Figurative, spontaneous, interactive and potentially offensive: Three projects with rich visual and linguistic stimuli

Cognitive Science 2006 Workshop: What have eye movements told us so far, and what is next?

Daniel C. Richardson (dcr@ucsc.edu**)**
Department of Psychology, University of California, Santa Cruz
273 Social Sciences 2, Santa Cruz, CA 95064

Teenie Matlock (tmatlock@ucmerced.edu)
Cognitive Science Program, UC Merced
PO Box 2039, University of California, Merced, CA 95344

Jennifer Randall Crosby (jlr@psych.stanford.edu)
Psychology Department, Stanford University
Stanford, CA 94305

Rick Dale (rdale@mail.psyc.memphis.edu)
University of Memphis, Department of Psychology
3693 Norriswood Ave., Memphis, TN 38152-6400

Abstract

We describe three independent projects with novel applications of eye movement technology. In each, we have sought to expand the range of stimuli and tasks typically used in eye movement research, and have been rewarded with a number of interesting theoretical findings. We present pictures, figurative speech, and videos to participants, who watch the displays, form opinions, have discussions and play games. The first two projects use standard looking time measures. In the first we examined how people process figurative speech and other forms of implicit spatial language. In the second we investigated why people look at members of a minority group when forming their opinions about potentially offensive remarks. In the third project we use a mathematical technique called cross recurrence analysis to quantify the temporal coupling between two people's eye movements. We eye track two conversants simultaneously while they talk about TV, art, politics and match ambiguous figures. We are making interesting discoveries about the role of common ground and the coordination of visual attention. On the basis of these three projects, we argue that eye movement research can employ rich, ecologically valid tasks and stimuli yet still yield rigorous empirical results.

Introduction

Studies of spoken language and eye movements typically

Focus on literal language
Present static images or scenes
Look at eye movements to objects not people
Use speech that is a terse, scripted monologue

This not the content nor the context of our everyday language use (Clark, 1996).

Note we are not making the criticism that typical eye movement research is not ecologically valid, and therefore its conclusions are limited. Far from it. In the spirit of this workshop, our point is that the range of stimuli and tasks in eye movement research, and hence the range of theoretical questions, can be dramatically broadened. Here we present three initial forays into figurative, potentially offensive, interactive and spontaneous language use, and argue that eye movements can provide rich theoretical insights.

Figurative language

Even though figurative language is pervasive in all cultures and all settings (Gibbs, 1994), eye movement research has focused on literal language. In recent work, we explored how figurative language would affect the process through which we perceive the world. In one project, we investigated how a scene would be perceived when it was described by forms of literal and figurative language that are reported to have equivalent meaning. We reasoned that any differences in eye movement patterns would tell us about both the distinct mental representations that are evoked by figurative language, and the scope of the integration between visual and verbal processing. We chose to examine fictive motion, a pervasive form of figurative language in English and other languages

(1a) The road runs through the desert
(1b) The fence follows the river

These descriptions are figurative because they contain a motion verb but describe no motion (Talmy, 2000). On the surface, fictive motion (FM) descriptions are equivalent to literal spatial descriptions (non-FM) such as

(2a) The road is in the desert
(2b) The fence is next to the river

Evidence from reading times, temporal judgments, and drawing studies suggests that FM descriptions engage motion representations (Matlock, 2004; Matlock, Ramscar, & Boroditsky, 2005). Given this, how would comprehending FM descriptions interact with visual processing?

Matlock and Richardson (2004) presented participants with simple drawings of paths (linear objects) such as roads, rivers and pipelines and tracked their gaze. The same scene was shown to participants as they heard either an FM or non-FM description of the path, counter balanced between participants (Figure 1). The FM and non-FM sentences were of equivalent length, and were judged by an independent set of participants to have equivalent meaning. The FM descriptions caused participants to have a longer gaze duration within the region of the path. One could argue, of course, that FM descriptions are simply more interesting forms of speech, and caused participants to be generally more interested in the pictures in front of them. On the contrary, our recent work has found evidence that FM sentences specifically evoke representations of motion.

