

Modeling Learning of Relational Abstractions via Structural Alignment

Subu Kandaswamy (subu@u.northwestern.edu)

Kenneth D. Forbus (forbus@northwestern.edu)

Qualitative Reasoning Group, Northwestern University, 2133 Sheridan Road
Evanston, IL 60201 USA

Abstract

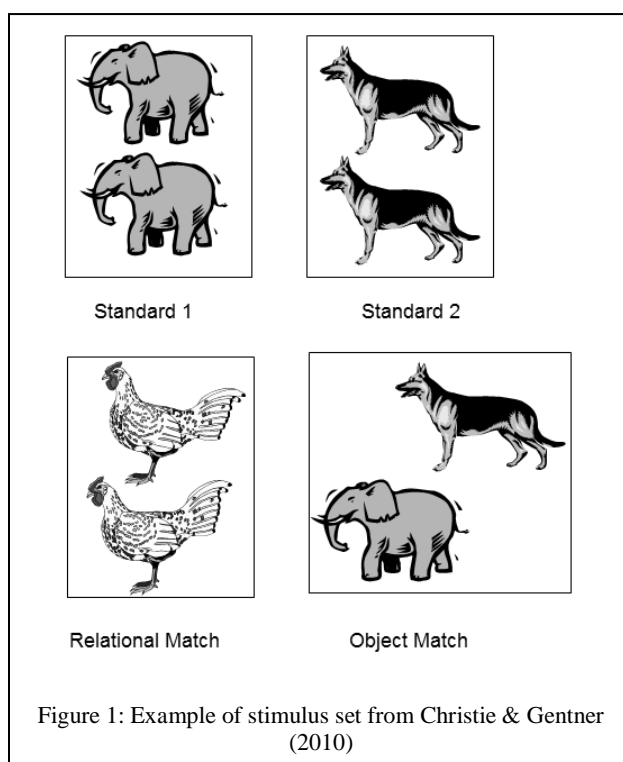
Learning abstract relationships is an essential capability in human intelligence. Christie & Gentner (2010) argued that comparison plays a crucial role in such learning. Structural alignment highlights the shared relational structure between compared examples, thereby making it more salient and accessible for subsequent use. They showed that 3-4 year old children who compared examples in a word-extension task showed higher sensitivity to relational structure. This paper shows how a slight extension to an existing analogical model of word learning (Lockwood et al 2008) can be used to simulate their experiments. This provides another source of evidence for comparison as a mechanism for learning relational abstractions.

Introduction

Our ability to abstract and reason with relations between objects is an essential part of our intelligence. As children, we acquire a variety of relations, including spatial relations such as *above*, and *on*, and functional relations like *edible* and *dangerous*. How children acquire and use such relational abstractions is an important question in cognitive development. Gentner (2003) has argued that comparison promotes learning new relational abstractions. The idea is that structural alignment highlights common structure, which becomes more salient and available for subsequent use.

One line of evidence for this theory comes from an experiment by Christie & Gentner (2010). To show that children (ages 3-4) were learning new relations, they used novel spatial relational categories in a word extension task, as illustrated in Figure 1. Here the relationship might be characterized as “An animal above another identical animal”. In the Solo condition, children were shown a single standard (here, Standard 1) and told it was a novel noun (e.g. “Look, this is a jiggy! Can you say jiggy?”). In the Comparison condition, children were invited to compare two examples (e.g. “Can you see why these are both jiggies?” when presenting Standard 1 and Standard 2 simultaneously). In both conditions, children were then presented with a forced-choice task, where they had to choose which one of the alternatives is a jiggy (e.g. “Which one of these is a jiggy?” when presented with the relational match and object match cards). Children in the Solo condition preferred the object match, while those in the Comparison condition chose relational matches twice as often as object matches. This provides evidence that comparison can lead to learning new relational abstractions. In a second experiment, a third condition, Sequential, was added, where children saw two standards serially, to test

whether or not simple exposure to more examples was sufficient to promote learning. They found significant differences between Sequential and Comparison, and between Solo and Comparison, but the difference between Sequential and Solo were not significant. This provides additional evidence that it is comparison specifically that is promoting learning.



