UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
Modeling word learning and its impact in reading comprehension

Permalink
https://escholarship.org/uc/item/4mm9h3nn

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 36(36)

ISSN
1069-7977

Author
Vale-Lisboa, Juan

Publication Date
2014

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Developmental Changes in the Semantic Organization of Living Kinds

Layla Unger (LUnger@andrew.cmu.edu)
Carnegie Mellon University, Department of Psychology, 5000 Forbes Avenue
Pittsburgh, PA 15214 USA

Anna V. Fisher (Fisher49@andrew.cmu.edu)
Carnegie Mellon University, Department of Psychology, 5000 Forbes Avenue
Pittsburgh, PA 15214 USA

Christopher J. MacLellan (CMaclell@cs.cmu.edu)
Carnegie Mellon University, Human-Computer Interaction Institute, 5000 Forbes Avenue
Pittsburgh, PA 15214 USA

Abstract

Semantic knowledge contains information about both individual concepts and relationships between concepts. Relationships come in many forms, including taxonomic and thematic, and are critical for converting collections of attributes known about each entity into an interconnected web of semantic knowledge. According to computational modeling studies, increasing knowledge about entities and their relationships should support increasing elaboration in the organizational structure of semantic knowledge. In contrast, extant empirical research has presented a static picture of the developmental trajectory of semantic organization, in which concepts remain organized according to thematic relations into adulthood. The current developmental study introduces methodological innovations designed to overcome limitations that may have skewed the developmental trajectory described in prior studies. The picture that has emerged from this study reflects dramatic changes in semantic organization from preschool age to adulthood, in which an initially limited grasp of multiple types of relations expands and becomes increasingly robust with age.

Keywords: Semantic Knowledge; Cognitive Development

Introduction

Over the course of development, children learn a substantial amount of information about both entities in the world around them, and the relationships between these entities. These relationships play a critical organizational role by converting individual collections of attributes known about each entity into a structured, interconnected network of semantic knowledge. The relationships that link entities fall into many types. For instance, thematic relationships link entities based on their co-occurrence in the same environment, whereas taxonomic relationships capture the degree to which entities are of the same kind based on shared properties. Computational models suggest that semantic development entails an increasingly rich organization of concepts into a network of clusters of related entities (Kemp & Tenenbaum, 2008; McClelland & Rogers, 2003). Early clusters are diffuse and strongly influenced by more superficial relations, such as environmental co-occurrence or perceptual similarity. With access to increasingly rich information, the networks become organized into discrete clusters according to kind-based relations. For example, given a small number of features for a set of animals, Kemp and Tenenbaum's (2008) structure learning algorithm identified crude divisions between water animals, land herbivores, land carnivores, insects, and birds. Given increasingly large numbers of features, the algorithm identified both fine-grained structures within these groups, such as a distinction between water mammals, water reptiles, and fish, and links between different groups (see also Hills, Maouene, Maouene, Sheya, & Smith, 2009; McClelland & Rogers, 2003).

This trajectory is consistent with evidence about the types of relationships that children recognize presented by studies using match-to-sample tasks, in which children are asked to identify which of multiple entities matches a target entity according to some criterion. This research has revealed that when task demands are low, children can make principled choices according to both thematic and taxonomic relations from an early age (Nguyen, 2007). This ability grows dramatically over the course of development: With age, children can succeed on higher-demand versions of the task, flexibly switch between making choices based on different relations, and provide articulate justifications for their choices (e.g., Blaye, Bernard-Peyron, Paour, & Bonthoux, 2006; Fisher, 2011).

The increasing competence in identifying different kinds of relationships among entities is particularly pronounced in children's understanding of taxonomic relationships (Blaye et al., 2006; Hashimoto, McGregor, & Graham, 2007). However, empirical investigations of the organizational structure of the semantic knowledge network have revealed a static developmental picture, in which concepts remain organized according to thematic relations throughout childhood and even into adulthood (Crowe & Prescott, 2003; Storm, 1980; Winkler-Rhoades, Medin, Waxman, Woodring, & Ross, 2010). The paradoxically static developmental trajectory revealed by these studies may in part stem from aspects of their methodology that limited the scope of their data to capture only a narrow facet of semantic organization. The following section describes these limitations, and the methodological approach of the present study designed to overcome them.

