Preface

What are the central questions we should pose about humans' use of language for communication? And which methods should we use to investigate these issues?

In his book *Arenas of Language Use*, Clark (1992) noted that although there have been many different approaches to answering these questions, two distinct traditions have emerged within psycholinguistics, which he called the ''language-as-product'' and ''language-as-action'' traditions.

The product tradition, which has dominated psycholinguistics, has its roots in George Miller's (1962) synthesis of the then-emerging information-processing approach to cognition with Chomsky's (1957, 1959) revolutionary approach to linguistic knowledge as a cognitive system of rules and representations. Clark labeled this the language-as-product tradition because it focuses on the cognitive processes by which listeners recover, and speakers create, linguistic representations—the ''product'' of comprehension.

The second tradition sketched by Clark, the language-as-action tradition, has its roots in work by the Oxford philosophers of language use (e.g., Austin 1962, Grice 1957, and Searle 1969), and work on conversational analysis (e.g., Schegloff and Sachs 1973). The action tradition in language processing has been extended by psycholinguists focusing primarily on pragmatics and by computational linguists working in dialogue. This approach focuses on how people use language to perform acts in conversation, arguably the most basic form of language use. Psycholinguistic research within the action tradition focuses primarily on investigations of interactive conversation using natural tasks, typically in settings with real-world referents and well-defined behavioral goals.

This edited volume represents what we hope is the early stages of a movement to merge these traditions. Before we turn to a review of the contents of the book, we briefly describe the product and action traditions in more detail and lay out some of the reasons we believe that combining these traditions is both desirable and tractable.

Differences between the Product and Action Approaches

Although a broad range of perspectives can be found in the product tradition, most researchers in this tradition share a common set of methodological and theoretical assumptions. For instance, language comprehension and production are treated almost entirely as cognitive processes. As a result, experimental investigations focus on how individual ''comprehenders''—that is, readers or listeners—assemble the (primarily syntactic) linguistic representations necessary for interpretation and how individual ''speakers'' translate thoughts into linguistically structured utterances. The emphasis on syntactic representations can be traced to the influence that generative linguistics has had on the study of language comprehension. Linguistics has focused on syntactic phenomena in large part because of key recursive functions that are believed to underlie the productive aspects of language, especially at the level of sentence descriptions.

Because the product tradition is also heavily influenced by information-processing approaches to cognition, many of the important theoretical issues are also rooted in questions about how information from different subsystems is processed and how information is integrated during different stages in information processing (e.g., Fodor 1983; Garfield 1987). One consequence is that most researchers in the product tradition consider ''online'' measures to be the methodological gold standard. For example, in sentence processing, researchers focus on the moment-by-moment syntactic choices made by readers and listeners (e.g., see Clifton, Frazier, and Rayner 1994) in order to evaluate linguistically motivated theories of sentence syntactic processing (e.g., Frazier 1989; Tanenhaus and Trueswell 1995; MacDonald, Pearlmutter, and Seidenberg 1994). Several decades of research in this area has established that processing decisions at these levels are closely time-locked to the input. With only a few notable exceptions (e.g., Altmann and Steedman 1988; Crain and Steedman 1985), theories of these processes have emphasized ''noncontextual'' linguistic contributions to language understanding. Context is viewed as a correlated constraint that can inform decisions at points of temporary ambiguity or is used to instantiate interpretations that build on context-independent linguistic representations (for discussion see Tanenhaus, Chambers, and Hanna, forthcoming).

In the action tradition, dialogue, including the form of utterances, is viewed as emerging from joint actions created by collaboration between interlocutors in a conversation (Clark 1996). From this theoretical perspective the processing of an utterance is inextricably intertwined with the place, time, and situation of its use. Participants in conversations are believed to establish and update their common ground, which forms

the backdrop against which utterances are generated and interpreted. As a result, psycholinguistic research in the action tradition has focused on conversational issues, including the assessment of interlocutors' intentions and the coordination processes necessary to establish reference to familiar and novel objects in the world (cf. Clark 1992, 1996). Unlike the product tradition, in which experimental approaches typically examine a single person in the act of reading or speaking, psycholinguistic studies in the action tradition have focused on the behavior of multiple participants engaged simultaneously in both speaking and listening. Moreover, because real-time measures of comprehension have not been well suited to the study of situated language use, most research in this tradition has relied on offline measures, with theoretical accounts focusing on the more global properties of dialogue and reference.

