View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Dagstuhl Research Online Publication Server

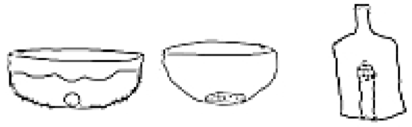
How can spatial language be learned?

Kenneth D. Forbus, Northwestern University

How languages are learned is one of the deepest mysteries of cognitive science. This question can be addressed from multiple perspectives. This position paper considers two of them: (1) How do people learn spatial language? (2) Given the wide range of spatial terms in language, how might we bootstrap the linguistic capabilities of intelligent systems that need spatial language to achieve wide and accurate coverage? We discuss each question in turn.

How do people learn spatial language?

Our hypothesis is that people learn spatial language via analogical generalization over qualitative spatial representations. The model of analogy we use is Gentner's (1983) structure-mapping theory, which describes analogy and similarity in terms of comparisons involving structured, relational representations. We simulate analogical matching via SME, the Structure-Mapping Engine. Analogical generalization is defined in the SAGE[1] model by using SME as a component. Every concept being learned by an organism, in this model, has a *generalization context* associated with it. A generalization context maintains a set of generalizations and a set of ungeneralized examples, and has a propositional *entry condition* associated with it. Roughly, SAGE works like this: Every incoming stimulus that matches the entry conditions of a generalization context G is added to G. When a new example E is added, SAGE sees if it is sufficiently similar to an existing generalization, using SME. If the similarity computed is over a threshold, that example is assimilated into that generalization. If E is not assimilated, then E is compared with examples in the list that G maintains. If it is close enough to one of them, a new generalization is formed. The formation of a generalization in either case consists of combining the matching facts and computing probabilities for them, based on frequency of occurrence in assimilated examples. When the probability of a fact gets low enough, it is dropped from the generalization. Thus facts which constitute noise eventually wither away, leaving only the strong commonalities in a generalization. Figure 1 illustrates.



Best Generalization IN
Size: 3
(*candle in bottle, cookie in bowl, marble in water*)
--DEFINITE FACTS:
(rcc8-TPP figure ground)
--POSSIBLE FACTS:
33%: (Basin ground)
33%: (Bowl-Generic ground)

Figure 1: Example of a generalization computed by SAGE for the English term *in*.

Note that the representations used in the generalizations are structured, relational statements. They include conceptual information (e.g., being a Basin or Bowl) but also qualitative spatial relationships (here, RCC8's Tangential Proper Part relationship). We believe that qualitative spatial representations

[1] SAGE = Sequential Analogical Generalization Engine, formerly known as SEQL. The simulation is the same, the name has been changed to prevent confusion with SQL.

Dagstuhl Seminar Proceedings 10131
Spatial Representation and Reasoning in Language : Ontologies and Logics of Space
http://drops.dagstuhl.de/opus/volltexte/2011/2733

are appropriate for modeling human spatial representations for several reasons. First, qualitative spatial representations have proven useful for conducting a wide variety of spatial reasoning, as illustrated by the efforts of the qualitative reasoning community. Second, human visual processing seems highly attuned to structured representations and qualitative distinctions. Finally, qualitative representations are a good fit for analogical reasoning and learning. The qualitative spatial representations we use are automatically computed by CogSketch (Forbus *et al* 2008), a sketch understanding system that models human visual and spatial processing.

SAGE has been used to model a variety of learning phenomena, including music classification, sketch classification, and counterterrorism. It has also been used to model learning of spatial prepositions (Lockwood et al 2008). Lockwood examined spatial prepositions of location which had been explored in a developmental study, specifically, the English prepositions *in* and *on*, and the Dutch prepositions *in*, *op*, *aan*, and *om*. One generalization context was created for each preposition, and examples of preposition use from the study were used to train the system. Classification was carried out by using SME to compare a new example against the generalizations and examples from each generalization context, selecting the context which had the most similar item as its answer (Figure 2). A leave-one-out cross validation was used to test accuracy, which was over 75% in all cases and statistically significant for all but English *in*.