This paper shows that this phenomena falls out of computational models of analogical generalization already proposed for word learning. We start by summarizing the necessary background, including the models of analogical matching and generalization used and how we use sketch understanding to reduce tailorability by producing portions of the input representations automatically. Next we describe an extension to a similarity-based word learning model (Lockwood et al 2008) that enables it to model this task. Then we describe two simulation experiments that demonstrate that this model is capable of exhibiting behavior consistent with that described in Christie & Gentner (2010), including sensitivity analyses to shed light on why it does so. After discussing related work, we close with a discussion of future work.

Background

Our model is based on Gentner’s (1983) structure-mapping theory of analogy and similarity. In structure-mapping, comparison involves a *base* and *target*, both structured, relational representations. Comparison results in one or more *mappings*, which contain a set of *correspondences* that describe how the elements in the structured representations align, a *score* that indicates the overall structural quality of the match, and possibly *candidate inferences* representing knowledge that could be projected from base to target (as well as from target to base). Our computational model of comparison is the *Structure-Mapping Engine*, SME (Falkenhainer et al 1989; Forbus et al 1994). Here SME is used both as a component in our model of analogical generalization (described below) and in making the decision in the forced-choice task. The score is normalized to be between zero and one, by dividing it by the score obtained for the maximum self-mapping of base and target.

Analogical generalization is modeled via SAGE (*Sequential Analogical Generalization Engine*), an extension of SEQL (Kuehne et al 2000) which incorporates probabilities and analogical retrieval. Information about concepts is stored in *generalization contexts* (Friedman & Forbus, 2008). Each generalization context maintains a set of examples of that concept, plus generalizations concerning it. Examples are provided incrementally. For each new example, the most similar prior examples and generalizations are retrieved via a model of analogical reminding (MAC/FAC, Forbus et al 1995). The retrieved items are compared, via SME, with the new example. For each comparison, if the score of the best mapping is over a threshold (the *assimilation threshold*), the compared items are assimilated into a generalization – a new one in the case of two examples, or an updated version of the existing generalization in the case of an example and a generalization. The assimilation process keeps the common structure of the mapping, replacing non-identical entities with abstract place-holders. Associated with each fact in generalizations is a probability, based on the number of times a statement aligning with it appears in an example (Halstead & Forbus, 2005). For example, in a generalization about swans, the fact that swans are birds might have a probability of 1.0, while the probability that their color is white might be 0.999 while the probability that their color is black might be 0.001. Non-overlapping facts are kept, albeit given a low probability (i.e., $1/N$, where N is the number of examples assimilated into that generalization). Facts whose probability drops below the *probability cutoff* are removed from the generalization. SAGE is the central component in our word-learning model, as explained below.

Tailorability is an important problem in cognitive simulation. To reduce tailorability, we use automatically constructed visual and spatial representations. These representations are computed by CogSketch (Forbus et al 2011), an open-domain sketch understanding system. CogSketch uses models of visual and spatial processing to

compute qualitative relationships from digital ink. For example, it automatically computes topological relationships (e.g. inside, touching) and relative positions (e.g. above, right of) for the entities in a sketch. It also includes a model of mental rotation, which uses SME to first do a qualitative shape match which then guides a quantitative match (Lovett et al 2009). This enables it to compute relationships such as *sameShape*, *reflectedOnXAxis*, and so on. Conceptual information can be introduced by adding attribute information to entities in the sketch. For example, the top entities in Standard 1 of Figure 1 might be described as identical elephants, one positioned above the other. The attributes are derived from a large, independently-developed knowledge base¹. The relationships automatically computed by CogSketch, along with the conceptual attributes provided for an entity, provide the inputs for our simulation. Moreover, CogSketch provides a mechanism for dividing a sketch into *subsketches*, which is what we use to combine all of the elements of a stimulus set onto the same sketch, for convenience. Figure 2 provides an example of a sketched stimulus set.