The methodological differences in the approaches can be illustrated by comparing two well-known experimental methods from the product and action traditions. Figure P.1 illustrates a schematic of a prototypical product-based task, cross-modal lexical priming (Swinney et al. 1978). Cross-modal priming builds on the classic finding that response times to a target word are faster when the target is preceded by a semantically related prime word (Meyer and Schvanaveldt 1970). The subject, who is wearing headphones, listens to sentences prerecorded by the experimenter. A sentence or short sequence of sentences is presented on each trial. At some point in the sentence a target letter string appears on a computer monitor, allowing for experimenter control over the timing of the probe with respect to the input. The subject's task is to make a forced-choice lexical decision indicating whether the letter string is a word or not. The pattern of lexical-decision times is used to assess comprehension processes. For example, when

Figure P.1
Schematic of prototypical product experiment: Cross-modal lexical priming with lexical decision.

the target word follows *testified*, a verb whose object, *doctor*, has been fronted in a relative clause, lexical decisions on words that are associatively related to the fronted object are faster than lexical decisions on unrelated target words. Comprehension questions or a memory test are presented to ensure that the subject attends to the sentence.

A prototypical example of an action-based task is the referential communication task originally introduced by Krauss and Weinheimer (1966). A schematic of a well-studied variant of this task introduced by Clark and his colleagues (e.g., Clark and Wilkes-Gibbs 1986) is illustrated in figure P.2. Two naive participants, a matcher and a director, are separated by a barrier. Each has the same set of shapes arranged in different positions on a numbered grid. These objects are not very ''codable'' in the sense that a single word does not typically come to mind for all participants when describing such objects. Their goal is for the matcher to rearrange the shapes on his grid to match the arrangement on the director's grid. The resulting conversation can then be analyzed to provide insights into the principles that guide interactive conversation.

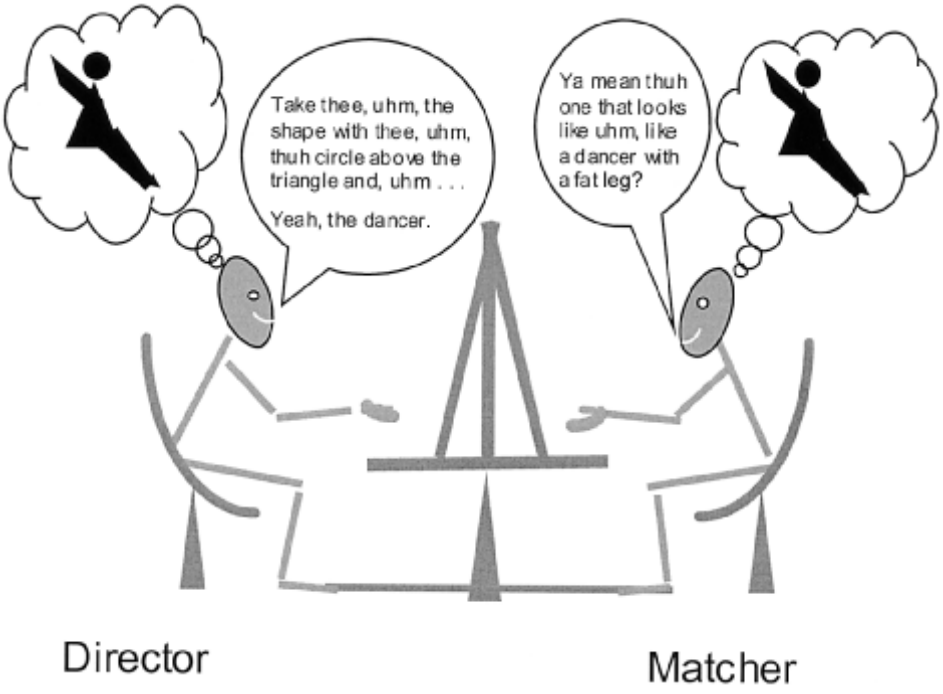

Figure P.2
Schematic of prototypical action task: Referential communication task with Tangrams.