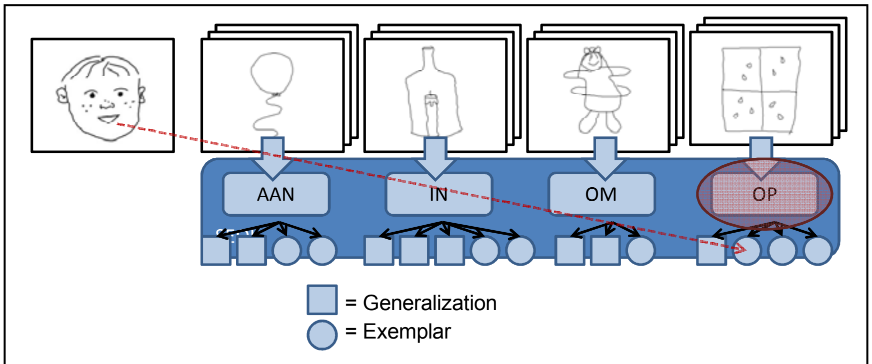


Figure 2: Classification of "freckles ? face" as *op*, based on SAGE's model

There are a number of limitations with this study, of course. First, it only involves a small number of prepositions, and only a few examples. (One advantage of SAGE is that it learns rapidly, typically needing orders of magnitude fewer examples than traditional statistical learning algorithms, and more like what is required for human learning.) Scaling up to a broader range of spatial terms, and with a larger range of examples, will undoubtedly raise new challenges. Second, the stimuli were sketched, instead of being visually processed. We postpone the modality issue until the next section. Third, the sketched stimuli only involved two objects, a figure and ground, both explicitly labeled. Learning from language describing complex scenes would both be more natural and very likely more difficult. Despite these limitations, we think this approach has potential as a model for human spatial language acquisition, and language acquisition more broadly.

How can we bootstrap the spatial language capabilities of intelligent systems?

Let us turn to the practical matter of building intelligent systems with spatial language. Developing a reusable resource for the meanings of spatial terms, just as WordNet is a reusable resource for lexical information and OpenCyc is a reusable resource for conceptual knowledge, seems like a very desirable goal. One reason to think that this can be done is that there is evidence suggesting that human spatial representations are *amodal* (Avraamides et al 2004). That is, initial levels of encoding are modality-specific (e.g., visual, haptic, auditory, linguistic), but these inputs are integrated into a common representation of space that is modality-independent. This suggests that we can build up resources with one pair of modalities, such as language and sketching, which can potentially be used to ground other modalities (e.g., vision, haptics) by using these same representations and reasoning techniques in an amodal core for spatial reasoning.

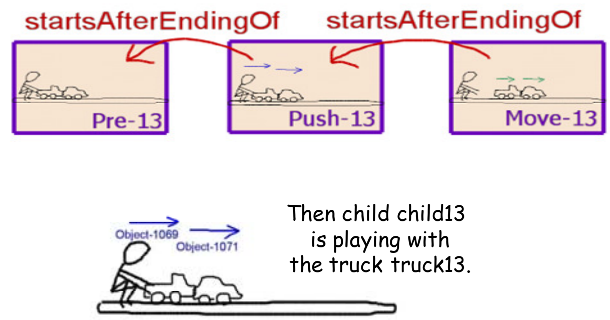


Figure 3: Comic graphs combine sketches and language to describe behaviors over time.

We are pursuing this approach by building up a corpus of sketches tied to spatial language, from which generalizations can be learned by the method described in the previous section. In addition to single sketches, we are using *comic graphs* created with CogSketch that combine multiple subsketches, each annotated with simplified English that our NLU system can process. Figure 3 illustrates a stimulus used in a model that learns concepts like pushing, pulling, and motion (Friedman et al 2009). We believe this approach could be scaled up to create a corpus that could be used to provide a way to ground spatial language more broadly. We are exploring ways to crowd-source the creation of such a corpus, most likely via knowledge capture games, to create a resource for the research community.

References

- Avraamides, M., Loomis, J. Klatzky, R. & Golledge, R. 2004. Functional equivalence of spatial representations derived from vision and language: Evidence from allocentric judgments. *JEP:Learning, Memory, & Cognition* **30**(4), 801-814.
- Forbus, K., Usher, J., Lovett, A., Lockwood, K., & Wetzel, J. (2008). CogSketch: Open-domain sketch understanding for cognitive science research and for education. In *Proceedings of the Fifth Eurographics Workshop on Sketch-Based Interfaces and Modeling*. Annecy, France.
- Friedman, S., Taylor, J. & Forbus, K. (2009). Learning Naïve Physics Models by Analogical Generalization. In *Proceedings of the 2nd International Analogy Conference*. Sofia, Bulgaria.
- Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7, 155-170.
- Lockwood, K., Lovett, A., and Forbus, K. 2008. Automatic classification of containment and support spatial relations in English and Dutch. *Spatial Cognition 2008*. Freiburg, Germany