Past and Present Methodological Approaches

Unlike match-to-sample tasks, studies that investigate the underlying organizational structure of semantic knowledge collect pairwise similarity judgments for a full set of presented entities. This approach allows researchers to go beyond assessing whether participants can recognize a given type of relation by deriving a picture of the structure of the subset the semantic knowledge network that represents these entities. Examinations of this structure can reveal the influences of different types of relationships, relationships that vary in strength, and overlapping relationships (as when entities are both taxonomically and thematically related).

The majority of these studies have used the *semantic fluency* paradigm to collect similarity judgments (Crowe & Prescott, 2003; Storm, 1980; Winkler-Rhoades et al., 2010). In this paradigm, participants are asked to list as many items as they can think of from a target category, such as "Animals", within a short period of time. The proximity of items to each other across lists is taken as a measure of their perceived similarity and therefore the proximity of their corresponding conceptual representations in semantic knowledge. These data are then submitted to analyses designed to derive representations of the structure underlying semantic organization. For example, hierarchical cluster analysis identifies local clusters of items linked by short distances, that are in turn embedded within higher order clusters of items linked by greater distances. An examination of the relationships that apparently determine whether concepts are closely linked is taken to reveal the relationships that organize semantic knowledge. For example, Crowe and Prescott (2003) determined that the hierarchical clustering solutions for children aged 5 to 10 were all predominantly organized according to thematic relationships. Similarly, Storm (1980) found no significant differences between hierarchical clustering solutions for age groups ranging from Kindergarten to Adults, although the relationships that governed these solutions were not clear.

This static developmental trajectory for semantic organization from early childhood to adulthood stands in stark contrast to the pronounced changes observed in both computational modeling studies and developmental studies that used match-to-sample tasks. However, it is possible that the semantic fluency task may yield skewed data due to two critical limitations. First, over 90% of the animals participants listed across studies were mammals. Data collected from within a taxonomic category have limited potential to reveal distinctions between concepts based on taxonomic relations, thus contributing to the apparent absence of taxonomic relations from semantic organization. Second, participants may invoke a response strategy in which they intentionally search for items associated with the same environment. In fact, the semantic fluency task has been used to investigate the development of children's ability to flexibly invoke response strategies (Snyder & Munakata, 2010). From this perspective, participants may draw on a biased sample of the relations that organize semantic knowledge when completing this task.

To overcome these limitations, this project developed a methodological approach that does not rely on spontaneous production data. Instead, this methodology entails the selection of a predetermined set of items drawn from overlapping taxonomic and thematic groups. For example, the set of stimuli used in this study includes items such as 'seaweed' (thematically related to other water items, and taxonomically related to other plants), and 'chicken' (thematically related to other farm items, and taxonomically related to other birds). Participants were asked to arrange these items on a grid in a task modeled on the Spatial Arrangement paradigm introduced by Goldstone (1994). Participants arranged items multiple times according to three different sorting prompts, in which they were asked to put together things that "go together", "are the same kind of thing", or "match". The use of multiple prompts was designed to allow participants to draw on multiple relations when making arrangement judgments, as past research has demonstrated that prompt wording influences the relationships on which participants focus (Lin & Murphy, 2001; Waxman & Namy, 1997). The proximity of items to each other on sorting trials was taken as an index of the proximity of their corresponding mental representations.

This task was administered to participants from five age groups, ranging from pre-school age children to adults. In order to track the development of semantic organization across this range, data were submitted to analyses designed to assess participants' sensitivity to different types of relationships, and to analyses designed to derive a picture of the structure underlying semantic organization.

Method

Participants

The total sample of 90 participants included 18 participants in each of five age groups: Preschool (M_{age} = 4.5 years, SD = 0.83), Kindergarten (M_{age} = 5.6 years, SD = 0.52), First Grade (M_{age} = 6.5 years, SD = 0.44), Second Grade (M_{age} = 7.6 years, SD = 0.72), and Adults (M_{age} = 19 years, SD = 1.84). Approximately half of the participants in each age group were female.