Given the theoretical concerns and methodological tools used by the action and product traditions, it is not surprising that each has viewed the other with a certain amount of skepticism. For instance, researchers in the action tradition have criticized the product tradition for running ''phone-booth'' psycholinguistic studies (Clark 1992) in which willing-but-confused human participants are seated in a dark room and bombarded with written or spoken material that is devoid of much relevant context, or any possibility for assessing the speakers' goals. Thus, one may question whether experiments conducted within traditional paradigms using relatively ''decontextualized'' materials will generalize to more normal modes of language use. Likewise, the product tradition has criticized the action researchers for taking what might be called a ''let-all-flowers-bloom'' approach to psycholinguistics, in which research problems are left open ended, theoretical accounts are rarely mechanistic, and fewer links to linguistic or computational formalisms are made.

These biases aside, it is clear that detailed complex linguistic knowledge is a central component of human-language comprehension and production. Moreover, language is comprehended and generated in real time using basic information-processing mechanisms. However, it is *also* clear that much of this process is necessarily intertwined with the ongoings of the ambient world, as well as the intention of language users to communicate perspectives on that world. It therefore seems that attention to both sides of the language coin, the cognitive and social, will be important for understanding how language is processed in natural settings.

A Bridging of Traditions

A confluence of methodological and theoretical developments in psycholinguistics, linguistics, and computational linguistics, all related to the goal of providing mechanistic accounts of language use within rich referential environments, suggest that the time is ripe to bridge the product and action traditions.

Within psycholinguistics, the opportunity of such a merger has arisen with the advent of world-situated eye-tracking techniques (e.g., see Cooper 1974; Tanenhaus et al. 1995). In this technique, a listener's eye gaze is followed as he or she responds to spoken instructions to move objects about in the world (e.g., Tanenhaus et al. 1995), generates utterances (e.g., Eberhard 1998; Griffin and Bock 2000), or listens to spoken descriptions of visually copresent scenes (e.g., Altmann and Kamide 1999). This technique has allowed researchers to conduct studies that, from the subject's perspective, are like the contextually rich conversation studies devised in the action tradition, but from the experimenters' perspective, are studies that (behind the scenes) generate

linguistically time-locked behavioral data relevant to the product tradition. This Wizard-of-Oz approach, in which the product tradition's experimental gadgets and gizmos are hiding behind a curtain, permits the study of the inner workings of the comprehension and production machinery while manipulating factors central to the action tradition. This work has already revealed examples of amazingly rapid coordination of the listener's linguistic knowledge with his or her knowledge about the relevant visual-referent world (e.g., Tanenhaus et al. 1995; Spivey et al. 2002; Sedivy et al. 1999; Altmann and Kamide 1999), the listener's assessment of the speaker's perspective on that same world (Keysar, Barr, Balin, and Paek 1998; Keysar, Barr, and Horton 1998; but see also Hanna, Tanenhaus, and Trueswell 2003), discourse organization (Arnold et al. 2000), and even the development of these interactive and integrative mechanisms in young children (Trueswell et al. 1999).

However, research using these techniques has barely scratched the surface in terms of addressing issues that arise in studying language generated in natural conversational interactions. It has also been largely uninformed by new theoretical developments within computational linguistics and formal semantics. Until recently, most implemented computer models of conversation operated in only very restricted domains, typically processing a handful of scripted dialogues. These systems did not make contact with psycholinguistic phenomena at a useful grain. However, the dramatic increase in the speed and power of computers, along with improvements in real-time speech recognition and visual-pattern recognition, has made it possible to explore a new generation of conversational agents that engage in interactive conversation with people in practical dialogue (e.g., task-oriented dialogue). For example, Allen and his colleagues at Rochester have developed a system that engages in cooperative problem solving using unrestricted spoken language to coordinate and plan the most efficient train routes in a model world, given existing constraints of that world (e.g., Allen et al. 1995, 1996; Heeman and Allen 1999). Cassell and colleagues at MIT, in collaboration with Stone and his colleagues at Rutgers, have been exploring embodied conversational agents that coordinate gestures, utterances, and postural signals in generating and understanding interactive conversation (e.g., Cassell, Stone, and Yan 2000). These working systems incorporate theoretical proposals about real-time cross-modality integration and generation that cry out for experimental evaluation. As computer systems become more realistic, they are likely to serve as useful hypothesis-testing domains for evaluating how interactional variables influence human-language comprehension and production. Indeed, Brennan and colleagues have pioneered such an approach, examining computer-dialogue performance in an experimental setting (Brennan and Hulteen 1995; Brennan 1998).

