Discourse Studies
http://dis.sagepub.com/

Views from a cognitive scientist: cognitive representations underlying discourse are sometimes social

Arthur C. Graesser

Discourse Studies 2006 8: 59
DOI: 10.1177/1461445606059555

Views from a cognitive scientist: cognitive representations underlying discourse are sometimes social

Discourse Studies
Copyright © 2006
SAGE Publications.
(London, Thousand Oaks, CA and New Delhi)
www.sagepublications.com
Vol 8(1): 59–66.
10.1177/1461445606059555

ARTHUR C. GRAESSER
UNIVERSITY OF MEMPHIS

ABSTRACT Most areas of the cognitive and social sciences assume that knowledge representations are constructed and used during communication and that much of its content is social. Those of us who build computer models of comprehension and conversation are forced to be explicit about the nature of these knowledge representations and affiliated processes. There are some conditions when knowledge is not sufficiently social, and other conditions when knowledge is overly grounded in social mechanisms. The argument is advanced that constraints, coherence, and precision are very much at the heart of an explanation of the nature and amount of social knowledge. Asocial content reigns supreme when the referential world knowledge is highly constrained, coherently structured, and precisely specified. The social context of knowledge construction becomes progressively more influential to the extent to that world knowledge is vague, open-ended, imprecise, underspecified, and fragmentary.

KEY WORDS: *animated conversational agents, cognitive science, computational linguistics, speech acts, tutoring*

I might be the token renegade in this special issue. The slate of contributors is fully loaded with pioneers of discourse analysis who have carefully analyzed many of the social foundations of discourse and who have a head start in dissecting (or rejecting) those links between conversation and cognition that are reflected in recent publications of this community (e.g. te Molder and Potter, 2005; van Dijk, 1998). I will be adorning two hats that, if successful, might stir up cognitive disequilibrium in some readers and perhaps comic relief in others. Regarding the first hat, as former editor and current associate editor of the journal *Discourse Processes*, it is imperative that I embrace the view that cognitive representations underlie conversation and other forms of social interaction. That is what the vast majority of my community of researchers believe (Graesser et al., 2003; Kintsch, 1998), so that is the position that I feel compelled to defend. Attached to the brim of this first hat is the belief that the scientific method is the proper epistemological foundation for investigating these cognitive representations and processes.

Regarding the second hat, I will make the case that it is worthwhile to build computer models that simulate the discourse mechanisms. This second hat incorporates the fields of computational linguistics (Jurafsky and Martin, 2000) and computational discourse (Rich and Sidner, 1998). In my case, I build computer systems like AutoTutor (Graesser et al., 2001, 2004) which help students learn about Newtonian physics and computer literacy by holding a conversation with them in natural language. There are levels of analytical detail that a researcher is forced to worry about when building computer systems that attempt to comprehend natural language, produce natural language, and co-construct discourse. The systems that get built are always faulty, but the researchers learn a considerable amount from the limitations and errors of computer models.

For starters, cognitive scientists have routinely embraced social mechanisms in their cognitive theories. Whomever argued that cognitive representations are asocial was led astray down a scholarly trajectory that cut out mainstream developments in cognitive science. Consider the title of Bartlett's landmark 1932 book that launched the schema theories that still survive today: *Remembering: A Study In Experimental and Social Psychology*. Social content was extensively interwoven in the cognitive representations of discourse that were proposed in early discourse processing theories (van Dijk and Kintsch, 1983). One of the earliest models in artificial intelligence was Weizenbaum's (1966) Eliza program that simulated a Rogerian psychotherapist with only about 200 production rules; it is amazing how the illusion of comprehension could be created by such a simple mechanism. Subsequent models of dialogue in computational discourse attempted to track the beliefs, desires, and intentions (the BDI models) of participants in two-party dialogues between human and computer (Allen, 1995; Cohen and Perrault, 1979). Recent models with animated conversational agents attempt to generate speech, facial expressions, and gestures that help co-construct messages with the human through back channel feedback and filling in words (Cassell and Thorisson, 1999; Gratch et al., 2002). The notion that conversations are co-constructed by speech participants in joint activity is thoroughly familiar territory in cognitive science and discourse processes (Clark, 1996).

