Toward a computational model of creativity: Novel hypothesis generation from structural knowledge

Shohei Hidaka
School of Knowledge Science
Japan Advanced Institute of Science and Technology
1-1 Asahidai, Nomi, Ishikawa, Japan
shhidaka@jaist.ac.jp

Abstract— **Creativity, generation of a new idea from past experience and knowledge, is one of fundamental aspects of inferential process making progress in many scientific and non-scientific fields. Children's learning at their early development needs to be creative: by nature, they frequently encounter new situations in which they need to infer about things unfamiliar to them. In the present study, we attempt to review empirical and theoretical studies on creative inference in children's word learning. Two theoretical implications for creative cognition are discussed. A computational model of word learning offers a formal way to analyze the relationship between hypothesis generation and structural prior knowledge, which can potentially explain some aspects of empirical findings on new idea generation.**

Keywords-component; creativity; abduction; statistical learning; word learning

I. Creativity in Early Word Learning

Creativity is one of the fundamental aspects of the cognitive process, making progress in many fields such as product design, architecture, computer interaction, education, writing, sports, and science [1]. Thus, creativity has been studied in a broad range of fields and approaches: psychological experiments [2], psychometric approaches [3], [4], case studies [5], practical approaches [6], artificial intelligence [7] and so on. Decades of cumulative findings on many basic cognitive processes such as mental image, concept formation, categorization, memory retrieval, analogical reasoning, problem solving have implications to understanding on creativity. Finke, Ward, and Smith [1] have integrated these implications on basic cognitive processes and proposed a conceptual model of creative cognition. In their model called the "Geneplore" model, they highlighted repetitive interaction between generative and exploratory processes. In generative processes, pre-inventive structures, building blocks potentially used for general purposes, are prepared, and exploration processes combine, weigh and select generated structures. Based on the series of studies, Finke *et al.* [1] suggested constraints on generation and/or exploration play an important role in creativity of outcomes of the interactive process.

In the present study, we explore a possible computational model of creative inference by walking through empirical and theoretical studies on children's word learning. Word learning is not just memorization of a relationship between a word and its referents, but it is rather creative -- generating hypothesis and inductive validation on real-life categories. A word (e.g., "apple") typically refers a certain range of objects (e.g., a red apple, a sliced apple, minced apple, a pictorial apple). A learner needs to infer to what the potential range of referents a word may refer or generalize. This is essentially a process of hypothesis making - inductive and abductive inference from past experience. More importantly, children as well as adults can generalize a word to unexperienced thing - generating a novel hypothesis on what a word means Since children at early developmental stage are less experienced with their mother tongue, their word learning is typically creative- they need to make hypothesis on unexperienced things. This inferential process is one of the core components of creative cognition, which can be characterized as a type of abduction. Since abduction is a key concept through this entire paper, we briefly introduce its basic idea.

A. Deduction, induction and abduction

In his seminal works, Pierce (1931-1958) has established the concept of *abduction* distinguishing from the other two classes of logical inferences, deduction and induction. Deduction is logical inference deriving a result from a given assumption: All the beans in this bag are white (rule), and this is a bean from the back (case), then this bean must be white (result). Thus deductive inference creates nothing newer than the given assumption, so inference process is guaranteed errorless. On the other hand, induction is local inference generalizing a limited number of observations to a more general principle: This is a bean drawn from this bag (case), this bean is white (result), therefore, the rest of the beans in this bag should be white (rule). Induction derives a non-obvious statement from the limited observation, thus it may create new information which may or may not be true. Abductive inference derives a non-trivial statement as well as inductive inference, but unlike inductive reasoning it may derive hypothesis which may not be directly implied by observation: This bean is white (result), and there is a bag with white beans (rule), therefore, this white bean came from this bag (case). In this case, statement "this white bean came from this bag" is a hypothesis (but not only one) which makes sense of rule and result statements above if it is true. However, abductive inference, or hypothesis making, has no guarantee of trueness of its process. Abductive inference is most "illogical" process which derives a hypothesis that has never been observed (in the above case, drawing a bean from the bag is not observed). Therefore, abductive inference is the most creative process in the three classes of logical inference.

In what follows, we review developmental literature on empirical findings of children's novel word generalization. Then we walk through a computational model proposed as a mechanistic account for children's novel word generalization. Finally, we discuss the potential implication of word learning and its model to creative inference in a more general case.