Reading time studies (Matlock 2004) found that participants were quicker to process fictive motion target sentences after reading about terrains that were easy to traverse (e.g., The valley was flat and smooth) versus terrains that were not (e.g., The valley was bumpy and uneven). Critically, there was no difference for comparable literal target sentences without fictive motion (e.g., The road is in the valley). Following this logic, we (Richardson & Matlock, in press) presented participants with a picture and descriptions of easy or difficult terrains and then FM sentences or non-FM sentences. Terrain information modulated looking behavior with FM sentences, but not non-FM sentences (Richardson & Matlock, in press). Specifically, difficult terrain information and FM sentences produced longer gaze durations within the region of the path, and more saccades between points along the path.

FM: *The palm trees run along the road*
NFM: *The palm trees are next to the road*

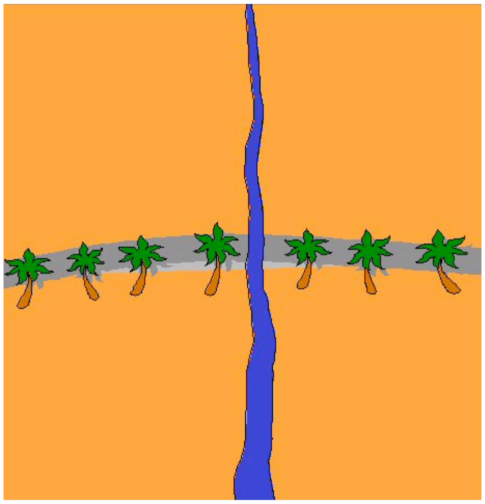

Figure 1. Example scene and spoken descriptions from Matlock & Richardson (2004)

Fictive motion descriptions drive our eyes across a visual image. We found that figurative language can evoke mental representations distinct from those of equivalent literal sentences, and these representations immediately interact with visual processing. We class figurative language as one form of implicit spatial language. Unlike explicit spatial language (e.g., X is above Y) or explicit referential language (e.g., Pick up the cube), implicit spatial information arises indirectly through implication, association, or metaphor. In ongoing research, we are using the lens of implicit spatial language to view the integration of language and vision.

Potentially offensive language

Imagine (or remember) being the only member of social group in the room. In everyone's earshot, a person makes a remark about your social group that borders on the offensive. What happens at this point? All eyes in the room turn to you. If you have ever experienced this, it is doubly unpleasant. Not only has your social group been besmirched, but suddenly you are the center of attention.

Why does this situation arise? One possibility is that when a potentially offensive remark is made, people practice social referencing - they determine if discrimination has occurred by measuring their own reaction against the reaction of an individual with perceived standing. We used eye movement research to find out if this anecdotal experience is a reliable phenomena, and to investigate the social referencing hypothesis (Crosby, Monin & Richardson, in submission).

In our experiment, participants were eye tracked as they watched a video of four males (three White and one Black) discussing university admissions. All four discussants were visible at all times. As one discussant voiced strong opinions against affirmative action, we measured the amount of time participants looked at the other discussants (Figure 2). If our anecdotal situation holds true, there will be more looks to the black individual at this point.

Of course, there are many strands of eye movement research that would make this prediction. Participants could direct their gaze towards individuals simply on the basis of any association between what is being said and what is in front of them. For example, if someone says that "the economy is in the red" and an individual is wearing a red shirt, we may look to this person simply because they fit into an accessible category. Eye movements often reveal such 'low-level' effects in which words, or parts of words, can trigger looks to potential referents in a scene (Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy, 1995), even when those part words are names from a different language (Spivey & Marian, 1999), or the referents have been removed and the locations are empty (Richardson & Spivey, 2000; Spivey & Geng, 2001). From this perspective, it would not be surprising at all if any discussion of racial issues would be enough to cause an eye movement to a black individual. We termed this possibility the 'association hypothesis'. In contrast, the 'social referencing hypothesis' holds that the minority individual is not looked at simply because they are broadly associated to

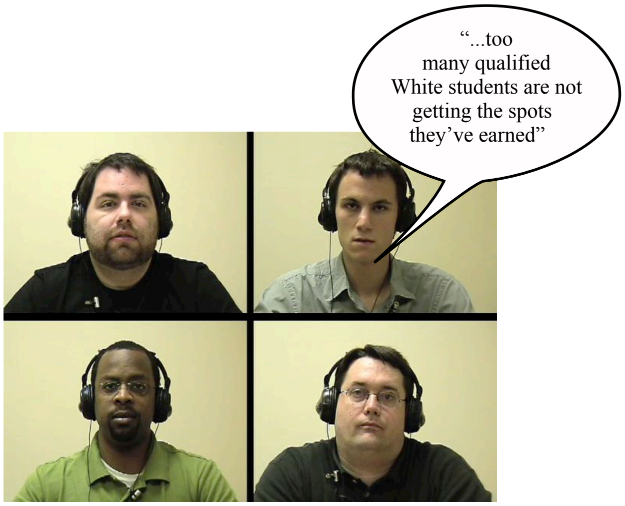
"...too many qualified White students are not getting the spots they've earned"

Figure 2. Potentially offensive comment (Crosby, Monin & Richardson, in submission)

the discussion, but specifically because the participants are seeking information about the potential offensiveness of the remark.

