

(Beyond) referential mechanisms in spatial language comprehension

Everyday interaction often involves comprehending spatial language such as when trying to check the time and being told “the clock is *on* the table”. It is well established that spatial language processing requires attention mechanisms (Carlson & Logan, 2005; Logan, 1994) but how precisely people deploy visual attention during real-time spatial language comprehension, is still unclear.

The Attention Vector Sum (AVS) model postulates that to comprehend sentence (1) people must shift their attention from the vase (‘reference object’) to the clock (located object, e.g., Carlson-Radvansky & Irwin, 1994; Carlson & Logan, 2005, Regier & Carlson, 2001). An alternative account from ‘visual world’ studies suggests people incrementally inspect objects as they are mentioned and thus for (1) inspect the clock followed by the vase (e.g., Tanenhaus et al., 1995). In sum, these two accounts predict opposing inspection orders although the visual world (but not the AVS) account specifies the time course of visual attention allocation. We examine gaze pattern to objects during spatial language comprehension, and evaluate their fit against predictions of the AVS (reference object -> located object) and visual world (located object -> reference object) accounts.

(1) “The clock is above the vase”.

We recorded eye movements while people listened to spatial descriptions (e.g., (1)) and verified whether the sentence matched (vs. didn’t match) the picture. We analysed fixations and inspections (consecutive fixations to an object) for the matching picture-sentence pairs. Shortly after people heard “above” they fixated the vase more often than the clock, corroborating the visual world account. In contrast, analyses of inspections show that people after hearing “above”, and after one inspection to the vase, look next to the clock on 70 percent of inspections. The distribution of fixations during “the vase” confirms this view, in that 30 percent of fixations are directed at the clock (vs. 60 percent to the vase vs. 10 percent to a third unrelated distracter object). In sum, gaze analyses after “above” revealed that (a) people anticipate the post-verbal object as predicted by the visual world account, but (b) after inspecting the vase, they next inspect the clock. While the AVS model cannot accommodate findings (a), the visual world account alone cannot accommodate findings (b), suggesting we need aspects of both accounts to accommodate the data.

Carlson, L. A. & Logan, G. D. (2005). *Attention and spatial language*. In L. Itti, G. Rees, & J. Tsotsos (Eds.), *Neurobiology of attention* (pp. 330-336). San Diego, CA: Elsevier.

Carlson-Radvansky, L. A. & Irwin, D. E. (1994). Reference frame activation during spatial term assignment. *Journal of Memory and Language*, 33, 646-671.

Logan, G. (1994). Spatial attention and the apprehension of spatial relations. *Journal of Experimental Psychology: HPP*, 1015-1036.

Regier, T. & Carlson, L. (2001). Grounding spatial language in perception: an empirical and computational investigation. *Journal of Experimental Psychology: General*, 130, 273-298.

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

Keywords: eye-tracking; language comprehension, semantics

Word count: 498