Generating and interpreting spatial relations

Group 2

To generate a spatial language, we need to consider factors other than the geometric information (the geometric representation of an object in space); these factors, as we understand from paper [1, 2] are as follows:

- Perceptual context: the phrase “to the left of” semantics depends on the intended perspective [3]; is it to the left of the speaker or the left of some grounded reference point.
- The functional geometric framework:
    - **Dynamic-kinematic routines**: how objects interact (kinematic) over time (dynamic). The phrase “an apple in a container” conveys the impression that the container contains and constrains the location of the apple over time [4], as shown in the apple/bowl experiment[5].
    - **Situational knowledge**: the learned experience of how the objects involved in the scene interact with each other or how they function affect both the visual and dynamic-kinematic routines. As we discussed above, containers have a function of constraining contents over time[4]. However, jugs are more associated with liquids, while bowls are associated with both liquids or solids. It is found that the propositional “in” is produced when a bowl contains a solid more than when a jug contains the same solid [4]. Both cases have the same geometric features.
    - Both of the dynamic-kinematic routines and situational knowledge formalize the **functional geometric framework**; the past learned situated interactions of the objects involved in the scene and how they interact over time establish the spatial language semantics [4].

In light of the above discussion, paper [1] tries to integrate the functional geometric framework in a system that predicts the propositional “above” and “over”, which are sensitive to the functional geometric framework. Paper [2] tries to integrate the geometric features extracted from a scene with the learned spatial language semantics from a corpus to predict the proposition connecting between the objects in the scene. It seems that paper [1] is more theoretically oriented towards cognitive science theories; that is why it is too limited to two propositions. Paper [2] focus more on the geometric effect. Still, one could argue that the effect of functional-geometric features, to some degree, could be learned from the image-corpus integration by an ML algorithm. But this, of course, depends on the algorithm itself and the dataset. We are unsure about the algorithms used in the paper [1]; however, could a deep neural network with attention mechanisms, like what we have discussed in the previous seminar, do that? How can we include the cognitive and language learning theory in such deep neural systems?

Paper [1] set-up a dynamic scene consists of a teapot pouring water into a cup with different geometric configurations. Such setup reflects the semantic of the propositional “above” and “over” within the functional geometric framework. The system consists of three modules:

1. Vision Processing module with an attention mechanism that able to process the interaction between the three objects from image sequences.
2. Elman Network; which is a recurrent network that takes the output of the vision processing unit to predict the objects’ dynamic; i.e. predict the objects’ states change from the image sequence.
3. Dual-Route Network; which is a dual feedforward neural network that learns from the objects’ dynamics and linguistics features the description for the scene or the description for each configuration of the objects.

Paper [2], detect entities in the scene using an object detector; the output of the object detector is used to extracts the geometric features. The textual embedding, CCN image features, and the calculated geometric feature are feed into classification algorithms to predict the proposition that combines the entities under study.

References:

1. Coventry, K. R., Cangelosi, A., Rajapakse, R., Bacon, A., Newstead, S., Joyce, D., & Richards, L. V. (2005). Spatial prepositions and vague quantifiers: Implementing the functional geometric framework. Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science), 3343, 98–110. https://doi.org/10.1007/978-3-540-32255-9_6

2. Ramisa, A., Wang, J., Lu, Y., Dellandrea, E., Moreno-Noguer, F., & Gaizauskas, R. (2015). Combining geometric, textual and visual features for predicting prepositions in image descriptions. Conference Proceedings - EMNLP 2015: Conference on Empirical Methods in Natural Language Processing, September, 214–220. https://doi.org/10.18653/v1/d15-1022

3. Dobnik, S., & Kelleher, J. D. (2018). Modular mechanistic networks: On bridging mechanistic and phenomenological models with deep neural networks in natural language processing. ArXiv, 1–18.

4. Fu, Z. (2016). Spatial Language Learning. Journal of Education and Practice, 7(8), 146–151.

5. Coventry, Kenny R., and Pedro Guijarro-Fuentes. “Spatial Language Learning and The Functional Geometric Framework.” Handbook of Cognitive Linguistics and Second Language Acquisition, First ed., Routledge, 2008.