We distinguished these two hypothesis by means of a two experimental conditions. In an introductory passage it was established that either all participants could hear each other (four person condition), or that the bottom two participants (which included the minority individual) had their headphones turned off (two person condition). Importantly, the conditions were identical once the discussion of affirmative action began. Whilst the association hypothesis predicts that the black individual would be looked at more in both conditions, the social reference hypothesis predicts this only in the four person condition, when he can hear the potentially offensive remark and make a potentially informative reaction.

We found that participants spent dramatically longer looking at the Black individual if and only if he could hear the potentially offensive comments. Participants showed no interest in this individual in the two person condition when they believed he could not hear what was being said. The simple 'association hypothesis' was disproven. Instead, we have strong behavioural evidence that members of a minority will be looked at during instances of suspected discrimination when it is possible that they provide an informative response.

This is a surprising result in the context of recent claims in the eye movement literature. Some researchers have suggested that listeners have a surprisingly shallow awareness of interlocutor's mental states (Keysar, Barr, Balin & Brauner, 2000). In contrast, we have found sharp differences in the way that identical video images are inspected that depend on participants' reasoning about an individual's knowledge state, and their reaction to socially loaded information. This indicates that participants' eye movements are influenced by a range of subtle linguistic and interpersonal factors (Hanna, Tanenhaus, & Trueswell, 2003; Metzing and Brennan, 2003).

Interactive and Spontaneous language

Imagine an argument over a map, a debate over a proof written out on a black board, or a civilized conversation about a painting at a gallery. In these cases, the stream of speech will be punctuated by hand waving and pointing to the shared visual scene, and perhaps even grabbing the map and turning it the right way up. During the 'joint activity' of language use (Clark, 1996), conversants will use many such means to coordinate their visual attention.

The relationship between language use and visual attention has typically been studied by one of two approaches. One set of researchers have used eye movement technology to explore the link between a speakers' eye movements and their language production (e.g., Tanenhaus et al 1995), and a listener's eye movements and their language comprehension (e.g. Griffin & Bock, 2000). The other set of researchers have studied interaction between participants and have focussed on the actions they use to coordinate attention, such as gestures and pointing (Bangerter, 2004; Clark & Krych, 2004).

Our approach to studying visual attention and language use is different. In contrast to the first approach, we do not track an individual's eye movements, but record the eye movements of two participants while they discuss a shared visual scene. In contrast to the second set of researchers, we do not measure the actions participants make to coordinate attention, we measured the coordination of attention itself. Using cross recurrence analysis we quantify the temporal coupling between the conversants' eye movements.

This approach allows us to investigate a number of interesting questions. In our paradigm, the conversants cannot see each other, and hence cannot use pointing actions to coordinate their attention. Nevertheless, will their visual attention be coupled? Previous research has found reliable links between an individual's eye movements and their language comprehension and production in the case of short sentences (e.g., Griffin & Bock, 2000; Tanenhaus et al 1995). Will these results generalize to cases of extended, spontaneous speech between two people? If so, what factors enable conversants to coordinate their visual attention by verbal means?

We began answering these questions using monologue version of our task (Richardson & Dale, 2005) recorded the speech and eye movements of one set of participants as they looked at pictures of six cast members of a TV sitcom (either 'Friends' or 'The Simpsons'). They spoke spontaneously about their favourite episode and characters. One-minute segments were chosen and then played back unedited to a separate set of participants. The listeners looked at the same visual display of the cast members, and their eye movements were recorded as they listened to the segments of speech. They then answered a series of comprehension questions.