B. Children's Inference On Novel Word

There are an infinite number of objectively correct descriptions of the features characteristic of anything. Quine [9] has pointed out uncertainty of meaning of a novel word in a unknown language. This is the very situation which children may face in everyday life. However, in contrast to this philosophical argument about the difficulty in inference of word meaning, children's learning is efficient. In the first years of life, children begin to comprehend and produce words. Between 8 and 16 months of age, children's receptive vocabulary nearly doubles in size every two months [10]. From 12 to 24 months, their expressive vocabulary follows a similar path of productive growth. It has been estimated that between 18-months and 18-years of age children acquire approximately ten new words per day, or one new word every hour and a half the child is awake [11]. The words a child learns in this time period include nouns, verbs, determiners, prepositions; however, nouns are acquired at a faster rate than other word classes. Accumulated empirical evidence of decades of developmental studies suggests that efficient word learning is likely to be driven by children's powerful inference on novel word meaning. These results derive from laboratory studies of how 2- and 3- year olds generalize a category to new instances given a single exemplar. In these experiments, children are given a novel never-seen-before thing, told its name ("This is a dax") and asked what other things have that name. The results show that children extend the names for things with features typical of animates (e.g., eyes) by multiple similarities, for things with features typical of artifacts by shape (e.g., solid and angular shapes), and for things with features typical of substances by material (e.g., nonsolid, rounded flat shape). The children systematically extend the name to new instances by different features for different kinds of objects [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26].

C. Theoretical explanation of novel word generalization

What processes underlie children's novel word generalization and their efficient learning? Recent theoretical studies have proposed statistical learning models in which children's word learning is viewed as a type of machine learning, such as neural network model [27], [28] and Bayesian statistical inferences [29]. Figure 1 depicts the statistical learning framework of early word learning. A word (or category) is formulated as a probabilistic distribution on feature space (hypothetical two dimensional space shown in Figure 1a). Word learning from instances of a category is viewed as a general statistical inference, estimation of parameters of probabilistic distribution (Figure 1a). Therefore, with more instances of a category, one can estimate probabilistic distribution (its latent equal-likelihood contour is shown as a broken ellipsis) more accurately. In generalization of a category (say, an instance depicted as an unfilled circle is given), estimated probabilistic distribution is used to infer what category a novel instance is likely to belong to. However, this general statistical inference alone does not work for a special case when the first and only instance is given (Figure 1b). Because no other instance is given, it is impossible to generalize it or estimate distribution (or variance) patterns of the category (Figure 1c). This is the computational formulation of novel word generalization which three-year-old children can do.

Statistical inference, or inductive inference, from instances is not sufficient to account for children's novel word generalization. Therefore, a theory on novel word generalization is required to explain how a word learner can infer probabilistic distribution of a novel category based on a single instance more accurately than chance. The past theoretical works have offered a few similar but different explanations about it. The basic idea common in the past models is that children learn not just within-category patterns but also across-category patterns (Figure 1d). Colunga and Smith [27] have proposed an artificial neural network model which learns statistical regularity among an artificial set of word categories. Similarly, Xu and Tenenbaum [29] have proposed a Bayesian statistical learning model which learns a hierarchical structure of word categories as well as each category. The former neural network model implicitly learns higher order statistical regularity across categories as well as within-category patterns, whereas the latter Bayesian model learns across-category patterns over explicitly given structure (i.e., hierarchy). These models, the neural network model or Bayesian model, can certainly learn across-category statistical patterns. However, the past studies to date, including the above two, are limited, since they have worked on "toy data" which is artificially generated and well-controlled. In contrast to the learning of the toy dataset, a natural set of word categories have rich and intriguing structure which is indeed essential - because its essence is not just a learning mechanism but interaction between learning and structure of what is to be learned. Therefore, Hidaka and Smith [30], [31] have proposed a statistical learning model for novel word generalization with abductive inferential mechanism called category packing which works on naturalistic word category and feature set collected in a psychological experiment.

II. HYPOTHESIS GENERATION BASED ON KNOWN WORDS

The category packing is a model of a kind of optimization of category organization and its generalization. According to the category packing model, there is some level of word categories in which category organization simultaneously optimizes two principles, discriminability and inclusion. Categories need to be discriminable: there are few instances belonging to multiple categories. Categories need to be inclusive: each category should be maximally used to refer to as many instances as possible. Simultaneous optimization of two principles is traded-off: more discriminable categories tend to be less inclusive or less discriminable categories tend to be more inclusive. Hidaka and Smith [30], [31] have shown that simultaneous optimization of discriminability and inclusion leads a certain type of category organization: similar categories tend to have similar generalization patterns, or, vice versa; dissimilar categories tend to have dissimilar generalization patterns.