Materials

The materials included a game board, a stimulus sheet, 15 stimulus cards, and a camera. The game board was a 10 x 10 grid of 2.5" squares. The stimuli were a set of 15 plants and

Table 1: Plants and animals used in sorting task.

	Birds	Mammals	Plants
Farm	Chicken	Pig	Corn
	Turkey		Lettuce
Water	Penguin	Whale	Seaweed
		Beaver	Water Lily
Wild/Zoo	Ostrich	Lion	Cactus
	Eagle	Kangaroo	

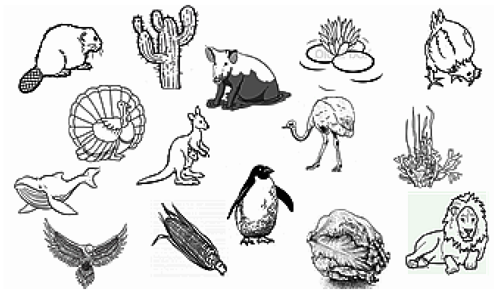

Figure 1: Images used in picture sorting task, presented in the configuration used for the stimulus sheet.

animals that were calibrated with a separate group of adult participants to be cross-classifiable into one of three taxonomic groups (mammal, bird, or plant) and one of three thematic groups (farm, water, and wild/zoo). The assignment of stimuli to thematic groups was based on the results from a questionnaire in which a separate group of adult participants rated the degree to which each possible pair of a broader range of 20 plants and animals belonged to each of the thematic groups on a 7-point Likert scale. Each of the five stimuli assigned to a given thematic group were those that achieved a rating of ≥ 5.5 for that group when paired with each of the other four (see Table 1).

The selected items were presented as black and white line drawings intended to reduce the superficial visual similarity of items to each other, particularly for items from the same group. The stimuli were arranged in a random order on the stimulus sheet (see Figure 1). Additionally, each organism was depicted individually on a separate stimulus card.

Procedure

Participants were tested individually in a quiet space. To begin, participants were told that they were going to play a game in which their job was to help a character, Zibbo, organize his favorite things in “a few different ways”. The experimenter then showed participants the stimulus sheet, named each organism, and removed the sheet from view. The experimenter then administered three sorting trials. Each sorting trial began with a prompt to put together plants and animals that: “Go together”, “Are the same kind of thing”, or “Match”. The order of these prompts was counterbalanced across three between-subjects versions of the task that were otherwise identical. The experimenter then placed one of the 15 stimulus cards on a central square of the game board, and then named and presented each of the remaining cards to the participant for the participant to place on the board. Each sorting trial presented stimuli in a different, pre-determined, pseudo-random order in which no more than two organisms from the same taxonomic or thematic group were presented consecutively. During a given trial, participants were allowed to move cards that they had placed earlier in the trial. After each trial, the experimenter took a photograph of the game board to record the locations of the 15 cards, then removed the cards from the game board. The average duration of each sorting trial was approximately 4 min.

Results

Because no effects of prompt wording were observed across the sample ($F(2,34)=1.33$, p=.28), the following analyses are presented for data collapsed across sorting trials for each participant.

Scoring

The photographs taken following each sorting trial were scored by treating the 10x10 game board as a coordinate plane and calculating the Euclidean distance between each card. The range of possible distances was 1 (cards adjacent) to 12.73 (cards on opposite corners of the board).

Sensitivity Analyses

To determine the extent to which participants in each age group were sensitive to different types of relationships, each of the pairwise distances between the 15 stimuli (121 in total) were assigned to one of four Relationship conditions: 1) Both taxonomic and thematic, 2) Taxonomic, 3) Thematic, and 4) Unrelated. This assignment was determined on the basis of the allocation of stimuli to groups given in Table 1. These data were then submitted to a set of ANOVAs in order to assess the effect of Relationship on the full sample as a whole and on each age group. Bonferroni-corrected pairwise comparisons were used to further explore condition differences.

