Semantic Packing: An Account for Category Coherence

Shohei Hidaka (hidaka@cog.ist.i.kyoto-u.ac.jp)
Graduate School of Informatics, Kyoto University;
Yoshida-Honmachi, Sakyo-ku, Kyoto, 606-8501, Japan

Jun Saiki (saiki@cv.jinkan.kyoto-u.ac.jp)
Graduate School of Human and Environmental Studies, Kyoto University;
Yoshida Nihonmatsu-cho, Sakyo-ku, Kyoto 606-8501, Japan

Linda B. Smith (smith4@indiana.edu)
Department of Psychological and Brain Sciences, Indiana University;
1101 East Tenth Street, Bloomington, IN 47405-7007, USA

Abstract

It is argued by a computational model that intuitive groupings for natural categories exist, called category coherence. The relationship between property and generalization of category, referred to as smoothness, is proposed as a quantitative measure of category coherence. A survey on early acquired noun categories suggests that a basic-level category has smoothness. Similarity-based categorization models can not account for category coherence, because the definition of category and feature selection involves an essential circularity. We propose here a semantic packing theory that produces smoothness in categories. In the model, the learner's two conflicting constraints on discrimination and generalization are optimized simultaneously, as if packing knowledge into memory. The equivalence between packing and smoothness is mathematically proven under a specific simplification. Furthermore, the simulation showed that the packing constraint could reproduce the original category organization based on partial survey data. These results indicate that semantic packing could be a core mechanism of fast mapping, a process by which children generalize a novel instance without the need for trial and error.

Introduction

Why Does "Category Coherence" Emerge?

Murphy and Medin (1985) define "category coherence" as the intuitive and useful groupings that characterize natural categories, and claim that this coherence is one of most important aspects of semantic cognition. However, how we learn about feature selection in category coherence remains to be clarified, as the relevant features and clusters of features for a category depend on the kind of category it is. For example, color is more important for discrimination when the item is a pea rather than when it is a ball. How do we learn this? And how do we use the knowledge that we have learned? How, when we see an object-a potential pea or ball-do we know to attend to color or not?

This problem of feature selection has played a key role in theoretical discussions of the mechanisms that underlie category learning. Category-dependent feature selection is difficult to explain in terms of simple similarity-based accounts such as feature lists (Murphy & Medin, 1985; Gelman & Williams, 1997). Murphy and Medin (1985) claim that categorization based only on similarity and its correlation is not enough to solve the coherence problem, and they proposed the importance of folk theory based on causal induction. Their major criticism of similarity-based accounts is their circularity: categorization depends on feature selection, and feature selection depends on the category. For example, "zebra" and "barber's symbol" are similar when the property "has stripe" is heavily weighted; thus some set of property weighting is necessary to identify "zebra" to "horse". As similarity *per se* can not determine categorization, one must already know how to weigh the properties to know what category an entity belongs to. Most category learning studies ignore this circularity and implicitly assume the categories and the relevant properties. Resolving this circularity issue of category property is essential in semantic cognition.

Theory Outline

The main goal of this study is to propose a computational theory of internal semantic representation that accounts for category coherence. The big idea of the model is that a constraint on the learner should be considered instead of category, similarity, or property. [1] .

This is because learners have practical reasons for why a particular category structure is preferable, for example, so that two categories can be discriminated with minimal memory used to represent category and property information. "Efficient" categorization is of practical importance to learners, who after all must deal with everyday events in the world. Thus, as a solution of the circular issue, we will argue what kind of learner efficiency should be considered. More specifically, the essential trade-off between generalization and discriminability is argued to be learning efficiency. We will refer to the constraint on efficiency as "semantic packing", because it is analogous to packing various sized and shaped containers (categories) into a larger but finite container (memory and the attentional and retrieval processes that apply to memory). We present the theory metaphorically and then the formal mathematical specification later.