Steps toward Formalizing Situated Language Use

The writing of the chapters in this book arose out of a special session of the 2001 CUNY Human Sentence Processing conference, held at the University of Pennsylvania with the generous support of the National Science Foundation (BCS-0096377) and the Institute for Research in Cognitive Science. Indeed, the signature conference for the product tradition has been the CUNY conference originally founded by Janet Fodor in 1987. More than 200 linguists, cognitive psychologists, and computational linguists now gather annually at this conference to address issues in sentence processing.

The goals of the special session were to bring together key researchers from the product and action traditions who shared the interest of further connecting cognitive and social approaches to language processing within dynamic models of language use. Chapters based on both the invited and submitted presentations of this session appear in this book.

Part I of the book, titled ''Reviews and Theoretical Perspectives,'' features a set of four review/position papers. In chapter 1, Tanenhaus and Trueswell argue for the importance of conducting real-time studies that investigate action-type variables. They outline methodological desiderata for such approaches and argue that eye tracking meets the central criteria for a product-based measure, while generalizing to action-based paradigms. They conclude with a review of how this technique can be used to study a wide range of issues that bridge the product and action traditions.

Next, in chapter 2, Stone lays out a representational and computational framework for utterance interpretation in human dialogue, specifically within a Gricean view of language use as an intentional activity. We believe this chapter makes significant theoretical advances for how best to connect Gricean observations to formal language-processing systems. Stone begins the chapter by considering a set of task-oriented dialogues, which he uses to motivate the need for detailed pragmatic representations that interface with linguistic representations. These dialogues highlight the complexity of the problem facing interlocutors in almost every exchange, but Stone provides some rather elegant solutions to these problems via a set of representations pertaining to pragmatic interpretations. He then cashes in on the advantages of such a representational system when he explores how utterance understanding and utterance production can be viewed as operations on these pragmatic representations. Stone discusses how this formalism might operate within a constraint-satisfaction system, and even considers connections to experimental work presented in other sections of this book.

In chapter 3, Keysar and Barr review a more specific attempt at bridging the product and action traditions within a theory of coordinated reference among interlocu-

tors. The authors address directly the important issue of how to reconcile context-independent linguistic representation and processes (the mainstay of the product tradition) and context-dependent mechanisms (emphasized in the action tradition). Based on a range of eye-gaze and other online data, they propose a staged model of reference resolution by both speakers and listeners in which initial reference computation is determined using relatively little conversational knowledge regarding a fellow interlocutor's perspective. This early egocentric stage generates representations that may only later be evaluated against broader conversational knowledge, in particular the referential common ground established between interlocutors. This line of research has generated a lively debate about the time course with which common-ground information is used during definite and pronominal reference, with other researchers arguing against staged approaches in favor of the simultaneous application of multiple constraints (see in particular Hanna and Tanenhaus, chapter 5, this book).

In chapter 4, Brennan details her view of reference coordination, which has been heavily influenced by the action tradition. Brennan reviews the Clark and Wilkes-Gibbs (1986) reference-contribution model, which defines reference as a coordination process between interlocutors. Brennan expands on this account by providing more detailed predictions about the dynamics of reference coordination during conversation. And, in an interesting twist, she reviews a previously unpublished study from the 1980s where she collected real-time comprehension measures of interlocutors engaged in a referential communication task. This work foreshadowed much current research activity and sheds light on some present controversies in the field.

The remaining sections of this book consist of shorter reports of experimental findings in the literature. Taken together, these chapters offer a snapshot of current work that begins to bridge the product and action approaches. We have organized these chapters into four groups.

The first group, comprising Part II "Speakers and Listeners as Participants in Conversations," examines language-processing issues as they occur in natural and seminatural conversational settings. Multiple research methods are represented here. Hanna and Tanenhaus (chapter 5) and Brown-Schmidt, Campana, and Tanenhaus (chapter 6) both examine reference using eye-gaze measures; Bard and Aylett (chapter 7) provide spoken-corpus data; and the contributions from McLean, Pickering, and Branigan (chapter 8) and from Schafer and Speer (chapter 9) examine linguistic and behavioral measures in dialogue settings. These chapters have a common thread in that all explore conversational phenomena that have competing explanations from product and action approaches to language use.