Those who build computer systems that simulate conversation are convinced that it is impossible to build a smooth conversation partner without the system having internal representations of the knowledge, beliefs, goals, intentions, plans, norms, values, and other cognitive representations of both speech participants. These representations include social knowledge and possibly emotions and attitude. Among the pressing theoretical questions are what content and how much context is needed in these cognitive representations for the conversational systems to function. Consider, for example, a symbolic structure of the speech act that is functionally a *request* (Cohen and Perrault, 1979).

REQUEST (S, H, ACT)
[Meaning the speaker requests the hearer to perform an action]
Constraints: **Speaker(S) Λ Hearer (H) Λ ACT(A) Λ H is agent of ACT**
Precondition: **WANT (S, ACT (H))**
[Meaning the speaker wants the hearer to do the action]
Effect: **WANT (H, ACTION (H))**
[Meaning the hearer wants to do the action after the request]
BELIEVE (H,WANT(S (ACT(H))))
[The hearer believes the speaker wants the hearer to do the action]

This analysis is aligned with Searle's (1969) seminal philosophical treatment of speech acts. It is important to acknowledge that a speech act constitutes a request if this complex conceptual pattern of conditions holds up in the context of the speech act. A speech act will not function as a request if the speaker does not want action A to occur, and is ineffective if the listener does not realize the speaker wants it to be done. It is a separate, irrelevant matter whether the speaker directly expresses the word *request* or the other primitive ontological verbs like *want*, *believe*, and *know*. Cognitive theories have never assumed that humans can reliably articulate the conceptual knowledge that underlies their social interactions. Rather interestingly, there are now computer systems that can recognize requests and other categories of speech acts without the need for the ontological verbs to be explicitly expressed and without the need for speakers to articulate verbal expressions precisely (Gratch et al., 2002; Olney et al., 2003). The goal, plan, and speech act recognition problems are difficult challenges computationally, but significant progress is being made by statistical algorithms that induce patterns from large corpora of speech samples (Jurafsky and Martin, 2000).

Another conclusion made by those who built computer models of conversation is that the successful systems can handle very specific tasks in social and physical settings that were highly constrained. Examples of these settings are customers asking the computer system about train schedules, callers interacting with telephone operators, customers making airline reservations, a guide helping a lost soul navigate in space, and a bartender interacting with customers. The setting my research group has been developing is tutoring, where the student and computer tutor collaboratively answer difficult questions. It is perhaps illuminating that the successful systems are highly constrained to a specific social-physical-task context, as opposed to being generic. The systems work when speech participants know their roles, the goals of the exchange are obvious, the relevant tasks are known, and so on. In a similar fashion, humans get socialized to understand characteristics of and differences among particular classes of settings, registers, genres, or whatever packages of social-physical-task context the discourse analyst has in mind. A child may have learned and mastered 10 of these classes of context whereas an adult may have mastered a thousand. These classes of context get recognized, launched, and recruited during the stream of everyday activity. Carried with each context is an array of cognitive representations that are instantiated and that govern interactions.

The tutoring context is an interesting one because the shared knowledge between tutor and student is low. My colleagues and I (Graesser et al., 1995) spent nearly a decade analyzing 100 hours of video-taped tutoring sessions. Some of the sessions were students in a middle school being tutored on mathematics by older students, whereas other sessions were college students being tutored on research methods. We tracked the discourse patterns and knowledge expressed in the student–tutor interactions, as well as their mastery of the material. Many of these discourse patterns were eventually incorporated into AutoTutor, the computer system that simulates tutorial dialogue. One might view this project as a synthesis of methodologies from discourse analysis, discourse processing, and computational discourse. These sorts of fusions are very much at the forefront of future research, I believe.