In the next section, for a more quantitative description, we propose a working definition of category coherence, which we refer to as "semantic smoothness". De-

[1] In this study, "category" and "property" ("feature") are defined as a particular memorized set of entities from a universe. A "category" is linked to a set of "properties" , and the pattern of the set "properties" is "generalization" ("property weighting") of the category. "Similarity" is defined as psychological distance between a pair of "categories", "properties" or "generalization" in property space.

velopmental studies have indicated children to be efficient learners , who can systematically generalize even a novel noun category. Furthermore, a survey of children's early acquired noun categories suggested that their systematic generalization is consistent with the correlation between property and generalization. Therefore, children show category coherence (i.e., they share the systematic intuition to novel entities), and the survey provided a quantitative measure of this coherence.

Working Definition: "Semantic Smoothness" in Natural Categories

Many developmental studies using novel word generalization tasks have shown that children systematically attend to different properties when generalizing different types of entities, a process known as "fast mapping". For example, children generalized solid artifacts and non-solid substances based on the similarity of shape and material, respectively (Soja, Carey & Spelke, 1991). Therefore, children seem to solve the circular problem, but neither younger children nor late talkers show the same generalization pattern (Jones, 2003). This finding suggests that these differential weighting patterns are learned. An adult survey study of the similarities that characterize the first 300 nouns learned by children showed that their attentional biases in noun extension tasks reflect the regularities in the corpus of early noun categories. Specifically, there is a high correlation between category generalization (i.e., shape- or material-based category organization) and property (i.e., solidity) (Samuelson & Smith, 1999). In other words, this survey indicated that "property" (i.e., solid or non-solid) of natural categories predicted "generalization" (i.e., property weighting: shape- or material- based generalization), and vice versa. The semantic space would be in our terms "smooth" if the property-generalization correlation was universal in any semantic domain as well as "solidity of early acquired noun categories", , that is, the property difference between any two categories would be correlated to any difference in how those categories are generalized (see also Equation 11 for the mathematical definition). Furthermore, the smooth semantic space would form clusters that have a correlated property-generalization relationship; in other words, similarly distributed categories are basically grouped near each other (i.e., domain specific property weighting: Figure 1 (b)) [2]. Thus "smoothness" of the semantic space may be considered as a quantitative measure of category coherence. Here, in an empirical study, we investigate the smoothness of the semantic space of early-acquired nouns. The results indicate that natural categories have "smoothness", that similar categories (e.g., "cat" and "tiger") share a similar property and generalization pattern (whereas"cat" and "chair" share a dissimilar property and generalization pattern). Before a more detailed presentation of the theory, we consider how one recent theory of category development has dealt with this issue.

Insufficiency of a Previous Model

Rogers and McClelland (2004) proposed a statistical learning model that learns "coherent" categories more easily than "incoherent" categories. More specifically, they compared learning performance of their Parallel Distributed Processing model, which is sensitive to statistical properties of the learning set, when given a set of categories with correlated or randomly arranged properties. The model showed better learning of the correlated set than of the random set. They claimed that the statistical learning model accounts for category coherence because "coherent" categories (i.e., the correlated set) were easier to learn than those with randomly organized properties. However, there are three problems associated with their claim. First, they did not define the " category-property list" learned by the model. Recall that the original question of category coherence is why and how we "discover" such correlated organization in natural categories (i.e., the category-property list). Is the correlated "category-property list" in internal or external representation? Second, their model shows that correlated sets imply easy learnability, but their model does not show that easy learnability implies that the data set is correlated. Third, the model does not account for the deeper mechanism of category coherence because the PDP model is, without analysis of internal structure, a black box. What aspect of the model causes easier learning of the correlated set? In other words, category coherence was simulated at a surface level, but the essential question of why category coherence matters was left unanswered.

Efficiency in Semantic Cognition

Rosch, Mervis, Gray, Johnson & Boyes-Braem (1976) claimed that natural categories, particularly at the intermediate level, called the "basic level", trade off between having the most information (i.e., they are more abstract and include more classes) and being more differentiated (i.e., they are more concrete and include less classes). We argue that learning efficiency under the trade off between generalization and discrimination is a core mechanism of category coherence. We define "semantic memory" for the purposes of this paper as a set of categories and their features. A reasonable question to ask is what features should be represented for any category. In general, one would think that discriminating features should be represented. For example, both dogs and cats have "four limbs" so this feature does not discriminate between two categories. One solution to this is to store only exceptional features (or conjunctive features that as conjunctions are unique, e.g., "four-limbed, eyed, a pet, meat-eating, gather in crowds, ⋯" as a single feature only for "dog"). However, this means that what one knows about one category can not help in learning or making decisions about another. In the ideal choice of discriminating features, each category would be defined by a single feature that is unique to that category. This insight indicates that discriminability and generalization from knowledge about one category to another trade off.