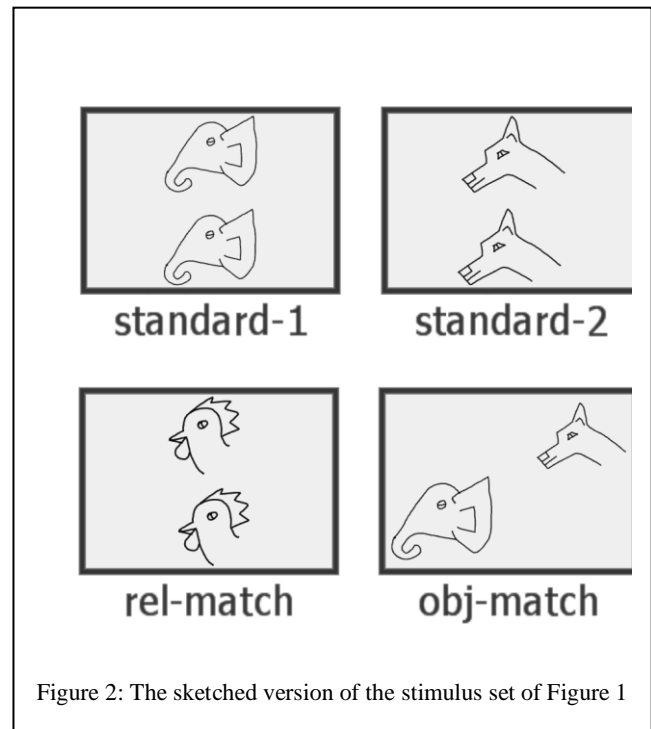


Figure 2: The sketched version of the stimulus set of Figure 1

¹ OpenCyc, see www.opencyc.org

Word Learning via Analogical Generalization

We model the learning of words as follows. For each word, there is a generalization context. Every time the word is used, an appropriate subset of the world is encoded to capture information about what that word denotes, and is provided to the generalization context for that word as an example. The generalizations constructed by SAGE can be considered as the meanings for the words. Note that such meanings can be probabilistic, since SAGE computes frequency information for every statement in the generalization. The ability to track multiple generalizations provides a mechanism for handling multiple senses of a word. The ability to store unassimilated examples provides a means of handling edge cases, and helps provide noise immunity in the face of changes in the underlying distribution of examples of a concept.

This account has been used to successfully model spatial propositions of contact in English and in Dutch (Lockwood et al 2008). It makes no commitment to the particulars of encoding, because this is a complex issue, especially since evidence from studies of novice/expert differences suggest that encoding strategies evolve with learning (Chi et al 1981). When using CogSketch as a source of stimuli, we use an entire subsketch as the relevant material to encode.

Simulation Experiments

Now let us see how this model can be used to simulate the experiments of (Christie & Gentner, 2010). We begin with their Experiment 1.

Simulation Experiment 1

Recall that Experiment 1 used two conditions to show children the new concept, followed by a forced-choice task. We model these as follows:

- **Forced-choice task:** Each of the choices is used with the generalization context for the word to retrieve the most similar generalization or example for that choice. The choice whose similarity score is highest constitutes the decision. For simplicity, We start with an empty generalization context for every novel word used.
- **Solo Condition:** The single example is added to the generalization context for the word.
- **Comparison Condition:** The two examples are added to the generalization context, but since the experimenter has asserted that they are both examples of the concept, we assume that the child is more likely to assimilate them into a generalization, which is modeled by lowering the assimilation threshold from its default of 0.8 to 0.1. We also assume that the probability cutoff is 0.6, so that facts which do not appear in the shared structure will be eliminated from the generalization.

The original experiment used 8 stimulus sets. We encoded 8 sketches of animals, using CogSketch. Each element of the stimulus set (e.g. Standard 1, Standard 2, etc.) was drawn as a separate subsketch. CogSketch’s default encoding methods were used, plus an additional query to ascertain if any of the entities in a subsketch had the same shape as any other, and if so, what transformation held between them (where no transformation implies the relationship `sameShape`). Moreover, filters were used to automatically remove three types of information: Redundant information (e.g. `given (rightOf B A), (leftOf A B)` is redundant), irrelevant information (e.g., global estimates of glyph size like `MediumSizeGlyph`), and bookkeeping information (e.g. relationships describing timestamps). The table below shows the final encoding for the sketches stimulus set (Figure 2) and the resultant generalization.

Table 1: Encoding for the sample sketch.