This optimization process is described with analogy to packing physical objects in a limited space in which multiple objects cannot occupy the same space (i.e., exclusivity or discriminability of categories) while we want to pack them in relatively small space (i.e., relatively large inclusion of categories in feature space). Remember our heuristics in packing stuff in a small briefcase: sort out and organize things so that they tend to have similar shape and that is the best heuristics to save room in your briefcase. This is not just a rule of thumb, but Hidaka and Smith [31] have shown that it is indeed the case in probabilistic packing of categories in feature space. Figure 2 shows probabilistic category representation in hypothetical two dimensional feature space in which instances are to be categorized spread over this feature space. In Figure 2a, probabilistic distribution shown as an elliptic contour represents how likely the category has its instances in the feature space. In Figure 2b, a well-packed set of categories are shown: generalization patterns (elliptic contour of likelihood) of nearby categories tend to be similar. On the other hand, in Figure 2c, a non-optimal set of categories are shown: multiple categories may include similar instances (thus less discriminable and have overlapped elliptic contours between them) and have more gaps between categories in which those instances are not classified in any categories.

Now let us consider the advantage of well-packed category organization in the case of novel word generalization. Suppose that a learner who encounters the first instance (shown as a star in Figure 2b and 2c) of an unknown category (shown as broken elliptic contour) needs to infer what the new category is like. This is the schematic formulation of likely situation children may face to in their learning of novel word. Since the learner is given only one instance, it is impossible to inductively (or statistically) infer the whole category structure by its own. However, with known categories organized in a certain way, it is possible to infer a likely generalization pattern of the novel category even from its first and only instance. Specifically, with well-packed known categories (Figure 2b), due to correlation between similarity (location in feature space) and shape of category likelihood (elliptic contour), the learner may correctly infer the novel category from its similar and known categories. On the other hand, with non-packed known categories, due to its irregularity of category organization, it is difficult to correctly infer how the novel category would be generalized. In fact, Hidaka and Smith [31] have collected feature data on a set of natural nouns, and have shown that it is possible to infer a likely category generalization from an only instance of novel noun category.

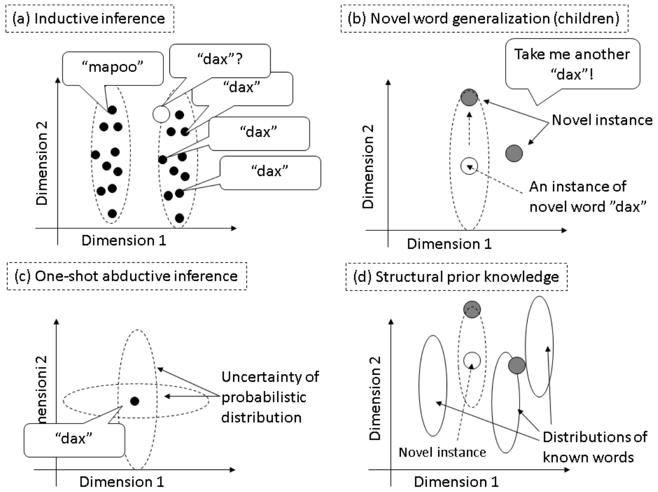

Figure 1: Theoretical framework of statistical learning in early development

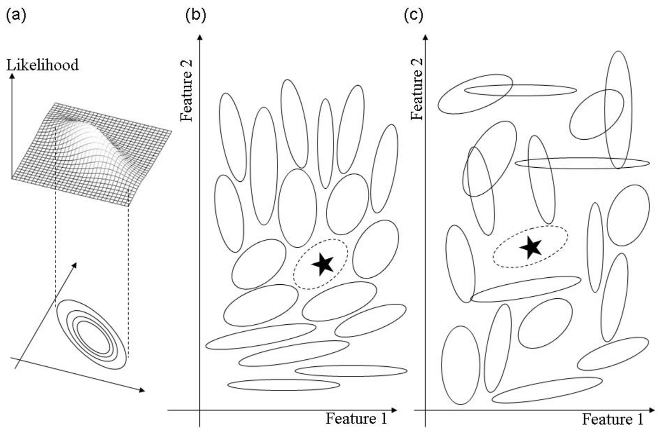

Figure 2: (a) Each ellipsis indicates the equal-likelihood contour of a category, and two schematic illustrations of (b) well-organized and (c) non-well-organized categories. The broken and solid ellipsis in each figure respectively indicate the equal-likelihood contour of the unknown and known category. The star indicates a given first instance of the unknown category.