One of the critical challenges we faced in our tutoring research was determining whether a student knew some piece of knowledge. In essence, or perhaps put in a crassly simple way, how do we know whether 'Student S knows Proposition P.' For example, in our physics tutor, one proposition P is Newton's law *net force equals mass times acceleration*. Novice tutors often ask 'comprehension gauging' questions (*Do you understand?, Do you follow?*) in order to ascertain whether the student has knowledge of propositions like this one, or knowledge of larger chunks of material. Most students answer *yes* to such questions either to be polite, to save face, or to gird a posture of competence. However, such answers are not to be trusted. Empirical data reveal that there is a negative correlation between grades (or mastery of the content) and the likelihood that students answer yes. So it is the students who have deeper knowledge who answer that they don't understand.

Another approach to determining whether Student S knows P is to get them to articulate P. In essence, the tutor gives hints or prompts in an effort to get the student to express P verbally. A persistent tutor produces a series of speech acts, often badgering the student, until the student expresses the words in P to the tutor's satisfaction. Unfortunately this may not be sufficient to conclude the student knows P because they often fail to articulate P later on while answering another question that requires the articulation of P. Articulating P in one context hardly insures that P will be articulated in another related context. Research in cognition and instruction has substantiated the generalization that students are notoriously poor in transferring technical knowledge from one context to structurally similar contexts. Evidence that a student knows P would be on firmer ground if the student expressed P in all appropriate contexts, but not inappropriate contexts.

Yet another approach is getting the student to perform an action that presupposes mastery of proposition P. In essence, doing is better than saying. So they might perform actions in an interactive simulation or game that presupposes mastery of P. One current version of AutoTutor has physics microworlds with interactive simulation (Graesser et al., 2005). Many researchers in cognition and education are most convinced that the student knows P when they consistently make decisions and perform actions that presuppose P. Such active knowledge is believed to be superior to inert knowledge that might be expressed verbally but is not readily applied to practical problems (Scardamalia and Bereiter, 1985). The main point to be conveyed is that cognitive scientists of today have multiple indicators of whether students know something, including answers to questions, verbal expression, action, gestures, intonation, facial expression, and body posture (Craig et al., 2004; Graesser et al., 2005). Developers of intelligent tutoring systems use Bayesian statistical analyses from various indicators to infer whether or the degree to which Student S knows P (VanLehn et al., 2002).

Students often want confirmation from the tutor that what they say or do is correct. So with a rising intonation the student might express *Isn't force equal to mass times acceleration?* They seek a confirmation or ratification from the tutor that their knowledge is correct. So the tutor might say *that's right* or a more sophisticated tutor might ask *What do you think?* or *How might we find out?* It is informative to note that this short pedagogical feedback by the tutor is often misleading when the student expresses incorrect or vague information. They may say *already*, *okay*, or even *yeah* verbally (positive feedback) but with an intonation or facial expressions full of skepticism (negative feedback). When this occurs, the linguistic record is misleading whereas the paralinguistic channels provide accurate feedback to the students. This mixed feedback is presumably generated either as a face saving tactic or as a form of support to keep the student from being discouraged. The important conclusion, from the present standpoint, is that the human tutors and AutoTutor incorporate these multiple channels and modalities. The more general point is that cognitive science embraces multi-channel communication, emotions, and bodily action.

The major thrust of my argument so far is that cognitive scientists have tried to account for many of the interesting challenges put forth by those in discourse analysis who might have doubted the value of a cognitive theory. It remains to be seen whether the advances in cognitive science are sufficiently impressive to convince the community in discourse analysis and conversation analysis.

At this point I want to shift perspective a bit. My goal is to advance the claim that there are some conditions when knowledge is not sufficiently social, and other conditions when knowledge is overly grounded in social mechanisms. If this is claim correct, then some of the radical assertions about the links between conversation and cognition are overstated. For example, the notion that knowledge is merely a social construction may be giving too much credit to social systems. There are times when knowledge is asocial or minimally social, with constraints and coherence that give it a reality of its own.

