment seem utterly bizarre.

This criticism is puzzling because there also exists the opposite intuition that the task of the original experiment is trivially easy. How difficult could it be to keep in mind that the top object is shared and the bottom one is not? Participants in the experiment did not seem at all confused. Perhaps, then, it *is* natural to dissect ‘perceptual displays into shared and unshared information, with the dissection achieved only by virtue of community membership’. Indeed, simple examples do come to mind. For instance, recall how Linda looked out the window at the cars outside and commented on the family car. Every time she looks out the window and talks about the car to her husband, she is naturally participating in a situation analogous to the experimental task. She is looking at a perceptual array (i.e. many visible cars) that includes an object (i.e. the family car, whenever not stolen) which is mutually known by virtue of community membership (i.e. the fact that she always parks the car there). In addition, she is also looking at other, privileged objects (i.e. other parked cars, all recently parked by non-residents) and her husband knows that there are other cars, but doesn’t know which. Although it sounds complicated, it is actually extremely easy to partition the world in this way, and people do it routinely.

The specific lesson from all these examples and intuitions is that the original task is not unusual. More generally, the lesson is that how difficult it *seems* depends on what examples come to mind and what intuition follows. Perhaps the lesson should not be that this intuition or that intuition is correct, but that the field of pragmatics has been driven too much by intuitions and appeals to common sense, without the benefit of experimental backing. As it stands now, the task used by Horton and Keysar is the best we have, but one should certainly strive to develop better ones.

4. Analytic considerations: what’s an error?

The rationale of the Horton and Keysar experiment rests on the assumption that,

²The original Monitoring and Adjustment model assumes that the monitoring process is both slower and more effortful than planning. The experimental result could be explained by either or both assumptions.

under time pressure, speakers' utterances better reflect their initial planning. Polichak and Gerrig raise two concerns with respect to this assumption. The first concern is that instead of reflecting the initial plan, time pressure may ruin the very process of planning. If this is true, then one cannot use this methodology to study the nature of the plan. Yet hurried speakers in the experiment did not show signs of corrupted plans; their speech was as fluent and natural as that of the unhurried speakers. Another possibility is that time pressure selectively impairs only audience design. In order to show this, one must compare the attention required by using mutual knowledge to that required by other aspects of planning, and explain why attention required by common ground would be affected while other attention-demanding planning is not.

The more compelling argument against the logic of the original experiment has to do with the use of error data in production. Polichak and Gerrig again provide an intuitively compelling example: when a speaker under time pressure says 'cuf of coffee' instead of 'cup of coffee' (Fromkin, 1971), should we conclude that the speech error is a reflection of the initial plan? Should we assume that the speaker intended to say 'cuf'? Clearly, the answer is no, the speaker did not intend to say 'cuf'.

The general issue is very important, especially because of the extensive use of errors in the study of speech production (e.g. Fromkin, 1973; Fromkin, 1980). Do overt errors reveal the initial intention? We suggest that the answer is that some do, some don't. In the 'cuf' example the error does not, while in the 'small circle' example it does reflect on the initially-planned intention.

Perhaps a more illuminating example could come from considering other things for which speakers monitor (Levelt, 1989). For instance, people typically monitor their speech for social appropriateness, taking into account the social circumstances. Under pressure, though, people sometimes express themselves inappropriately. That would constitute a type of error, a social gaffe that would have been monitored out with the luxury of more time. Does this error reflect on the initial plan? Of course it does. Where else would it have come from?

In the same way that people monitor for social appropriateness, they might be monitoring for appropriateness vis a vis mutual knowledge. In the same way that time pressure does not always allow sufficient adjustment for social appropriateness, it might not allow sufficient adjustment for mutual knowledge. Both insufficient adjustments would result in an error and both errors would directly reflect on the initial plan.

It is important to note that anticipations, like saying 'cuf' for 'cup', are a very different kind of speech error from errors of social appropriateness. Anticipation typically results in a non-word, which is clearly not what the speaker intended. In contrast, social inappropriateness violates a social norm and reflect what the speaker wanted to say but should not have. Errors of violation of mutual knowledge are more similar to the latter than the former. Indeed, the only reason to consider them 'errors' is that they violate the *normative* standard of the principle of optimal design.

5. Conclusion: the challenge

Recent work in the processing of pragmatics (e.g. Brown and Dell, 1987; Dell and Brown, 1991; Ferreira and Dell, 1996; Horton and Keysar, 1996; Keysar, 1997, 1998; Keysar et al., 1996, Keysar et al., 1998a,b) both enriches and challenges the traditional pragmatic view. It enriches our understanding of pragmatic principles by revealing the mental processes that underlie pragmatic phenomena. So far, in considering the mental processes that underlie the use of mutual knowledge in production, two models have been compared experimentally: the Initial Design model and the Monitoring and Adjustment model. Polichak and Gerrig propose a modified Initial Design model. One could think of a variety of alternative models, and it would be important to contrast them with critical experiments.