[2] Consistent with the mathematical term "smooth", "smoothness" here refers to the probabilistic degree of local linearity of manifold (category-feature space) where generalization σ_i is curvature around μ_i .

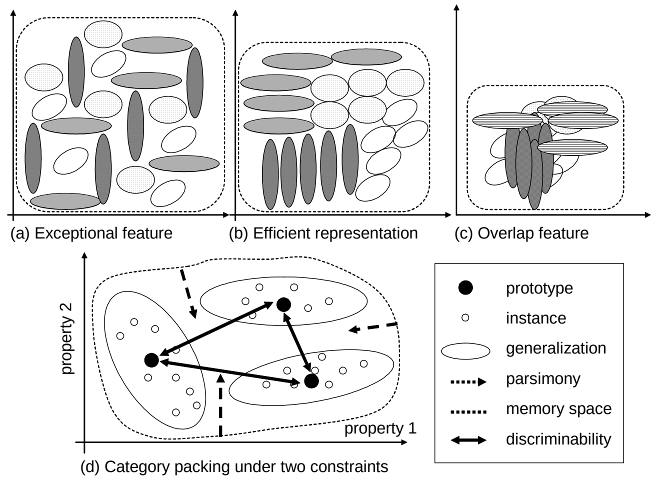

Figure 1: 20 categories (ellipses represent contour of probabilistic distributions) in property space. (a) Category coding with exceptional feature: more property space (dashed line) is needed for categorization. (b) Efficient category coding: less property space and more discriminability. Categories with similar generalization are localized. (c) Category coding with overlapped feature: feature is useless for categorization. (d) Category packing is considered the balanced optimal state under two constraints; memory parsimony (dashed arrow) and discriminability of category (bidirectional arrow).

An efficient semantic memory may try to optimize both, that is, it may work at some midlevel between these two extremes.

We describe this trade-off relationship in a simple computational form below. Assume that a category is formulated as a probabilistic distribution in psychological feature space (Figure 1). If one category is represented by an exceptional feature, the category representation must be sparse (Figure 1 (a): G gets smaller, but E_1 gets bigger in Equation 5), because other discriminable categories are unlikely to have this feature. In this case, the system needs more features and thus memory space (i.e., larger dashed-line enclosure). If, however, a category is represented by features that overlap with other categories, the category would not be discriminated well (Figure 1 (c): E_1 gets smaller, but G gets bigger in Equation 5). In this case, though the system needs less memory space, the categories are not discriminable. Efficient category coding should be a balanced state that emerges from the trade-off between the two limits (Figure 1 (b)). What then would emerge in such a case?

We describe the optimal state as "category packing", where the system packs categories of a particular shape close together, thus taking up less space overall (Figure 1 (d): both G, E_1 and L get smaller in Equation 5). Assume one creates optimally organized categories by moving the prototypes with fixed distributions. Alternatively, one could move distributions with fixed prototypes (note that this is not "category learning" but is just exploring optimal organization). This process is analogous to packing things into smaller space (categories or things avoid probabilistic or solid "collision", respectively). What category organization is most efficient for this packing? The most efficient packing of different sizes and shapes of things (or categories, as we propose here) involves packing similar shaped things together (e.g., angular things should be next to angular things, and rounded things should be next to rounded things). For categories, this means that categories with similar feature distributions should be closer in the feature space (Figure 1 (b) and Equation 11). In other words, semantic smoothness would emerge as a result of category packing and, vice versa, a smooth semantic space optimizes packing constraint. Next we briefly introduce the detailed formulation of our theory in a simple case in which categories are defined by prototypical representations.

