Describing Objects in Visual Scenes: Is Visual Salience Like Conversational Salience?

March 1, 2013

When describing an object in a visual scene (referring expression generation; (Krahmer & Deemter, 2012), a speaker must devise an expression which allows a listener to quickly and accurately locate the target. Speakers often do so by describing both the target itself and other objects (landmarks); previous work has shown that objects chosen as landmarks are frequently visually salient– easy to find by visually searching the scene (Viethen & Dale, 2011). However, the relationship between visual salience and the syntax of the resulting linguistic description is still unclear. While objects that are salient because of recent conversational mentions are known to be described differently and in different positions (Prince, 1981), it is not clear whether the same patterns hold true for objects that are salient for purely visual reasons.

In this study, we present preliminary results on the syntax of a corpus of descriptions of cartoon people in crowded visual scenes drawn from the childrens' book series "Where's Wally". We evaluate visual salience using a computational model (Torralba, A. Oliva, & Henderson, 2006). We investigate speakers' strategies for selecting landmarks, describing them, and positioning them in the description. Finally, we draw comparisons between strategies for describing and positioning visually salient and conversationally salient objects.

References

Krahmer, E., & Deemter, K. van. (2012, March). Computational generation of referring expressions: A survey. *Computational Linguistics*, *38*(1), 173–218.

Prince, E. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical pragmatics* (pp. 223–255). New York: Academic Press.

Torralba, A., A. Oliva, M. C., & Henderson, J. M. (2006). Contextual guidance of attention in natural scenes: The role of global features on object search. *Psychological Review*, *113*, 766-786.

Viethen, H., & Dale, R. (2011). GRE3D7: A corpus of distinguishing descriptions for objects in visual scenes. In *Proceedings of the workshop on using corpora in natural language generation and evaluation.* Edinburgh, Scotland.