Consider the conditions when we wish knowledge was more socially grounded. As all of us know, there is an unfortunate tendency for many children and adults to read textbooks uncritically and to passively accept whatever is written down as truth. We might prefer these readers to critically evaluate the epistemological status of claims and the social foundations that led to the claims. Comprehension mechanisms can change rather noticeably when these readers are trained to adopt a critical stance. One example of a systematic training program is Questioning the Author (Beck et al., 1997). Children are trained to ask themselves the following questions as they read: *What evidence is there for this claim? Who is the author? What expertise does the author have? Why did the author say that? How is this assertion relevant to that conclusion?* The author is viewed as an imperfect creature who is there to be challenged. After many months of training with Questioning the Author, the children end up comprehending the text more deeply. One might expect and hope there would be improvements in their reasoning skills and critical inquiry, but that is an open question for empirical research. The important implication, from the present standpoint, is that many readers without this training simply get caught up in the content of the material to be studied and minimally consider the social foundations, if at all. In this case, the textual and referential content dominates to the point where the constraints of social context are anemic.

Constraints, coherence, and precision are very much at the heart of the matter. When the world knowledge is highly constrained, coherently structured, and precisely specified, then asocial content reigns supreme. On the other hand, the extent to which world knowledge is vague, open-ended, imprecise, under-specified, and fragmentary is the extent to which the social context of knowledge construction becomes more and more influential. At times the social context becomes so influential that individuals can be convinced that virtually any claim is true. Studies that demonstrate the malleability of respondents in surveys, clients in therapy sessions, or public servants in interviews by the media are quite expected when the referential knowledge lacks constraints, coherence, and precision. This is a testable generalization, of course, one that can be investigated through the scientific method. And, who knows, the generalization just might be wrong.

ACKNOWLEDGEMENTS

The Tutoring Research Group (TRG) is an interdisciplinary research team comprised of approximately 35 researchers from psychology, computer science, physics, and education (visit [http://www.autotutor.org]). The research on AutoTutor was supported by the National Science Foundation (SBR 9720314, REC 0106965, REC 0126265, ITR 0325428) and the DoD Multidisciplinary University Research Initiative (MURI) administered by ONR under grant N00014–00–1-0600. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of DoD, ONR, or NSF.

REFERENCES

Allen, J. (1995) *Natural Language Understanding*. Redwook City, CA: Benjamin/Cummings.

Bartlett, F.C. (1932) *Remembering: A Study in Experimental and Social Psychology*. Cambridge: Cambridge University Press.

Beck, I.L., McKeown, M.G., Hamilton, R.L. and Kucan, L. (1997) *Questioning the Author: An Approach for Enhancing Student Engagement with Text*. Newark, DE: International Reading Association.

Cassell, J. and Thorisson, K. (1999) 'The Power of a Nod and a Glance: Envelope vs. Emotional Feedback in Animated Conversational Agents', *Applied Artificial Intelligence* 13 : 519–38.

Clark, H.H. (1996) *Using Language*. Cambridge: Cambridge University Press.

Cohen, P.R. and Perrault, C.R. (1979) 'Elements of a Plan-based Theory of Speech Acts', *Cognitive Science* 3: 177–212.

Craig, S.D., Graesser, A.C., Sullins, J. and Gholson, B. (2004) 'Affect and Learning: An Exploratory Look into the Role of Affect in Learning', *Journal of Educational Media* 29: 241–50.

Graesser, A.C., Gernsbacher, M.A. and Goldman, S. (eds) (2003) *Handbook of Discourse Processes*. Mahwah, NJ: Erlbaum.

Graesser, A.C., Lu, S., Jackson, G.T., Mitchell, H., Ventura, M., Olney, A. and Louwerse, M.M. (2004) 'AutoTutor: A Tutor with Dialogue in Natural Language', *Behavioral Research Methods, Instruments, and Computers* 36: 180–93.

Graesser, A.C., Olney, A., Haynes, B.C. and Chipman, P. (2005) 'AutoTutor: A Cognitive System that Simulates a Tutor that Facilitates Learning through Mixed-initiative Dialogue', in C. Forsythe, M.L. Bernard and T.E. Goldsmith (eds) *Cognitive Systems: Human Cognitive Models in Systems Design*. Mahwah, NJ: Erlbaum.