A. As a model of hypothesis making in categorization

Although Hidaka and Smith [31] have proposed the packing model originally as a computational mechanism of children's novel word generalization, their model may have important implications for the creative inference process. First, we reformulate the model in terms of abductive inference. The category packing model consists of two major parts. One part of the model is a general statistical inference in which the model estimates likely parameters of probabilistic distribution from a given set of instances. This is inductive inference – derive the probabilistic rule on how a category can be generalized from multiple instances of the category. The other part of the model, which is not a typical component of statistical model, is prior belief (derived from packing optimization is abductive inference - it derives a likely probabilistic rule even with the only instance which has substantially no information on generalization. The first inductive part is more accurate with more samples in a category, meanwhile the second abductive part is more accurate with the well-structured other known categories regardless of the number of samples in each category, since the category packing model optimizes category organization. Therefore, in this framework, the power of abductive inference comes from meta-generalization or generalization of generalization patterns of multiple categories. In developmental literature, this is called the second order generalization [27], [24] or over-hypothesis [29], [32], [33].

III. IMPLICATIONS

Here we summarize implications of empirical and theoretical studies of children's word learning for building a computational model of creativity. First, word learning at early development is efficient and creative: intense vocabulary growth [10], [11] and generalization of the first and only instance of novel words [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26]. Despite their limited knowledge of their mother tongue, their vocabulary growth is incredible, and this efficient growth is likely to be supported by generalization of their past experience [24]. Second, theoretical studies have suggested that such efficient learning would be realized not only by ordinary statistical inference (inductive inference from samples) but also by higher order generalization (abductive inference) of known category structure [30], [31], [27], [29]. Third, according to Hidaka and Smith [30], [31], the higher order generalization may be understood as a gap-filling operation which integrates and utilizes distributed and structural information drawn from known words. That is, coherent category organization in which similar categories tend to have similar generation enable to form a likely hypothesis on what a novel word refers to.

A. *Toward a computational model of creativity*

Given the empirical and theoretical implications of early word learning altogether, we draw two insights toward a computational mechanism of creative inference. First, there is a theoretical tradeoff between predictability and flexibility of knowledge representation which holds in any knowledge-based inference including word learning. Second, our review also suggests what should be treated as a creative process which should be determined based on background knowledge - not just a new idea itself but also the other prior knowledge.

1) *Trade-off in knowledge representation*

Since theoretical studies on word learning have suggested that hypothesis making (on novel words) requires well-organized past knowledge, it naturally calls detailed characterization of a domain of interest. This is the other way to imply the "no free lunch" theorem: It is impossible to gain from no assumption on a problem [34]. Therefore, we need to start with construct representation of the problem - multi-dimensional Euclidian feature space was analyzed in order to model learning of early word categories. Depending on the domain of interest, it may not be immediately helpful to characterize the problem as multi-dimensional feature space. Nonetheless, we draw some general prescription for representation of the problem. That is the trade-off relationship between flexibility and predictability of framework. In order to make this point clear, we depicted three schematic representations of a problem (Figure 3): Euclidian space (or regular lattice), Riemannian manifold (or semi-regular lattice with local regularity), and a general graph with arbitrary connections. In general flexible representation such as a graph in which any two nodes may have an edge between them tends to have less predictability. Since any two nodes may or may not have edges, knowing the existence of an edge node A and B does not tell any information on the relationship between A and C. In contrast, Euclidian space is the most constrained (the least flexible) and thus more predictable. Any three points in Euclidian space hold triangle inequality - sum of the distance between A and B and that between B and C is larger than the distance between A and C. This is a type of constraint, and it also works to give predictability – knowing the distances AB and BC tells partial information on the distance between A and C even before knowing them. This transitivity of distances in Euclidian space may be powerful if one knows multiple pieces of partial information - integrating distributed partial information gives a rather accurate guess on an unknown item. This is, in principle, the idea of how the category packing model forms a likely hypothesis on a novel word. Well-organized categories reflect optimization of both predictability and flexibility in the representation of word knowledge. Therefore flexible-yet-predictable will be the middle ground in-between the dichotomy constrained or non-predictable. From this point of view, creative inference may not merely be "lucky guess" but an insight with implicitly or explicitly well-prepared background knowledge (or serendipity). This may be consistent to empirical findings on importance of divergent and convergent thinking (For example, [4]) or constraints generative and exploratory process [2], [1].