Standard-1	Standard-2
<pre>(sameShapes Object-99 Object-420) (above Object-99 Object-420) (isa Object-420 Elephant) (isa Object-99 Elephant)</pre>	<pre>(sameShapes Object-104 Object-425) (above Object-104 Object-425) (isa Object-425 Dog) (isa Object-104 Dog)</pre>
Generalization for “jiggy”	
<pre>(above (GenEntFn 1 0 jiggy) (GenEntFn 0 0 jiggy)) (sameShapes (GenEntFn 1 0 jiggy) (GenEntFn 0 0 jiggy))</pre>	

An interesting open parameter concerns the number of conceptual attributes that children might be encoding. While we suspect that a large number of attributes would be encoded², we do not know of data that provides specific estimates. Consequently, we perform a sensitivity analysis by running the simulation while varying the number of conceptual attributes to ascertain their impact on the results. Specifically, we varied the number of attributes from zero to nine. We assumed that encoding is reasonably uniform, i.e. that the same attributes would always be computed for identical objects. For simplicity, we further assumed that the set of attributes computed for one entity had no overlap with the set of attributes computed for another entity whose shape is different. Given these assumptions, we used synthetic attributes (e.g. `Uniquestandard-1MtAttribute8`) for convenience.

Figure 3 shows the results. From the data, we can see that the model chose the relational match 100% of the time for the Comparison condition. This is qualitatively consistent with the behavior of participants in the Comparison condition, where participants chose the relational match around 60% of the time. We believe that the lack of object matches in this simulation condition are due to the use of completely independent attributes for each entity type in the

² See the Specificity Conjecture (Forbus & Gentner 1989).

stimuli sets. Since they are independent, no attributes are left in the generalization after assimilation. The more overlapping attributes there are, the more likely an object match is to become possible.

Returning to Figure 3, in the Solo condition, as the number of attributes rises, the proportion of object matches rises (i.e., the proportion of relational matches falls). Again, this provides a good qualitative fit for the results of (Christie & Gentner 2010) Experiment 1. Since attributes are more salient to children, due to lack of relevant domain knowledge (Ratterman & Gentner, 1998), it is reasonable to assume that they would encode more attributes than relations, which is compatible with the simulation results.

Recall that we assume that the probability cutoff is set high enough that non-overlapping information is immediately filtered out. (Since these are novel concepts, there can be at most two examples in any generalization, and hence the probability of any fact not in the overlap would be 0.5, which is less than the 0.6 threshold.) Would adding in probabilistic information improve the fit of the model to human data? To determine this, we tried changing the probability cutoff to its usual default of 0.2. This leads to all attributes remaining in the generalization, which results in the score for the object match being boosted so high that it always wins over the relational match, regardless of the experimental condition used. This suggests that when children are invited to compare, they do indeed restrict themselves to keeping exactly the overlapping structure.

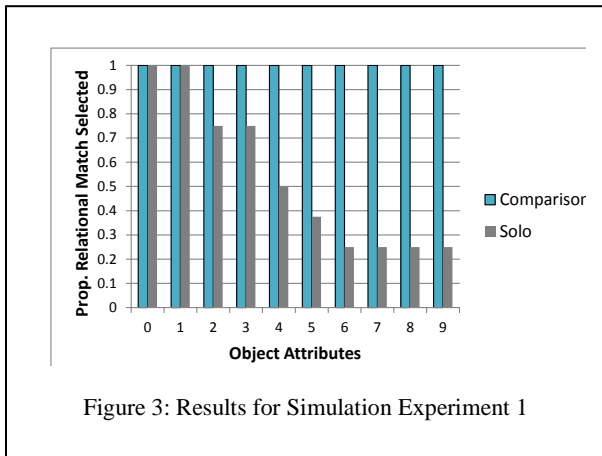


Figure 3: Results for Simulation Experiment 1

Simulation Experiment 2

Experiment 2 in (Christie & Gentner 2010) actually consists of two experiments. Both involved a new condition, the Sequential condition, designed to rule out non-comparison explanations. In Experiment 2a, fillers, in the form of pictures of familiar objects, were interposed between the serial presentation of the standards. No invitation to compare was issued. In Experiment 2b, no fillers were used, and the Solo and Comparisor conditions from Experiment 1 were added, by way of replication. In our

model, fillers would be added to some other generalization context, thus 2a and 2b look identical from the perspective of our model. We model the new condition as follows:

- **Sequential Condition:** The two examples are added to the generalization context, but with the default assimilation threshold 0.8.

Again we varied the number of conceptual attributes, in the same way as in Simulation Experiment 1.