The challenge that the original paper poses is *not* that the traditional view wrongly advocated the Initial Design model. Instead, it challenges the traditional assumption of optimal design. If some version of the Monitoring and Adjustment model turns out to be correct, then this would suggest that what looks like audience design is not really tailor-made speech. Recall that instead of assuming that planning occurs with mutual knowledge in mind, the model assumes that people monitor and revise plans that violate mutual knowledge. This is analogous to the difference between having a tailor custom-make a suit for you (i.e. Initial Design or effortful Initial Design) versus buying a suit from a store and making some adjustments if needed (i.e. Monitoring and Adjustment). Both suits might look custom-designed, but the second one isn't. It might be the case that speaking is more like altering an off-the-rack suit than like having one custom-made.

Acknowledgements

The writing of this paper was supported by a John Simon Guggenheim Memorial Fellowship and a PHS grant R29 MH49685 to Boaz Keysar. We greatly benefited from comments from Kate Baldwin, Dale Barr, Linda Ginzel, Rachel Giora and Anne Henly.

References

- Brown, P.M., Dell, G.S., 1987. Adapting production to comprehension: The explicit mention of instruments. *Cognitive Psychology* 19, 441–472.
- Clark, H.H., Carlson, T.B., 1981. Context for comprehension. In: J. Long, A. Baddeley (Eds.), *Attention and Performance IX*. Erlbaum, Hillsdale, N.J., pp. 313–330.
- Clark, H.H., Marshall, C.R., 1981. Definite reference and mutual knowledge. In: A.H. Joshe, B. Webber, I.A. Sag (Eds.), *Elements of Discourse Understanding*. Cambridge University Press, Cambridge, pp. 10–63.
- Clark, H.H., Schreuder, R., Buttrick, S., 1983. Common ground and the understanding of demonstrative reference. *Journal of Verbal Learning and Verbal Behavior* 22, 245–258.
- Dell, G.S., Brown, P.M., 1991. Mechanisms for listener-adaptation in language production: limiting the

- role of the 'model of the listener'. In: D. Napoli, J. Kegl (Eds.), *Bridges Between Psychology and Linguistics*. Academic Press, New York, pp. 105–129.
- Ferreira, V.S., G.S. Dell, 1996. Do speakers choose their words cooperatively? Investigating the production-comprehension interface. The 37th Annual Meeting of the Psychonomics Society. Chicago, IL.
- Fromkin, V., 1971. The nonanomalous nature of anomalous utterances. *Language* 47, 27–52.
- Fromkin, V.A. (Ed.), 1973. *Speech Errors as Linguistic Evidence*. Mouton, The Hague.
- Fromkin, V.A. (Ed.), 1980. *Errors in Linguistic Performance. Slips of the Tongue, Ear, Pen, and Hand*. Academic Press, New York.
- Fussell, S.R., Krauss, R.M., 1989. The effects of intended audience on message production and comprehension: reference in a common ground framework. *Journal of Experimental Social Psychology* 25, 203–219.
- Horton, W.S., Keysar, B., 1996. When do speakers take into account common ground?. *Cognition* 59, 91–117.
- Isaacs, E.A., Clark, H.H., 1987. References in conversation between experts and novices. *Journal of Experimental Psychology: General* 116, 26–37.
- Keysar, B., 1997. Unconfounding common ground. *Discourse Processes* 24, 253–270.
- Keysar, B., 1998. Language users as problem solvers: just what ambiguity problem do they solve? In: S.R. Fussell, R.J. Kreuz (Eds.), *Social and Cognitive Psychological Approaches to Interpersonal Communication*. Erlbaum, Hillsdale, NJ, in press.
- Keysar, B., Barr, D.J., Balin, J.A., Brauner, J.S., 1996. Common ground: an error correction mechanism in comprehension. The 37th annual meeting of the Psychonomic Society, Chicago, Illinois.
- Keysar, B., Barr, D.J., Balin, J.A., and Paek, T.S., 1998a. Definite reference and mutual knowledge: Process models of common ground in comprehension. *Journal of Memory and Language*, in press.
- Keysar, B., Barr, D.J., Horton, W.S., 1998b. The egocentric basis of language use: insights from a processing approach. *Current Directions in Psychological Science*, in press.
- Krauss, R., 1987. The role of the listener: addressee influences on message formulation. *Journal of Language and Social Psychology* 6, 81–98.
- Levelt, W.J.M., 1989. *Speaking: From Intention to Articulation*. MIT Press, Cambridge, MA.
- Polichak, J.W., Gerrig, R.J., 1998. Common ground and everyday language use: comments on Horton and Keysar (1996). *Cognition* 66, 183–189.