Part III, ''Language-Scene Interactions,'' examines how nonlinguistic information, gleaned from visual scenes, can be used by listeners to constrain and predict linguistic hypotheses. In chapter 10, Kamide, Altmann, and Haywood examine predictive linguistic processing in both English and Japanese listeners. This work shows that verbs and other lexical items, when interpreted with respect to visual scenes, allow for rapid predictions of upcoming, yet-to-be-heard, constituents. In chapter 11, Gennari, Meroni, and Crain examine how prosodic and visual-scene information interact to constrain the interpretation of quantifiers. In chapter 12, Arnold, Brown-Schmidt, Trueswell, and Fagnano examine referential issues from a developmental perspective, asking how linguistic and nonlinguistic cues contribute to the development of online pronoun interpretation.

The contributions in Part IV, ''Product Approaches to Action Variables,'' use measures from the product tradition to explore issues traditionally discussed in the action tradition. In chapter 13, Almor describes a computational model of reference that emphasizes how assumptions regarding information-processing load interact with pragmatic considerations. In chapter 14, Bailey and Ferreira examine disfluencies, and how they influence syntactic-ambiguity resolution. In chapter 15, Fitneva and Spivey examine how perceived speaker authorship constrains lexical-ambiguity resolution.

Part V, ''Gricean Phenomena,'' discusses how phenomena typically construed as examples of Grice's cooperative principle are instantiated in language use, focusing specifically on reference. In chapter 16, Barr presents a set of artificial-language simulations that explore how referential systems can emerge from formal-language users whose behavior is egocentric. Sedivy (in chapter 17), on the other hand, uses Gricean considerations to explain key online referential findings on the generation and interpretation of different classes of prenominal adjectives.

Clearly the body of work presented in this book represents only the first hesitant steps toward bridging the action and product traditions. We hope that the work presented here will motivate more researchers in the computational, psycholinguistic, and linguistic communities to pursue research that builds on and transcends these initial efforts.

References

Allen, J. F., Schubert, L. K., Ferguson, G., Heeman, P., Hwang, C. H., Kato, T., Light, M., Nathaniel, G., Martin, N. G., Miller, B. W., Poesio, M., and Traum, D. R. 1995. The TRAINS Project: A case study in building a conversational planning agent. *Journal of Experimental and Theoretical AI*, 7.

Allen, J. F., Miller, B. W., Ringger, E. K., and Sikorski, T. 1996. A Robust System for Natural Spoken Dialogue. In *Proceedings of the 1996 Annual Meeting of the Association for Computational Linguistics*, 62–70.

Altmann, G., and Kamide, Y. 1999. Incremental interpretation of verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247–264.

Altmann, G., and Steedman, M. 1988. Interaction with context during human sentence processing. *Cognition*, *30*, 191–238.

Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., and Trueswell, J. C. 2000. The rapid use of gender information: Evidence of the time course of pronoun resolution from eyetracking. *Cognition*, *76*, b13–b26.

Austin, J. L. 1962. *How to Do Things with Words*. Cambridge, MA: Harvard University Press.

Brennan, S. E. 1998. The grounding problem in conversations with and through computers. In S. Fussell and R. J. Kreuz, eds., *Social and cognitive psychological approaches to interpersonal communication*, 201–225. Mahweh, NJ: Erlbaum.

Brennan, S. E., and Hulteen, E. 1995. Interaction and feedback in a spoken language system: A theoretical framework. *Knowledge-Based Systems*, *8*, 143–151.

Cassell, J., Stone, M., and Yan, H. 2000. Coordination and context-dependence in the generation of embodied conversation. *INLG 2000*, 171–178.

Chomsky, N. 1957. *Syntactic Structures*. Gravenhage, Holland: Mouton.

Chomsky, N. 1959. *Review of Skinner's Verbal Behavior*. *Language*, *35*, 26–58.

Clark, H. H. 1992. *Arenas of Language Use*. Chicago: University of Chicago Press.

Clark, H. H. 1996. *Using Language*. New York: Cambridge University Press.

Clark, H. H., and Wilkes-Gibbs, D. 1986. Referring as a collaborative process. *Cognition*, *22*, 1–39.