Theoretical Formulation of Packing

We prove the equivalence between semantic packing and smoothness, under the simplification that each category is represented by its prototype and generalization pattern. Note that the packing process does not assume any predefined specific "property" or "category" in the packing process. $P_i(\theta)$, the probability of ith category occurrence given feature θ, is defined as a d-dimensional normal distribution, which has a mean vector (i.e., prototype) μ_i and covariance matrix (i.e., generalization or feature weighting) σ_i. The superscript t refers to transposition.

$$P_i(\theta) = ((2\pi)^d|\sigma_i|)^{-\frac{1}{2}} \exp(-\frac{1}{2}(\theta-\mu_i)^t\sigma_i^{-1}(\theta-\mu_i)) \quad (1)$$

The amount of overlap of multiple independent categories F is defined as the joint probability of category occurrences in the entire feature space Ω. The less F gets (i.e., the less overlap that occurs among categories), the more discriminability the feature space has.

$$F = \int_\Omega \prod_i^n P_i(\theta)^{\frac{1}{n}} d\theta \quad (2)$$

As normal distribution has reproductive property in multiplication, Equation 2 can be rewritten as a normal distribution $N(\theta|A^{-1}B, A^{-1})$ with mean $A^{-1}B$ and covariance A^{-1}. Assume that $A = \sum_i^n \sigma_i^{-1}$, $B = \sum_i^n \sigma_i^{-1}\mu_i$, $C = \sum_i^n \mu_i^t\sigma_i^{-1}\mu_i$, and $D = \exp(\frac{1}{2n}(B^tA^{-1}B - C))$. Then

$$F = D(|A|\prod_i^n |\sigma_i|^{\frac{1}{n}})^{\frac{-1}{2}} \int_\Omega N(\theta|A^{-1}B, A^{-1}))d\theta \quad (3)$$

. Assume that $G = \log(F)$, and the integral term should be one when Ω is whole space.

$$G = \frac{1}{2n}(B^tA^{-1}B - C - n\log|A| - \sum_i^n \log|\sigma_i|) \quad (4)$$

Optimization for only the constraint $\frac{\partial G}{\partial \mu_i}$ or $\frac{\partial G}{\partial \sigma_i}$ (i.e., discriminability in Figure 1) gives $(\mu_i - \mu_j)^t(\mu_i - \mu_j) \rightarrow \infty$ or $|\sigma_i| \rightarrow 0$, indicating an immense amount of feature space or an instance as a category (no generalization), respectively. Therefore, constraints to normal distributions $E_1 = \sum_i^n ||\mu_i||^2 = \sum_i^n \mu_i^t \mu_i$ and $E_2 = \log |A^{-1}|$ are necessary. For the cognitive process, the constraints E_1 and E_2 refer to the maintenance of constant memory space (i.e., parsimony in Figure 1) and generalization ranges, respectively. The Lagrange multiplier method is used for optimization of the constraints. The Lagrange equation with multiplier λ is $L = G+\lambda_1 E_1+\lambda_2 E_2$, which indicates semantic packing (L) optimizes both discriminability (G) and generalization (E_1 and E_2).

$$\frac{\partial L}{\partial \mu_i} = \frac{\partial}{\partial \mu_i}(G + \lambda_1 E_1) = -\sigma_i^{-1}(\mu_i - \bar{\mu}) + \lambda_1 \mu_i = 0 \quad (5)$$

where $\bar{\mu} = A^{-1}B = (\sum_i^n \sigma_i^{-1})^{-1} \sum_i^n \sigma_i^{-1}\mu_i$.

$$\mu_i = -(\lambda_1 \sigma_i - I)^{-1}\bar{\mu} \quad (6)$$

where I is the identity matrix. Therefore the relationship between a pair of categories when L is optimized as a function of μ is

$$\Delta\mu_{ij} = \lambda_1(\lambda_1\sigma_i - I)^{-1}\Delta\sigma_{ij}(\lambda_1\sigma_j - I)^{-1}\bar{\mu} \quad (7)$$

where $\Delta\mu_{ij} = \mu_i - \mu_j$ and $\Delta\sigma_{ij} = \sigma_i - \sigma_j$. Next, in addition to μ_i, L is optimized as a function of σ_i.

$$\frac{\partial 2G}{\partial \sigma_i} = -(\bar{\mu} - \mu_i)^t\sigma_i^{-2}(\bar{\mu} - \mu_i) + (A^{-1}\sigma_i^{-2} - \sigma_i^{-1}) \quad (8)$$