Cross-recurrence analysis (Zbilut, Giuliani, & Webber, 1998) quantified the degree to which speaker and listener eye positions overlapped at successive time lags. This speaker X listener distribution of fixations was compared to a speaker X randomized-listener distribution, produced by shuffling the temporal order of each listener's eye

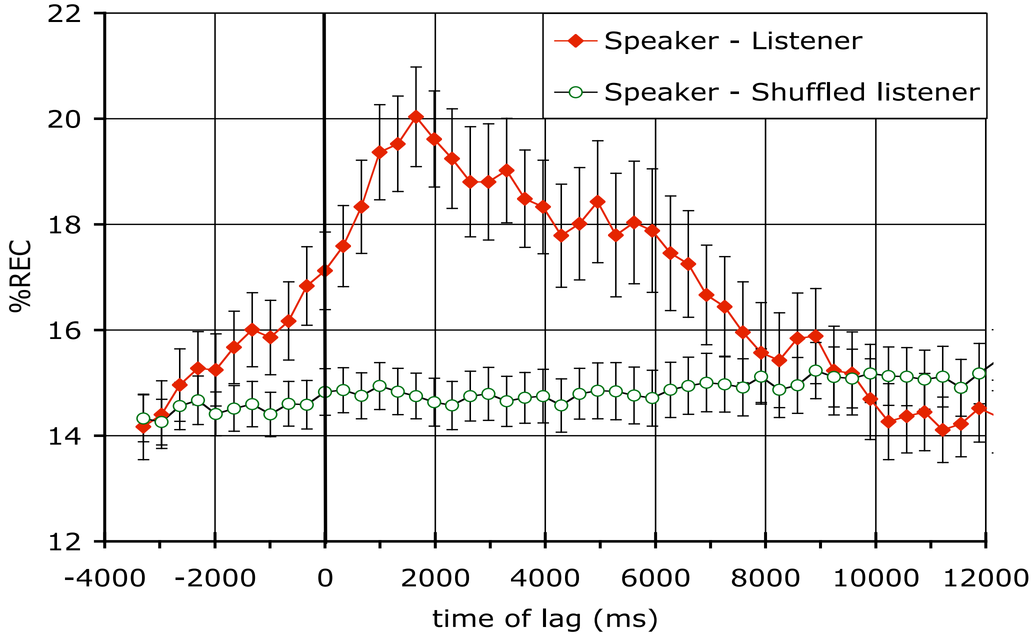

Figure 3. Cross-recurrence at different time lags between speaker and listener (Richardson & Dale, 2005)

movement sequence and then calculating the cross recurrence with the speaker.

From the moment a speaker looks at a picture, and for the following six seconds, a listener was more likely than chance to be looking at that same picture (Figure 3). The overlap between speaker and listener eye movements peaked at about 2000ms. In other words, two seconds after the speaker looked at a cast member, the listener was most likely to be looking at the same cast member. The timing of this peak roughly corresponds to results in the speech production and comprehension literatures. Speakers will fixate objects 800-1000ms (Griffin & Bock, 2000) before naming them, and listeners will typically take 500-1000ms to fixate an object from the word onset (Allopenna et al., 1998). Planning diverse types of speech appears to systematically influence the speaker's eye movements, and a few seconds later, hearing them will influence the listener's eye movements.

Importantly, this coupling of eye-movements between speaker and listener was not merely an epiphenomenal by-product of conversation. The cross-recurrence between individual speaker-listener pairs reliably predicted how many of the comprehension questions the listener answered correctly. This correlation was supported by a follow-up study that experimentally manipulated the relationship between speaker and listener eye movements. We found that by flashing the pictures in time with the speakers' fixations (or a randomized version) we caused the listeners' eye movements look more (or less) like the speakers', and influenced the listeners' performance on comprehension questions.

Though the language use in Richardson and Dale's (2005) study was spontaneous, it lacked a key element of everyday conversations - interaction. In a second set of studies, we tracked the gaze of two conversants simultaneously while they discussed TV shows, politics and surreal paintings. The results of some of these studies will presented in detail elsewhere during this conference (Richardson & Dale, 2006). We found that in the case of a live, interactive dialogue, conversants' eye movements continued to be coupled as they looked at a shared visual display. This coupling peaked at a lag of 0ms. In other words, the conversants were most likely to be looking at the same thing at the same point in time. As in the monologue results, this coupling was at above chance levels for a period of around six seconds, suggesting that conversants may keep track of a subset of the depicted people who are relevant moment-by-moment (Brown-Schmidt et al., 2004). We demonstrated experimentally that this coupling was related to the degree of knowledge that participants shared. Coordination of attention increased if prior to a discussion of a painting, participants heard the same (versus different) background information.