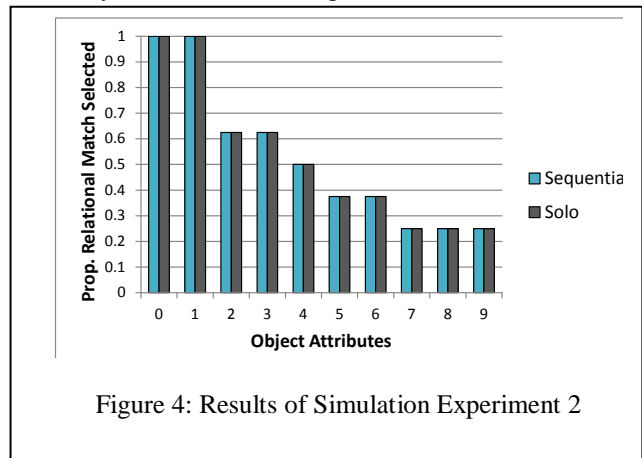


Figure 4: Results of Simulation Experiment 2

Figure 4 illustrates the results. As anticipated, the results for the Sequential condition are identical to the results the model generates for the Solo condition. This is because of the model does not generalize the two standards, and hence the choices will be compared to the exemplars in the generalization context. This makes the results of the

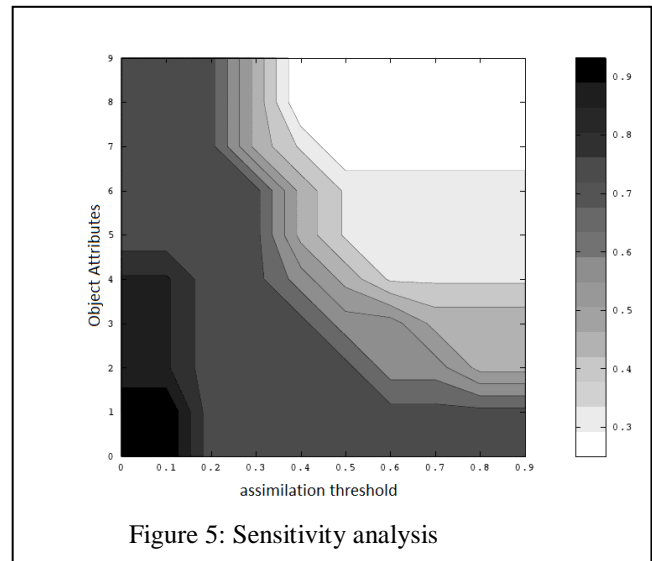


Figure 5: Sensitivity analysis

Comparison condition be the same as the Solo condition.

We know of no psychological evidence that would provide constraints on the value of the assimilation threshold. Consequently, we performed a sensitivity analysis by varying the assimilation threshold between 0.1 and 0.9, while varying the number of attributes from zero to nine. Figure 5 illustrates the results. The region marked as

black indicates a high proportion of relational match choices and then the contour fades down gradually.

The slope of the contour indicates that the model readily generalizes the standards when both the assimilation threshold and the number of object attributes are low. This can be interpreted as follows. A low assimilation threshold corresponds to a higher willingness to accept the standards as belonging to the same category, which fits the assumptions of our model. A low number of object attributes indicates a leaner encoding i.e. not enough attention was paid to the object, or it may be unfamiliar. This is a second possible explanation for why some children chose the relational match for the Sequential condition.

Related Work

There have been several prior computational models of word learning. For example, Siskind (1996) developed an algorithm for cross-situational learning of word/meaning mappings. He used synthetic conceptual representations and lexicons to examine its scaling properties and noise immunity. Our use of arbitrary predicates is similar to his use of synthetic conceptual representations, but our visual representations are grounded in prior cognitive science research. It is an open question whether Siskind's algorithm would work on realistic conceptual representations, and similarly, it is an open question as to whether our word learning algorithm can scale to the size of vocabularies that his does. Another model, described in (Roy and Pentland 2002) uses speech and vision signals as input, to tackle the problem of how children learn to segment these perceptual streams while at the same time learning word meanings. A particularly novel aspect of their approach was modeling a corpus of infant-directed speech they gathered, to ensure their inputs were naturalistic. Our use of sketch understanding is motivated by the hypothesis that it forces us to incorporate high-level vision, while factoring out most of the complexities of signal processing. The relatively crude visual processing techniques used in Roy & Pentland's system, compared to mammalian vision systems, suggests that theirs, too, is an approximation, albeit a more signal-rich version than ours. Neither of these models, nor any other word learning model that we are aware of, has tackled the role of comparison in learning relational abstractions.