Graesser, A.C., Person, N.K. and Magliano, J.P. (1995) 'Collaborative Dialogue Patterns in Naturalistic One-to-one Tutoring', *Applied Cognitive Psychology* 9: 3591–28.

Graesser, A.C., VanLehn, K., Rose, C., Jordan, P. and Harter, D. (2001) 'Intelligent Tutoring Systems with Conversational Dialogue', *AI Magazine* 22: 39–51.

Gratch, J., Rickel, J., Andre, E., Cassell, J., Petajan, E. and Badler, N. (2002) 'Creating Interactive Virtual Humans: Some Assembly Required', *IEEE Intelligent Systems* 17: 54–63.

Jurafsky, D. and Martin, J.H. (2000) *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Upper Saddle River, NJ: Prentice Hall.

Kintsch, W. (1998) *Comprehension: A Paradigm for Cognition*. New York: Cambridge University Press.

Olney, A., Louwerse, M.M., Mathews, E.C., Marineau, J., Mitchell, H.H., and Graesser, A.C. (2003) 'Utterance Classification in AutoTutor', in J. Burstein and C. Leacock (eds) *Building Educational Applications Using Natural Language Processing: Proceedings of the Human Language Technology – North American Chapter of the Association for Computational Linguistics Conference 2003 Workshop*, pp. 1–8. Philadelphia, PA: Association for Computational Linguistics.

Rich, C. and Sidner, C.L. (1998) 'COLLAGEN: A Collaborative Manager for Software Interface Agents', *User Modeling and User-adapted Interaction* 8: 315–50.

Scardamalia, M. and Bereiter, C. (1985) 'Fostering the Development of Self-regulation in Children's Knowledge Processing', in S.F. Chipman, J.W. Segal and R. Glaser (eds) *Thinking and Learning Skills*, Vol. 2, pp. 563–77. Hillsdale, NJ: Erlbaum.

Searle, J.R. (1969) *Speech Acts*. London: Cambridge University Press.

te Molder, H. and Potter, J. (eds) (2005) *Conversation and Cognition*. Cambridge: Cambridge University Press.

van Dijk, T.A. (1998) *Ideology*. London: Sage.

van Dijk, T.A. and Kintsch, W. (1983) *Strategies of Discourse Comprehension.* New York: Academic Press.

VanLehn, K., Lynch, C., Taylor, L., Weinstein, A., Shelby, R., Schulze, K. et al. (2002) 'Minimally Invasive Tutoring of Complex Physics Problem Solving', in S.A. Cerri, G. Gouarderes and F. Paraguacu (eds) *Proceedings of the Sixth International Conference on Intelligent Tutoring Systems 2002*, pp. 367–76. Berlin: Springer Verlag.

Weizenbaum, J. (1966) 'ELIZA – A Computer Program for the Study of Natural Language Communication between Man and Machine', *Communications of the ACM* 9: 36–45.

ART GRAESSER is presently a Professor in the Department of Psychology and Computer Science at The University of Memphis. He is co-director of the Institute for Intelligent Systems and chair of the Department of Psychology. Dr Graesser received his PhD in psychology from the University of California at San Diego. His primary research interests are in discourse processing and cognitive science. More specific interests include text comprehension, reading, inference generation, conversation, question asking and answering, tutoring, education, artificial intelligence, computational linguistics, and human-computer interaction. He is currently associate editor (and previous editor) of the journal *Discourse Processes* and was first editor of the recent *Handbook of Discourse Processes*. In addition to analyzing discourse theoretically and empirically, he has designed software in learning, language, and discourse technologies, including AutoTutor, Coh-Metrix, Question Understanding Aid (QUAID), QUEST, and Point&Query. ADDRESS: Department of Psychology, 202 Psychology Building, The University of Memphis, Memphis, TN 38152–3230, USA. [email: a-graesser@memphis.edu]