Next, in addition to μ_i, L is optimized as a function of σ_i. As $\frac{\partial L}{\partial \sigma_i} = \frac{\partial}{\partial \sigma_i}(2nG + \lambda_2 E_2)$, thus applying $\frac{\partial 2nG}{\partial \sigma_i} = \sigma_i^{-1}(\bar{\mu} - \mu_i)(\bar{\mu} - \mu_i)^t\sigma_i^{-1} + n\sigma_i^{-1}A^{-1}\sigma_i^{-1} - \sigma_i^{-1}$

$$\sigma_i\frac{\partial L}{\partial \sigma_i}\sigma_i = (\bar{\mu} - \mu_i)(\bar{\mu} - \mu_i)^t + (n + \lambda_2)A^{-1} - \sigma_i \quad (9)$$

As $\sigma_i\frac{\partial L}{\partial \sigma_i}\sigma_i - \sigma_j\frac{\partial L}{\partial \sigma_j}\sigma_j = 0$

$$\Delta\sigma_{ij} = \sum_{k=i,j}(-1)^{\delta_{ki}}(\hat{\mu} - \mu_k)(\hat{\mu} - \mu_k)^t \quad (10)$$

where $\delta_{ii} = 1$ when $i = j$, otherwise $\delta_{ij} = 0$.

Notice that σ_i is constant in Equation 7, and Equation 10 is $\Delta\sigma_{ij} \cong O(\Delta\mu_{ij})$. Consequently, the approximate monotonic relationship between $\Delta\mu_{ij}$ and $\Delta\sigma_{ij}$ with a given constant α (i.e., "smoothness") emerges, when $\frac{\partial L}{\partial \mu_i} = 0$ or $\frac{\partial L}{\partial \sigma_i} = 0$ (i.e., "packing").

$$||\mu_i - \mu_j|| \approx \alpha||\sigma_i - \sigma_j|| \quad (11)$$

In other words, semantic smoothness, which is the correlation between feature and generalization (Equation 11), is approximately equivalent to semantic packing. A learning system with smooth categories that optimize the packing principle, and vice versa.

An analytic solution to $\frac{\partial L}{\partial \mu_i} = 0$ is demonstrated as follows. Assume $E_1' = \sum_i^n \nu_i^t\nu_i$ where $\nu_i = \mu_i - A^{-1}B$ to be the constraint instead of E_1, and note that the replacement does not lose generality. Solving the Lagrange equation $L = \frac{G}{2} + \frac{\lambda}{2}E_1'$, we get $\frac{\partial L}{\partial \mu_i} = -\sigma_i^{-1}\nu_i + \lambda\sum_j^n(\delta_{ij} - \sigma_i^{-1}A^{-1})\nu_j$ where $\delta_{ii} = 1$ when $i = j$, otherwise $\delta_{ij} = 0$. Let $\nu = (\nu_1, \nu_2, ..., \nu_n)^t$ be the d-by-n-dimensional vector with ν_i as its ith elements, let $\mathbf{\Sigma}$ be the super matrix with σ_i as its ith diagonal elements, and $\mathbf{A}^{-1}$ be a super matrix with n^2 A^{-1} as its all elements. Then, we have

$$\nu - \lambda(\mathbf{\Sigma} - \mathbf{A}^{-1})\nu = 0 \quad (12)$$

Thus, Equation (5) ($i = 1, ..., n$) can be solved by ν as an eigenvector of $(\mathbf{\Sigma} - \mathbf{A}^{-1})$ in Equation (12).

Method

Survey Procedure

The first step in the simulation study was to collect data on the similarities of 48 nouns that are among the earliest learned by children (Fenson et al, 1993). To determine the relevant similarities across a broad range of properties, 104 Japanese undergraduates rated each noun category using 16 pairs of adjectives (Hidaka & Saiki, 2004). These adjective pairs are the potential features. Subjects used a 5-point scale to indicate how well the pair of adjectives described the items in the category (e.g., *large* = 5, *small* = 1). The 16 pairs of adjectives were selected by a pilot survey using 41 pairs collected from prior studies. We created questionnaires of 5 different orderings to cancel out the order effect. Participants completed the survey within one hour.

