Language and Vision: A Case Study of Interaction between Two Systems

Banchiamalck Dessalegn (banchi@cogsci.jhu.edu) & Barbara Landau (landau@cogsci.jhu.edu)

Cognitive Science Department, 3400 N Charles Street
Baltimore, MD 21218 USA

Keywords: Vision; Language; feature binding

Introduction

Cognitive systems that are thought to be modular produce outputs that interact. Consider vision and language. Although many theorists argue that vision and language are modular, it is also true that their outputs must interact enabling us, e.g., to talk about what we see. What are the consequences of such interactions? For example, when language and vision interact is the resulting representation qualitatively different from representations that are formed without an interaction?

On several accounts language plays a powerful role in altering representations by, e.g., automatically heightening attention to key visual properties (Smith, Jones &Landau, 1996) or allowing abstract analysis (Loewenstein & Gentner, 2005).

In the current experiments we investigated the role of language for a specific language-vision interaction. The specific case we considered was binding of visual features. Many have argued that binding of visual features, e.g. color and location, requires focused attention (for a review see Treisman, 1999). Thus, when children fail to bind visual features, we ask whether language helps, and if so, how?

Methods, Results, and Discussion

Four year-olds saw squares split by color and matched them after a delay. In Experiment 1, children saw the target and were told either "Look, this is a *blicket*" (Label Condition, N=12) or "Look!" (NoLabel Condition, N=12). Then, three choices appeared: the target replica, its mirror image, and another square that had a different internal split.

Overall, children performed better than chance (percent correct M = 65%). However, errors were almost exclusively mirror image, suggesting that children failed to maintain conjunction of color and location. There was no difference between the NoLabel and Label conditions; i.e. novel labels did not help children maintain the conjunctions of features.

Given that the results in Experiment 1 reflect failure to robustly encode the conjunction of color & location, Experiment 2 (N=24) asked whether sentences with *directional* terms, that specified the color & location of a part, would help. Experiment 2 was the same as Experiment 1 except that children were told, e.g. "The red is on the left." Performance was reliably better than Experiment 1, suggesting that these sentences with directional terms do enhance the maintenance of conjunctions.

Additional experiments showed that neither *neutral* relational labels (e.g. "The red is touching the green") nor *nonlinguistic* attentional manipulations (e.g., flashing the red part, making the red part bigger) improved performance.

Conclusion

Our findings show that sentences with directional linguistic terms can help children bind visual properties, and that this effect is not attributable to general attentional effects or to the relational nature of the terms alone.

We propose that the visual representation of the target, in combination with the linguistic labels, is used to accurately maintain the conjunction of color and location. One possibility is that long-term, stable knowledge of the linguistic labels can be used to maintain the conjunction. This would work for children who have a solid long-term representations of the terms left/right. For these children, we propose that the directional labels provide the representational scheme necessary to robustly encode the conjunction of color & location, allowing children to overcome problems associated with fragile encoding of these features in short-term memory.

Another possibility is that children might use the directional terms as 'temporary directional pointers'. This would work if children have relatively fragile understanding of the terms left/right, but can use the temporary context of hearing the terms to hold onto the relationship. We found that 4 year-olds fell roughly equally into groups with strong and fragile knowledge of the terms left/right, suggesting support for both possible mechanisms. In summary, we show a specific case of vision-language interaction – language is used to supplement visual representations such that fleeting distinctions, e.g., among mirror images, are made more robust.

Acknowledgments

We thank James Hoffman and members of the Language and Cognition lab at JHU for helpful comments and discussions. This research was supported by March of Dimes Grant 04-46, and NSF IGERT Grant to Cognitive Science Dep., JHU.

References

Treisman, A. (1999). Solutions to the binding problem: Progress through controversy and convergence. *Neuron*, 24, 105-110.

Smith, L., Jones, S., & Landau, B. (1996). Naming in young children: A dumb attentional mechanism? *Cognition*, 60, 143-171.

Loewenstein, J., & Gentner, D. (2005). Relational language and the development of relational mapping. *Cognitive Psychology*, 50(4), 315.