First, the data were submitted to a 4 x 5 Mixed ANOVA, with one within-subjects factor of Relationship (4 levels), and one between-subjects factor of age group (5 levels). This analysis revealed an overall main effect of Relationship, $F\ (2.27, 188.41) = 120.21$, $p < .001$, no main effect of Age, $F\ (4, 83) = 1.35$, $p = .26$), and a significant interaction between Relationship and Age, $F\ (9.08, 188.41) = 10.71$, $p < .001$). Pairwise comparisons between Relationship conditions revealed that all conditions were significantly different from each other (all ps < .05).

Data for each age group were then submitted to separate Repeated Measures ANOVAs, with Relationship as the within-subjects factor. The effect of Relationship was significant in all age groups, all Fs > 7.52, ps < .01. Pairwise comparisons were used to further explore these differences (see Figure 2). Adults differentiated between pairs of items from all four Relationship conditions, with both taxonomically and thematically related pairs placed closest (mean distance = 2.16), followed by taxonomically related pairs (mean=2.85), then thematically related pairs (mean=3.76), with unrelated pairs placed farthest from each other (mean=4.27), all ps < .01.

In contrast, Preschoolers placed items that were both taxonomically and thematically related (mean distance =2.96) closer together than all other pairs of items (means=3.40, 3.33, 3.42 for thematic, taxonomic, and unrelated pairs, respectively), all ps < .05, but did not differentiate between pairs from any other conditions, all ps

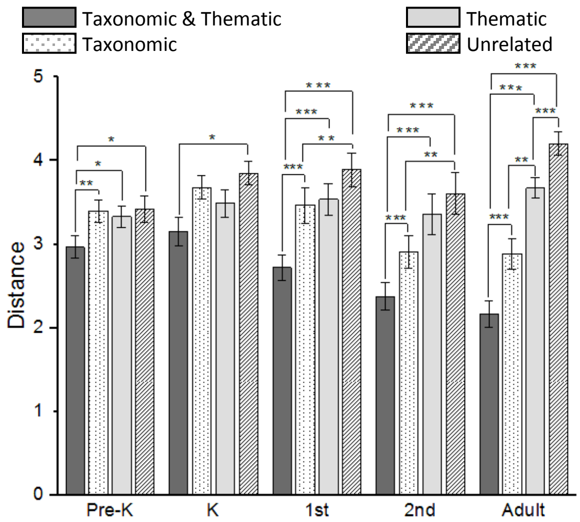

Figure 2: Mean distances between pairs of items in each Relationship condition for each age group.

> .9. Kindergarteners only placed items that were both taxonomically and thematically related (mean=3.15) closer than unrelated items (mean=3.85), $p < .05$. First and Second Grade students showed an intermediate pattern in which they placed items that were both taxonomically and thematically related (means=2.72 and 2.38) closer than all other pairs of items, all *ps* < .001, and placed items that were taxonomically related (means=3.47 and 2.91) closer than unrelated items (means=3.90 and 3.61), all *ps* < .05. For Second Grade students, the differences between taxonomic and thematic pairs (mean distance = 3.36) and between thematic and unrelated pairs were marginally significant, *ps* = .07 and .06, respectively.

Structure Derivation Analyses

For these analyses, the full set of average pairwise distances between the 15 items for each age group were first submitted to hierarchical cluster analysis, with distance between clusters calculated using the Average method. This form of analysis assigns items that are linked by short distances to local clusters that are in turn embedded within higher-order clusters of more distantly linked items. The resultant hierarchical clustering solutions are presented as a dendrogram for each group in Figure 3. These dendrograms are composed of nodes (represented by vertical lines), and arcs linking these nodes (represented by horizontal lines). The 15 items appear as the external nodes of each dendrogram. These external nodes are connected by arcs to form internal nodes that group the 15 items into hierarchical clusters. The height of the nodes captures both the similarity of items within a cluster and the distinctiveness of items in different clusters. Shorter nodes reflect greater similarity.