Stimuli

- **Adjective pairs (linguistic scales)**
  dynamic-static, wet-dry, light-heavy, large-small, complex-simple, slow-quick, quiet-noisy, stable-unstable, cool-warm, natural-artificial, round-square, weak-strong, rough hewn-finely crafted, straight-curved, smooth-bumpy, hard-soft.

- **Noun categories**
  butterfly, cat, fish, frog, horse, monkey, tiger, arm, eye, hand, knee, tongue, boots, gloves, jeans, shirt, banana, egg, ice cream, milk, pizza, salt, toast, bed, chair, door, refrigerator, table, rain, snow, stone, tree, water, camera, cup, key, money, paper, scissors, plant, balloon, book, doll, glue, airplane, train, car, bicycle

Analysis and Simulation

Correction of survey data The rating value was corrected by a logistic function to make the correlation between mean and variance zero. The original rating showed a small positive correlation between the deviation from the median and the variance, because an extreme rating (i.e., a rating near one or five) has smaller variance than a rating near the median. More specifically, the parameters of the logistic function are estimated to have zero correlation between $|x-b|$ and a standard deviation of x (x is rating, $b=3$, and $c=1.2$ in Equation (13)). The corrected mean and variance is used for analysis and simulation.

$$f(x) = (1+\exp((x-b)c^{-1}))^{-1} \quad (13)$$

Index of semantic smoothness Semantic smoothness, as predicted by Equation 11, was specifically calculated by norms of the mean vector and covariance matrix in the model. The mean vector and covariance (or correlation) matrix represent the mean and covariance across the 16-adjective ratings for a subject in the human survey data. The correlation and contribution of the norms of the mean vector and the covariance were used as an index of smoothness. The contribution of the major axis is calculated by principal component analysis, because the norms of both the mean and the covariance have variances. In other words, the coefficient of determination in regression analysis underestimates contribution because it supposes that only the dependent variance has error.

Simulation of packing category Two simulations were run. The first simulation involved semantic packing of randomly generated categories for specific visualization of coherent categories, and the other simulation involved reproducing category organization based on human survey data for testing the packing effect by fast mapping.

In the semantic packing simulation, we optimized the mean and covariance of several categories, in which the initial mean and covariance were generated randomly (i.e., the gradient method, in which the parameters were updated based on Equations 5 and 9). The smoothness index was measured after updating was performed 100 times. The updated final state refers to optimization in terms of balanced constraints.

In the simulation of survey data, the means of categories were reproduced by a solution of Equation 12 for a given covariance matrix of survey data. This simulation investigates the predictability of a prototype configuration of real data based only on a generalization pattern. The results were evaluated based on the correlation between the distances between all pairs of categories in the reproduced configuration and the original prototype configuration. The estimated number of degrees of freedom of the configuration is 752 (number of categories without pivot of rotation by property dimension is $(48-1)\times 16$).

Results

Figure 2 shows the relationship between the mean norm and the correlation norm of the survey data. The correlation and contribution are 0.466 and 0.733, respectively. The correlation and contribution of the smoothness index using covariance, rather than correlation, are 0.357 and 0.688, respectively. These results suggest that the investigated category set has smoothness. Figure 4 shows a simulated "packing" of 20 categories with randomly initialized prototypes and generalization in two-dimensional property space. Adjacent prototypes have

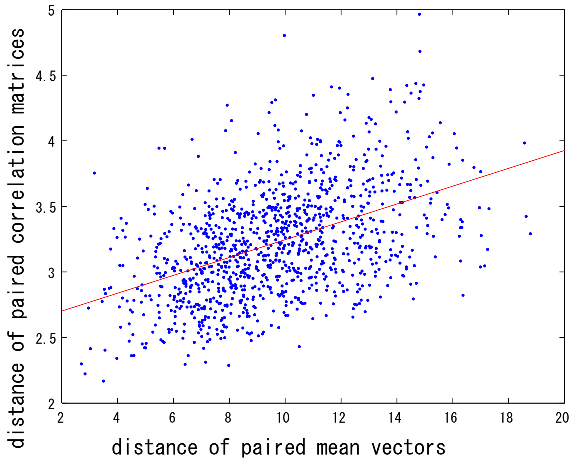

Figure 2: Scatter plot of mean vector norm (x axis: prototype dissimilarity) and correlation matrix norm (y axis: generalization dissimilarity) in survey data ($R = 0.466$)

similar generalization patterns (the result of a simulation is shown in Figure 4). Moreover, the mean norm and covariance norm of paired categories are shown (Figure 3). The average and standard deviation of the smoothness index for 100 simulations were .490 and .181 respectively. The correlation and contribution between the reproduced mean matrix in the simulation and the mean of survey data were 0.430 and 0.715, respectively. These results suggest that semantic packing could reproduce half of the data from only their generalization without any knowledge of the category configuration.