2) *What does it mean to be "creative"*

Second, the category packing model implies a specific relationship among creativity, prior knowledge and its constraints. This also suggests a potential computational formulation for what to be creative. According to the category packing model, a novel word with well-organized known categories is more predictable. This is a formulation how a new idea to be evaluated in terms of its background knowledge. We explain this in context of a specific case study on creativity of visual image composition.

Finke [2] has shown that, in his experiments of visual object composition, subjects' generated objects by composition of given parts of objects were more creative if subjects were more constrained to generate them. In this study, creativity for each composed visual image was defined as high and consistent rating by multiple independent evaluators. In his experiments, he have shown that subjects tended to generate creative ideas (1) if subjects work on their own primitive ideas as building blocks for a combined new idea rather than they work on others' primitive ideas given, and (2) if they are given a specific subject to generate a new idea rather than being free to choose a subject. These first and second points are seemingly contradictory to each other, since both are some kind of constraints but work for creativity in the other ways.

However, this is naturally understood in terms of the category packing model.

According to the category packing model, structural patterns in known knowledge is a primary source of abductive inference. Therefore, the point (1) above is natural that combining one's own primitive ideas is more beneficial for creating a new idea. In the other words, different people may have different organizations of knowledge structure – person A's relationship between concept A1 and A2 is not like person B's relationship between concept A1 and A2. Interestingly, the point (2) above is also naturally understood according to the category packing model. Again, the primary source of abductive inference in the packing model is the gap-filling mechanism among prior knowledge structure. This means that a narrower gap – or fully-constrained space – is actually more informative and predictable, and such narrow-gap-filling new idea makes more sense than any-way-possible-gap-filling idea in open space. It is not surprising that "make-sense" yet new idea would be evaluated as a creative one. This is a potential account for creativity which is not just new (and random) but also convincing.

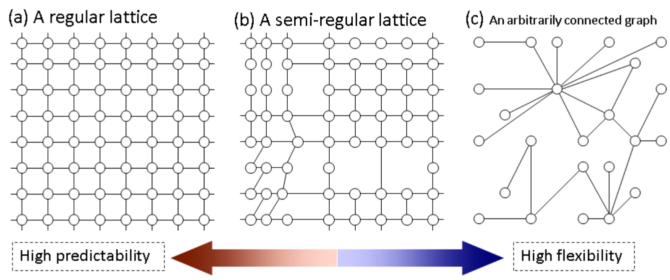

Figure 3: three representations of problem spaces with different predictability and flexibility

B. *Conclusion*

In the present study, we viewed children's creative inference on novel words as abductive hypothesis generation, and drew some implications to studies on creative cognition. It is by no mean sufficient. A potential mechanistic account for creative inference needs to address the other aspects which we did not explicitly explain in the present study. However, we believe this is a step toward establishing computational model of creativity.

ACKNOWLEDGEMENT

This study was supported by Artificial Intelligence Research Promotion Foundation and Grant-in-Aid for Scientific Research B No. 23300099.

REFERENCES

[1] R. A. Finke, T. B. Ward, and S. M. Smith, Creative Cognition: Theory, Research, and Applications. Cambridge: MIT Press, 1992.

[2] R. A. Finke, Creative Imagery: Discoveries and Inventions in Visualization. Hillsdale, NJ: Erlbaum, 1990.

[3] N. W. Kettner, J. P. Guilford, and P. R. Christensen, "A factor-analytic study across the domains of reasoning, creativity, and evaluation," Psychological Monographs: General and Applied, vol. 73, no. 9, pp. 1–31, 1959.

[4] J. P. Guilford, "Measurement and creativity," Theory Into Practice, vol. 5, no. 4, pp. 185–189, 1966.

[5] H. Gardner, Art, Mind, and Brain: A Cognitive Approach to Creativity. New York, NY: Basic Books, 1984.

[6] A. F. Osborn, Applied imagination: Principles and procedures of creative problem solving. Newyork, NY: Charles Scribner's Sons., 1963.

[7] M. A. Boden, "Creativity and artificial intelligence," Artificial Intelligence, vol. 103, no. 1-2, pp. 347–356, 1998.

[8] C. S. Peirce, Collected Papers of Charles Sanders Peirce, C. Hartshorne, P. Weiss, and A. W. Burks, Eds. Cambridge, MA: Belknap Press of Harvard University Press, 1958.