Qualitative assessment of these dendrograms reveals dramatic changes across age groups. The dendrogram for adult participants is highly organized at all hierarchical levels. At the highest level, the items are split according to a plants-versus-animals distinction, and within the animals

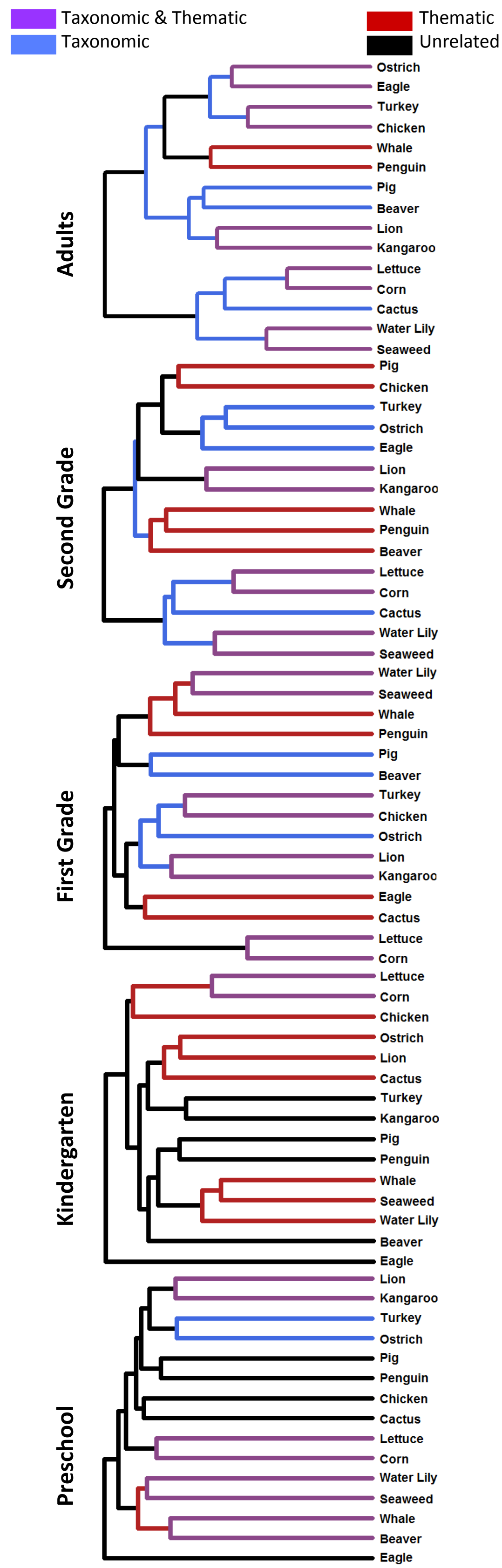

Figure 3: Dendrograms for each age group.

cluster, items are further split according to a mammals-versus-birds distinction (though note that Whale violates this distinction, and is clustered thematically with Penguin). Within these three overarching taxonomic groups, items are grouped into local clusters of both taxonomically and thematically related items. In contrast, dendrograms for the Preschool participants primarily contain local clusters of items that are both taxonomically and thematically related and are in turn embedded within idiosyncratic higher order clusters that do not reflect any readily classifiable relationships. Higher order structure begins to emerge in the dendrogram for Kindergarten participants, in which both local and higher order structure are governed by both thematic and idiosyncratic, non-classifiable relationships. The organization of the dendrogram for First Grade students is more principled: The influence of tight links between taxonomically and thematically related items is evident in local clusters, and the influences of purely taxonomic and purely thematic links are evident in both local and higher order clusters. This trend towards principled organization at multiple levels becomes more apparent in the dendrogram for Second Grade students, which is first split into two higher-order clusters that capture a plants-versus-animals distinction. Within the plants cluster, items are organized into local clusters of both taxonomically and thematically related items. The animals cluster includes a thematic water cluster (Whale, Penguin, and Beaver), a taxonomic bird cluster (Turkey, Ostrich, Eagle), and clusters of both taxonomically and thematically related items.