Clifton, C., Frazier, L., and Rayner, K., eds. 1994. *Perspectives on Sentence Processing*. Hillsdale, NJ: Erlbaum.

Cooper, R. M. 1974. The control of eye fixation by the meaning of spoken language. A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *6*, 84–107.

Crain, S., and Steedman, M. J. 1985. On not being led up the garden path: The use of context by the psychological parser. In D. Dowty, L. Karttunen, and A. Zwicky, eds., *Natural Language Parsing: Psychological, Computational, and Theoretical Perspectives*. Cambridge, UK: Cambridge University Press.

Eberhard, K. M. 1998. Watching speakers speak: Using eye movements to study language production. Invited paper presented at the Seventieth Annual Meeting of the Midwestern Psychological Association, Chicago.

Fodor, J. A. 1983. *The Modularity of Mind*. Cambridge, MA: MIT Press.

Frazier, L. 1989. Against lexical generation of syntax. In W. D. Marslen-Wilson, ed., *Lexical Representation and Process*. Cambridge, MA: MIT Press.

Garfield, J. L., ed. 1987. *Modularity in Knowledge Representation and Natural-Language Understanding*. Cambridge, MA: MIT Press.

Grice, P. 1957. Meaning. *Philosophical Review*, *66*, 377–388.

Griffin, Z. M., and Bock, K. 2000. What the eyes say about speaking. *Psychological Science*, *11(4)*, 274–279.

Hanna, J. E., Tanenhaus, M. K., and Trueswell, J. C. 2003. The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, *49*, 43–61.

Heeman, P., and Allen, J. 1999. Speech repairs, intonational phrases, and discourse markers: Modeling speakers' utterances in spoken dialog. *Computational Linguistics*, *25(4)*, 527–571.

Keysar, B., Barr, D. J., Balin, J. A., and Paek, T. S. 1998. Definite reference and mutual knowledge: Process models of common ground in comprehension. *Journal of Memory and Language*, *39*, 1–20.

Keysar, B., Barr, D. J., and Horton, W. S. 1998. The egocentric basis of language use: Insights from a processing approach. *Current Directions in Psychological Sciences*, *7*, 46–50.

Krauss, R. M., and Weinheimer, S. 1966. Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, *4*, 343–346.

MacDonald, M. C., Pearlmutter, N. J., and Seidenberg, M. S. 1994. The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*, 676–703.

Meyer, D. E., and Schvanveldt, R. W. 1971. Facilitation in recognizing words: Evidence of dependence upon retrieval operations. *Journal of Experimental Psychology*, *90*, 277–234.

Miller, G. 1962. Some psychological studies of grammar. *American Psychologist*, *17*, 748–762.

Schegloff, E. A., and Sacks, H. 1973. Opening up closing. *Semiotica*, *8*, 289–327.

Searle, J. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge, UK: Cambridge University Press.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., and Carlson, G. N. 1999. Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*, 109–147.

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., and Sedivy, J. C. 2002. Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, *45*, 447–481.

Swinney, D. A., Onifer, W., Prather, P., and Hirshkowitz, M. 1978. Semantic facilitation across sensory modalities in the processing of individual words and sentences. *Memory and Cognition*, *7*, 159–165.

Tanenhaus, M. K., Chambers, C. G., and Hanna, J. E. Forthcoming. Referential domains in spoken language comprehension: Using eye movements to bridge the product and action traditions. In J. M. Henderson and F. Ferreira, eds., *The Interface of Language, Vision, and Action: Eye Movements and the Visual World*. New York: Psychology Press.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., and Sedivy, J. C. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632–1634.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., and Sedivy, J. C. 1996. Using eye-movements to study spoken language comprehension: Evidence for visually-mediated incremental interpretation. In T. Inui and J. McClelland, eds., *Attention and Performance XVI: Integration in Perception and Communication*, 457–478. Cambridge, MA: MIT Press.

Tanenhaus, M. K., and Trueswell, J. C. 1995. Sentence comprehension. In J. L. Eimas and P. D. Miller, eds., *Handbook in Perception and Cognition, Volume 11: Speech, Language, and Communication*, 217–262. New York: Academic Press.

Trueswell, J. C., Sekerina, I., Hill, N. M., and Logrip, M. L. 1999. The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, *73(2)*, 89–134.