2018-06-14

Mind the Gap: Situated Spatial Language a Case-Study in Connecting Perception and Language

John Kelleher
Technological University Dublin, john.d.kelleher@tudublin.ie

Recommended Citation

Mind the gap: Situated spatial language a case-study in connecting perception and language. John D. Kelleher. In Proceedings of the Workshop on Dialogue and Perception 2018, page 9-11. Organized by CLASP the Centre for Linguistic Theory and Studies in Probability at the University of Gothenburg, Sweden. June 14-15, 2018.

Mind the Gap: Situated Spatial Language a Case-Study in Connecting Perception and Language

John D. Kelleher
ADAPT Research Centre,
Information, Communication, and Entertainment Institute,
Dublin Institute of Technology,
Ireland
john.d.kelleher@dit.ie

Situated language is spoken from a particular point of view within a simulated or physical context that is shared with an interlocutor. From theoretical linguistic and cognitive perspectives, situated dialog systems are interesting as they provide ideal testbeds for investigating the interaction between language and perception, at the same time there are a growing number of practical applications, for example robotic systems, where spoken interfaces, capable of situated dialog, promise many advantages (Kelleher, 2003) . An open challenge in this domain is the creation of computational models that appropriately ground the semantics of spatial terms within the shared perceptual context. This is partly because of the diversity of factors that impinge on spatial term semantics, including *geometry*, *world knowledge* (including functional roles and object dynamics), and human *perception*.

Many computational models of spatial semantics are based on the concept of a *spatial template* (Logan and Sadler, 1996). This standard model has been extended in a number of ways. For example, to include frame of reference ambiguity (Kelleher and Costello, 2005; Kelleher and van Genabith, 2006; Dobnik et al., 2014); the impact of distractor objects within the scene (Kelleher and Kruijff, 2005; Costello and Kelleher, 2006; Kelleher and Costello, 2009); and to include the role of human attention and visual perceptual factors in spatial reference resolution (Kelleher et al., 2005; Kelleher, 2006; Regier and Carlson, 2001; Kelleher et al., 2010). At the same time, other research has used corpus based analytics to explore the functional and geometric semantics of prepositions in visually situated spatial reference (Dobnik and Kelleher, 2014; Dobnik et al., 2018). However, to-date relatively little work has been focused on developing an integrated model that accommodates all of these factors.

In recent years, however, deep learning approaches have made significant breakthroughs in a number of areas. An exciting aspect of deep learning is the concept of representation learning from data. In particular, learning the projection of naturally discrete information (e.g. words) into continuous representations (e.g. word embeddings), and also learning vector based inter/multi-modal representations, such as those used in automatic image captioning systems. A number of shortcomings with current deep learning architectures have been identified with respect to their application to spatial language (Kelleher and Dobnik, 2017). However, adopting a modular mechanistic approach to training deep networks may offer a solution to these challenges (Dobnik and Kelleher, 2017).

In light of this, in this paper will review the literature on computational models of spatial semantics and the potential of deep learning models as an useful approach to this challenge.

References

Fintan Costello and John D. Kelleher. 2006. Spatial prepositions in context: The semantics of *Near* in the presense of distractor objects. In *Proceedings of the 3rd ACL-Sigsem Workshop on Prepositions*, pages 1–8.

Simon Dobnik, Mehdi Ghanimifard, and John D. Kelleher. 2018. Exploring the functional and geometric bias of spatial relations using neural language models. In *Proceedings of the First International Workshop on Spatial Language Understanding (SpLU-2018) at the 2018 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL-HLT)*.

Simon Dobnik and John D. Kelleher. 2014. Exploration of functional semantics of prepositions from corpora of descriptions of visual scenes. In *Proc. of the Workshop on Vision and Language*, pages 33–37.

Simon Dobnik and John D Kelleher. 2017. Modular mechanistic networks: On bridging mechanistic and phenomenological models with deep neural networks in natural language processing. *CLASP Papers in Computational Linguistics*, page 1.

Simon Dobnik, John D. Kelleher, and Christos Koniaris. 2014. Priming and alignment of frame of reference in situated conversation. *Proceedings of Dial-Watt-Semdial*, pages 43–52.

John D. Kelleher. 2003. *A perceptually based computational framework for the interpretation of spatial language*. Ph.D. thesis, Dublin City University.

John D. Kelleher. 2006. Attention driven reference resolution in multimodal contexts. *Artificial Intelligence Review*, 25(1):21–35.

John D. Kelleher and Fintan Costello. 2005. Cognitive representations of projective prepositions. In *Proceedings of the Second ACL-Sigsem Workshop of The Linguistic Dimensions of Prepositions and their Use in Computational Linguistic Formalisms and Applications*.

John D. Kelleher and Fintan Costello. 2009. Applying computational models of spatial prepositions to visually situated dialog. *Computational Linguistics*, 35(2):271–306.

John D. Kelleher, Fintan Costello, and Josef van Genabith. 2005. Dynamically structuring, updating and interrelating representations of visual and linguistic discourse context. *Artificial Intelligence*, 167(1):62–102.

John D Kelleher and Simon Dobnik. 2017. What is not where: the challenge of integrating spatial representations into deep learning architectures. *CLASP Papers in Computational Linguistics*, page 41.

John D. Kelleher and Josef van Genabith. 2006. A computational model of the referential semantics of projective prepositiosn. In *Syntax and Semantics of Prepositions*, pages 211–228. Springer.

John D. Kelleher and Geert-Jan Kruijff. 2005. A context-dependent model of proximity in physically situated environments. In *Proceedings of the 2nd ACL-SIGSEM Workshop on The Linguistic Dimensions of Prepositions and their use in Computational Linguistic Formalisms and Applications*.

John D. Kelleher, Robert J. Ross, Colm Sloan, and Brian Mac Namee. 2010. The effect of occlusion on the semantics of projective spatial terms: a case study in grounding language in perception. *Cognitive Processing*, 12(1):95–108.

G.D. Logan and D.D. Sadler. 1996. A computational analysis of the apprehension of spatial relations. In *Language and Space*, pages 493–530. MIT Press.

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