To quantitatively assess developmental changes apparent in the dendrograms, the hierarchical clustering solutions for each age group were compared to each other following the Fowlkes and Mallows (1983) method. In this method, each solution to be compared is cut at multiple levels to produce a set of flat partitions that contain 2 to n-1 clusters, where n is the number of items. The two solutions are then compared in terms of their similarity at each partition level *k*, where similarity is the degree to which the items are partitioned into the same clusters. The result of this comparison is the B(k) statistic, which ranges from 0 (minimally similar) to 1 (maximally similar). Values for B(k) can be computed both for the comparison itself, and to determine the range of B(k) that would be expected if the similarities between partitions were produced by chance. When the observed value of B(k) exceeds this range for a given partition level, one can conclude that the hierarchical clustering solutions are similar at that level. The outcomes of this analysis are summarized in Table 2.

Table 2: Proportion of levels on which dendrograms are similar.

	Adult	2nd	1st	K	Pre-K
Adult	---				
2nd	7/13	---			
1st	7/13	1/13	---		
K	0/13	0/13	7/13	---	
Pre-K	0/13	1/13	6/13	6/13	---

In contrast with findings reported by Storm (1980) and Crowe and Prescott (2003), findings summarized in Table 2 revealed that none of the age groups' hierarchical clustering solutions were similar to each other at more than half the partition levels. The age groups that were similar to each other at approximately half the partition levels included the youngest participants (Preschool, Kindergarten) and the oldest participants (Second Grade and Adults); participants in the middle of this range (First Grade) reached this criterion for comparisons to both the youngest participants and Adults. Conversely, the youngest and oldest age groups were not similar at all or all but one partition level.

Discussion

Different lines of inquiry into the developmental trajectory of semantic organization have yielded conflicting findings. Computational modeling studies have indicated that, over the course of learning, semantic organization progresses from a crude division of concepts into internally unstructured groups to a rich, interconnected network of concepts that captures both the structure of relations within and links between the original groups (Hills et al., 2009; Kemp & Tenenbaum, 2008; McClelland & Rogers, 2003). Similarly, empirical studies that have investigated the recognition of different types of relationships, such as those that have used match-to-sample tasks, suggest that an early, tenuous grasp of multiple types of relationships grows increasingly robust with age. This trajectory appears particularly pronounced with respect to taxonomic relationships (Blaye et al., 2006; Hashimoto et al., 2007). Conversely, empirical studies of the structure underpinning semantic organization have indicated that this structure remains unchanged over the course of development, remaining governed by the same relationships even into adulthood (Crowe & Prescott, 2003; Storm, 1980; Winkler-Rhoades et al., 2010).

The findings presented here provide evidence that the static developmental trajectory suggested by prior studies that used the semantic fluency paradigm is likely an artifact of their methodological approach. These studies' use of spontaneous production data may have provided a biased sample of the relationships that govern semantic organization by restricting the range of items submitted to structure derivation analyses and invoking participants' use of response strategies based on specific relationships. The present study was designed to overcome these limitations by using a task that incorporates a pre-determined set of items that are cross-classifiable into both taxonomic and thematic groups. The findings that emerged from this approach indicate that adults recognize an increasing degree of differentiation from unrelated to thematically related, taxonomically related, and both thematically and taxonomically related items. The structure of semantic organization for this sample is highly principled, with increasingly fine-grained taxonomic groups that are subdivided into clusters of both taxonomically and thematically related concepts. Conversely, Preschool-age children differentiate between items that are strongly linked on the basis of both taxonomic and thematic relationships versus all other items, but do not differentiate between items that are linked by only a single type of relationship versus unrelated items. Similarly, the structure of semantic organization at this age consists of local clusters of both taxonomically and thematically related concepts that are embedded within higher order clusters that are not governed by any discernable relationships. The intermediate age groups show a transitional pattern between these two endpoints. From the sensitivity analyses, Kindergarten students appear to treat pairs of items that are both taxonomically and thematically related as less distinct than Preschool children, and only differentiate between these and unrelated pairs. The structure derivation analyses revealed that children at this age are beginning to recognize some degree of higher order structure that is primarily influenced by thematic relationships. Children in First and Second Grade begin to recognize links between items that are based on a single relationship. Principled organization at multiple hierarchical levels according to both taxonomic and thematic relationships also emerges during this time. The influence of these relationships at higher order levels is particularly pronounced in Second Grade, when the overall structure becomes divided according to a plants-versus-animals distinction.