[9] W. V. O. Quine, Word and Object. Cambridge, MA: MIT Press, 1960.

[10] E. Bates, P. Dale, and D. Thal, Individual differences and their implications for theories of language development. Oxford: Basil Blackwell, 1995, pp. 96–151.

[11] P. Bloom, How Children Learn the Meaning of Words. Cambridge, MA: MIT Press, 2000.

[12] E. Colunga and L. B. Smith, "Flexibility and variability: Essential to human cognition and the study of human cognition," New Ideas in Psychology, vol. 26, no. 2, pp. 174–192, 2008.

[13] C. B. Mervis, Concepts and conceptual development: Ecological and intellectual factors in categorization. MA: Cambridge University Press, 1987, ch. Child-basic object categories and early lexical development, pp. 201–233.

[14] N. N. Soja, S. Carey, and E. S. Spelke, "Ontological categories guide young children's inductions of word meanings: object terms and substance terms," Cognition, vol. 38, pp. 179–211, 1991.

[15] S. S. Jones, L. B. Smith, and B. Landau, "Object properties and knowledge in early lexical learning," Child development, vol. 62, pp. 499–516, 1991.

[16] A. E. Booth and S. Waxman, "Word learning is 'smart': Evidence that conceptual information affects preschoolers' extension of novel words," Cognition, vol. 84, pp. B11–B22, 2002.

[17] E. M. Markman and J. E. Hutchinson, "Children's sensitivity to constraints on word meaning: Taxonomic versus thematic relations," Cognitive Psychology, vol. 16, pp. 1–27, 1984.

[18] L. Gershkoff-Stowe and L. B. Smith, "Shape and the first hundred nouns," Child Development, vol. 75, no. 4, pp. 1098–1114, 2004.

[19] S. S. Jones and L. Smith, "How children know the relevant properties for generalizing object names," Developmental Science, vol. 5, pp. 219–232,

2002.

[20] L. B. Smith and D. Heise, Percepts, Concepts and Categories. Amsterdam: Elsevier Science Publishers B. V., 1992, ch. Perceptual Similarity and Conceptual Structure.

[21] S. S. Jones and L. B. Smith, "The place of perception in children's concepts," Cognitive Development, vol. 8, pp. 113–139, 1993.

[22] L. K. Samuelson and L. B. Smith, "Early noun vocabularies: do ontology, category structure and syntax correspond?" Cognition, vol. 73, pp. 1–33, 1999.

[23] E. M. Markman, Categorization and naming in children: Problems of induction. Cambridge, MA: MIT Press, 1989.

[24] L. B. Smith, S. S. Jones, B. Landau, L. Gershkoff-Stowe, and L. Samuelson, "Object name learning provides on-the-job training for attention," Psychological Science, vol. 13, pp. 13–19, 2002.

[25] L. B. Smith, Self-organizing processes in learning to learn words: Development is not induction. Mahwah, New Jersey: Lawrence Erlbaum Associates, 1995, pp. 1–32.

[26] S. S. Jones, "Late talkers show no shape bias in object naming," Developmental Science, vol. 6, pp. 477–483, 2003.

[27] E. Colunga and L. Smith, "From the lexicon to expectations about kinds: A role for associative learning," Psychological Review, vol. 112, pp. 347–382, 2005.

[28] L. K. Samuelson, "Statistical regularities in vocabulary guide language acquisition in connectionist models and 15-20 month olds." Developmental Psychology, vol. 38, pp. 1016–1037, 2002.

[29] F. Xu and J. Tenenbaum, "Word learning as bayesian inference," Psychological Review, vol. 114, pp. 245–272, 2007.[30] S.

[30] Hidaka and L. B. Smith, "Acquisition of a single word to a population of words," Language Learning and Development, vol. 6, no. 3, pp. 206-222, 2010.

[31] ——, "Packing: A geometric analysis of feature selection and fast mapping in children 's category formation," Cognitive System Research, vol. 12, no. 1, pp. 1–18, 2011.

[32] C. Kemp and J. B. Tenenbaum, "The discovery of structural form," Proceedings of the National Academy of Sciences, vol. 105, no. 31, pp. 10 687–10 692, 2008.

[33] N. Goodman, Fact, Fiction, and Forecast. Cambridge: Harvard University Press, 1955.

[34] D. Wolpert and W. Macready, "No free lunch theorems for optimization," IEEE Transactions on Evolutionary Computation, vol. 1, pp. 67–82, 1997.