SPATIAL LANGUAGE COMPREHENSION: A COMPUTATIONAL INVESTIGATION OF THE DIRECTIONALITY OF ATTENTION

Thomas Kluth, Michele Burigo, Pia Knoeferle (CITEC: Cognitive Interaction Technology – Center of Excellence, Language & Cognition Group, University of Bielefeld)
tkluth@cit-ec.uni-bielefeld.de

How do people understand spatial language? How is the understanding of spatial language reflected in perceptual processes like visual attention?

In the visual world paradigm, participants' eye movements on a display are tracked while they listen to spoken utterances. Consider you are seeing a box above a sausage and hearing "The box is above the sausage". As you hear this sentence, your eye movements (i.e., your overt attention) will shift from the box to the sausage, as has been found by Burigo and Knoeferle (2015). That is, visual attention shifts from the *located object* (LO, the box) to the *reference object* (RO, the sausage). These findings from the visual world paradigm (see also Chambers et al., 2002) are consistent with studies in the field of cognitive neuroscience that have also found an attentional shift from the LO to the RO (e.g., Roth & Franconeri, 2012).

Contrary to these findings, researchers in the domain of spatial relation processing have argued that attention shifts from the RO to the LO, i.e., in the opposite direction (e.g., Logan & Sadler, 1996).[1] The directionality of this shift is reflected in the *Attentional Vector Sum* (AVS) model developed by Regier and Carlson (2001), which generates acceptability ratings of projective spatial prepositions given a RO, a LO, and a spatial preposition. The rating is computed as follows: First, attention is focused on the RO, resulting in an attentional distribution with an exponential decay. Then, at every point in the RO a vector is rooted, pointing to the LO and weighted by the amount of attention at this point. Next, all of these vectors are summed up. The vector thus obtained is compared to canonical upright (in the case of *above*): The higher the deviation, the lower the acceptability rating.

To computationally assess the importance of the directionality of the attention shift postulated by the AVS model, we have modified the direction of the attentional shift in the AVS model (that now goes from the LO to the RO) while maintaining the basic mechanism of the model – yielding the *reversed* AVS model (rAVS)[2]. We compared the ability of the AVS and the rAVS model to replicate empirical data. These data came from rating studies with which the AVS model was assessed (Logan & Sadler, 1996; Regier & Carlson, 2001). The rAVS model performed as well as the AVS model, i.e., both models were able to closely fit the empirical acceptability ratings as reflected by similar goodness-of-fit and simple hold-out (Schultheis et al., 2013) results. Thus, our model simulations do not favor any of the two models and accordingly, both directionalities of the attentional shift are equally supported.

Furthermore, the rAVS model suggests further details for the understanding of spatial prepositions. The proximal orientation and the center-of-mass orientation are known to affect the acceptability of spatial prepositions. The proximal orientation is the orientation of a line connecting the LO with the most proximal point on the RO, whereas the center-of-mass orientation is the orientation of a line connecting the LO with the center-of-mass of the RO. The AVS model incorporates both orientations within its vector sum. In the rAVS model, however, a *single* vector points either to the proximal point of the RO, the center-of-mass of the RO or in-between these two points (depending on the relative distance between the RO and the LO). As a consequence, the rAVS model has a lower computational complexity than the AVS model. Our simulations show that a single vector pointing from the LO to the RO captures all known effects on the acceptability of spatial prepositions without relying on a vector sum.

[1] Burigo and Knoeferle (2015) sometimes also found shifts from the RO to the LO, but these were not as frequent as shifts from the LO to the RO.

[2] The source code can be found at `https://bitbucket.org/kluth/avs`

References

Burigo, M., & Knoeferle, P. (2015). Visual Attention during Spatial Language Comprehension. *PloS ONE*, *10*(1), e0115758. doi: 10.1371/journal.pone.0115758

Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H., & Carlson, G. N. (2002). Circumscribing referential domains during real-time language comprehension. *Journal of Memory and Language*, *47*(1), 30–49.

Logan, G. D., & Sadler, D. D. (1996). A computational analysis of the apprehension of spatial relations. In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garrett (Eds.), *Language and Space* (pp. 493–530). The MIT Press.

Regier, T., & Carlson, L. A. (2001). Grounding spatial language in perception: An empirical and computational investigation. *Journal of Experimental Psychology: General*, *130*(2), 273–298.

Roth, J. C., & Franconeri, S. L. (2012, November). Asymmetric coding of categorical spatial relations in both language and vision. *Frontiers in Psychology*, *3*(464). doi: 10.3389/fpsyg.2012.00464

Schultheis, H., Singhaniya, A., & Chaplot, D. S. (2013). Comparing model comparison methods. In *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 1294 – 1299). Austin, TX: Cognitive Science Society.