It should be noted that the generalizability of this pattern to non-living entities is currently unknown. However, because taxonomic and thematic relations are comparably characteristic of associations across all ontological categories, there is no *a priori* reason to predict that the development of semantic organization will follow a different trajectory in other domains.

Overall, the picture that has emerged from this study therefore captures three developmental patterns. First, an early recognition of strong links between items that are related along multiple dimensions matures into a recognition of links between items that are related along only a single dimension. Second, higher order structure becomes increasingly governed by discernable relationships. Third, taxonomic relationships are gradually prioritized over thematic relationships, in terms of both the degree to which taxonomically related items are differentiated from other pairs of items, and the influence of taxonomic relationships on higher order structure. This picture is in line with both the increasing richness of semantic organization revealed in computational modeling studies and the expanding ability to recognize multiple types of relationships, particularly taxonomic, revealed in empirical match-to-sample paradigm studies.

Acknowledgements

This work was supported by a Graduate Training Grant awarded to CMU by the Department of Education, Institute of Education Sciences (R305B040063). We thank Charles Kemp, David Pfau, Andrew Bauer, and Erik Thiessen for their valuable input and comments. We thank the children, parents, teachers, and participating schools and institutions for making this project possible: Sacred Heart Elementary School, Community Day School, and Phipps Conservatory. Finally, we thank Manon Sohn for her help collecting data.

References

Blaye, A., Bernard-Peyron, V., Paour, J.-L., & Bonthoux, F. (2006). Categorical flexibility in children: Distinguishing response flexibility from conceptual flexibility. *European Journal of Developmental Psychology, 3*, 163-188.

Crowe, S. J., & Prescott, T. J. (2003). Continuity and change in the development of category structure: Insights from the semantic fluency task. *International Journal of Behavioral Development, 27*, 467-479.

Fisher, A. V. (2011). Processing of perceptual information is more robust than processing of conceptual information in preschool-age children: Evidence from costs of switching. *Cognition, 119*, 253-264.

Goldstone, R. (1994). An efficient method for obtaining similarity data. *Behavior Research Methods, Instruments, & Computers, 26*, 381-386.

Hashimoto, N., McGregor, K. K., & Graham, A. (2007). Conceptual organization at 6 and 8 years of age: Evidence from the semantic priming of object decisions. *Journal of Speech, Language and Hearing Research, 50*, 161-176.

Hills, T. T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009). Categorical structure among shared features in networks of early-learned nouns. *Cognition, 112*, 381-396.

Kemp, C., & Tenenbaum, J. B. (2008). The discovery of structural form. *Proceedings of the National Academy of Sciences, 105*, 10687-10692.

Lin, E. L., & Murphy, G. L. (2001). Thematic relations in adults' concepts. *Journal of Experimental Psychology: General, 130*, 3-28.

McClelland, J. L., & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience, 4*, 310-322.

Nguyen, S. P. (2007). Cross-classification and category representation in children's concepts. *Developmental Psychology, 43*, 719-731.

Snyder, H. R., & Munakata, Y. (2010). Becoming self-directed: Abstract representations support endogenous flexibility in children. *Cognition, 116*, 155-167.

Storm, C. (1980). The semantic structure of animal terms: A developmental study. *International Journal of Behavioral Development, 3*, 381-407.

Waxman, S. R., & Namy, L. L. (1997). Challenging the notion of a thematic preference in young children. *Developmental Psychology, 33*, 555-567.

Winkler-Rhoades, N., Medin, D., Waxman, S. R., Woodring, J., & Ross, N. O. (2010). Naming the animals that come to mind: effects of culture and experience on category fluency. *Journal of Cognition and Culture, 10*, 205-220.