In further studies, we are investigating how such common ground information might be created between conversants. Participants took part in three rounds of the tangram matching task (Clark & Brennan, 1991). They saw the same six abstract, humanoid shapes in different orders. One participant was instructed to describe his shapes in turn so that the other could find them. In the first round, participants typically established descriptors of the ambiguous shapes (e.g. 'the dancer', 'the skier'). This process of grounding and confirming descriptors is reflected in the eye movement recurrence. Typically, eye movement couplings increased during a trial until the matcher was fixating the right shape. At that point, a descriptor would be proposed. For the rest of the trial, the eye movement coupling decreased as both director and matcher looked at around at other shapes to see if the descriptor was a good one. In later rounds, these established 'conceptual pacts' task (Clark & Brennan, 1991) provided a quicker way to find the shapes, and eye movement recurrence peaked more quickly.

In all of our studies, eye movement couplings reveal an intimate relationship between discourse processes, attentional processes and the visual common ground. Just as eye movements reflect the mental state of an individual, the coupling between conversants eye movements reflects the success of their communication.

Conclusion

Psycholinguistics has profited greatly from eye movement research. It has allowed us to bridge the language as action and language as product traditions (Tanenhaus & Trueswell, 2004), and revealed the timecourse of particular types of language processing. We believe that eye movement research has even more to offer. Linguistic stimuli need not be literal descriptions or instructions. Figurative language has its own eye movement signature. Visual stimuli need not be just static scenes or arrays objects. Video stimuli can be used to explore how a listener considers the perspective and predicts the actions of others. Language use is more than individuals speaking or listening to monologues; it is a complex interaction between people. Eye movement techniques can capture this joint activity directly, by quantifying the coordination of conversants' attention.

Acknowledgments

The authors are indebted to Herbert H. Clark, Natasha Kirkham and Michael Spivey.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language. 38(4)*, 419-439.

Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science, 15(6)*, 415-419.

Brown-Schmidt, S., Campana, E., & Tanenhaus, M. K. (2004). Real-time reference resolution by naïve participants during a task-based unscripted conversation. In J. C. Trueswell & M. K. Tanenhaus (Eds.), *World-situated language processing: Bridging the language as product and language as action traditions.* Cambridge: MIT Press.

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

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127-149). Washington, DC: APA.

Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory & Language,* 50(1), 62-81.

Eberhard, K., Spivey-Knowlton, M., Sedivy, J. & Tanenhaus, M. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research,* 24, 409-436.

Gibbs, R. W., Jr. (1994). *The poetics of mind: Figurative thought, language, and understanding*. New York, NY: Cambridge University Press.

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

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory & Language, 49(1*), 43-61.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science,* 11 (1), 32-38.

Matlock, T. (2004). Fictive motion as cognitive simulation. *Memory & Cognition,* 32, 1389-1400.

Matlock, T., Ramscar, M., & Boroditsky, L. (2005). The experiential link between spatial and temporal language. Cognitive Science, 29, 655-664.

Richardson, D.C. & Dale, R. (2005). Looking To Understand: The Coupling Between Speakers' and Listeners' Eye Movements and its Relationship to Discourse Comprehension. *Cognitive Science, 29,* 1045–1060.

Richardson, D.C. & Dale, R. (2006). Grounding dialogue: eye movements reveal the coordination of attention during conversation and the effects of common ground. *Proceedings of the 28th Annual Cognitive Science Society Conference*

Richardson, D.C. & Matlock, T. (in press). The integration of figurative language and static depictions: An eye movement study of fictive motion. *Cognition*

Spivey, M. & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science,* 10, 281-284.

Talmy, L. (2000). *Toward a cognitive semantics (Volume 1: Concept structuring systems)*. Cambridge, MA, US: The MIT Press.

Tanenhaus, M. K., Spivey Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268(5217)*, 1632-1634.

Trueswell & J. C. Tanenhaus M. K. (2004), *World-situated language processing: Bridging the language as product and language as action traditions*. Cambridge: MIT Press.

Zbilut, J. P., Giuliani, A., & Webber, C. L., Jr. (1998). Detecting deterministic signals in exceptionally noisy environments using cross-recurrence quantification. *Physics Letters,* 246, 122-128.