Discussion

We have shown that a model of word learning based on analogical generalization, using automatically encoded sketches augmented by conceptual information, can simulate the behavior found in (Christie & Gentner 2010). The invitation to compare, we argue, leads the child to aggressively attempt to form a generalization between the two new exemplars, as modeled by lowering the assimilation threshold and only keeping overlapping structure. This finding is robust across a wide range of choices for the number of object attributes. Moreover, serial presentation to the model, as with humans, does not lead to

relational learning, as measured by responses in the forced-choice task.

There are several lines of future work that suggest themselves. First, we intend to explore if the model can handle closely related phenomena (e.g. Gentner & Namy 1999, who used a similar experimental paradigm but with pre-existing concepts instead of novel concepts). Second, we plan on exploiting more of CogSketch's automatic encoding capabilities, by using it to automatically decompose object-level spatial descriptions into edge-level representations. The sketches in the stimuli will be represented by a set of constituent edges, their attributes and relations that hold between them (e.g. (`isa edge2 StraightEdge`) (`edgesParallel edge2 edge4`)). For example, a square can be segmented into four constituent edges. These more detailed spatial representations will contain more shared attributes and relations and hence would naturally introduce more overlap between entities. This would provide a test of our hypothesis that such overlap is responsible for participants in the Comparison condition sometimes choosing the object match. Finally, we plan to extend this model to explore how object labels promote uniform relational encoding and re-representation (Gentner, 2010).

Acknowledgments

This research was supported by the Spatial Intelligence and Learning Center (SILC), an NSF Science of Learning Center, SBE-1041707, and by a grant from the Socio-Cognitive Architectures Program of the Office of Naval Research.

References

- Chi, M. T. H., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5: 121-152.
- Christie, S. & Gentner, D. (2010). Where hypotheses come from: Learning new relations by structural alignment. *Journal of Cognition and Development*, 11 (3). 356-373.
- Falkenhainer, B., Forbus, K. and Gentner, D. (1989). The Structure Mapping Engine: Algorithm and examples. *Artificial Intelligence*, 41, 1-63.
- Forbus, K., Ferguson, R. and Gentner, D. (1994). Incremental structure-mapping. *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*, August.
- Forbus, K., Gentner, D., and Law, K. (1995). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science*, 19, 141-205.
- Forbus, K. and Gentner, D. (1989). Structural evaluation of analogies: What counts? *Proceedings of the Cognitive Science Society*.
- Forbus, K., Usher, J., Lovett, A., and Wetzel, J. (2011). CogSketch: Sketch understanding for Cognitive Science Research and for Education. *Topics in Cognitive Science*. pp 1-19.

Friedman, S. & Forbus, K. (2008). Learning Causal Models via Progressive Alignment & Qualitative Modeling: A Simulation. *Proceedings of the 30th Annual Conference of the Cognitive Science Society (CogSci)*.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7, 155-170.

Gentner, D. (2003). Why we're so smart. In D. Gentner and S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp.195-235). Cambridge, MA: MIT Press.

Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science* 34,752-775.

Halstead, D. and Forbus, K. (2005). Transforming between Propositions and Features: Bridging the Gap. *Proceedings of AAAI-2005*. Pittsburgh, PA.

Kuehne, S., Forbus, K., Gentner, D. and Quinn, B. (2000). SEQL: Category learning as progressive abstraction using structure mapping. *Proceedings of CogSci 2000*, August.

Lockwood, K., Lovett, A., and Forbus, K. (2008). Automatic Classification of Containment and Support Spatial Relations in English and Dutch. *Proceedings of Spatial Cognition*.

Lovett, A., Tomai, E., Forbus, K. & Usher, J. (2009). Solving geometric analogy problems through two-stage analogical mapping. *Cognitive Science*, 33(7), 1192-1231.

Ratterman, M., & Gentner, D. (1998) More evidence for a relational shift in the development of analogy: Children's performance on a causal-mapping task. *Cognitive Development*, 13, 453-478.

Roy, D. and Pentland, A. (2002). Learning words from sights and sounds: A computational model. *Cognitive Science*, 26,113-146.

Siskind, J. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition* 61,39-91.