Discussion

The results obtained for the smoothness index of the survey data supports the hypothesis that property-generalization clusters were formed in not only specific domains (e.g., solidity-shape Samuelson & Smith, 1999) but also more generally. In other words, the smoothness index succeeded in quantifying category coherence as argued in previous studies. The result could be due to the specific properties (i.e., adjective pairs) and noun categories selected. The nouns categories examined might be efficient in terms of "packing" because most are in basic level categories, which are assumed to be the most efficient in terms of discrimination and generalization (Rosch et al., 1976). Moreover, the selected adjective pairs were chosen to be variable across categories and thus to discriminate among categories. Therefore, the set of categories and properties used in the survey might be particularly "efficient", as assumed in the packing theory.

In the simulation, though the prediction is in the inverse direction from generalization to feature, since the relation between them is symmetric (Equation 11), the simulation reproduced survey data, which indicates successful generalization for a novel category. Therefore,

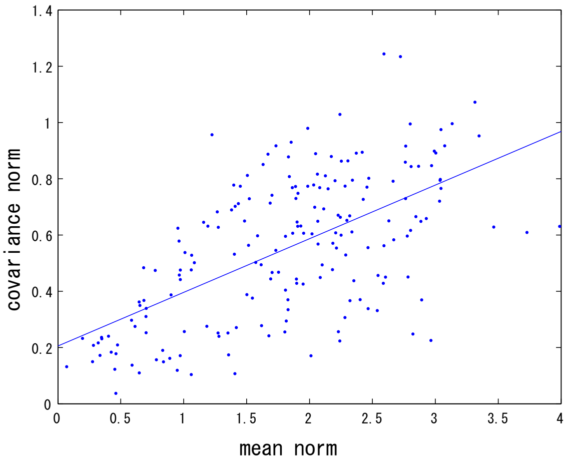

Figure 3: Scatter plot of mean vector norm (x axis: prototype dissimilarity) and correlation matrix norm (y axis: generalization dissimilarity) in simulation (smoothness index: 0.604).

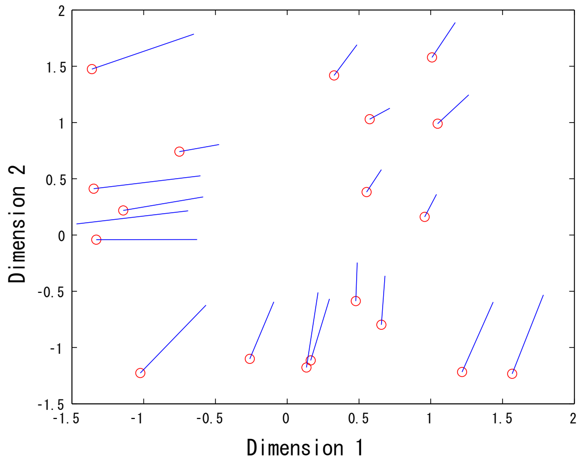

Figure 4: Two-dimensional feature space generated by optimization of 20 initially random categories. Circles and bars indicate prototype locations and generalization patterns (diagonal values of covariance matrices).

the partial success in reproducing the organization of early learned nouns suggests that a category system constrained by the semantic packing principle could generalize a category to new instances without the need for trial and error. This implies that the system would "know" the generalization pattern of a novel thing in a certain region of feature space. Young children show precisely this kind of knowledge in generalizing names for novel categories, typically referred to in the developmental literature as "fast mapping." Note that the system has no domain-specific meta-knowledge , as theory-theory claims, but smoothness, which emerges in efficient coding, works instead as meta-knowledge.

As a final conjecture, we propose semantic packing to be deeply related to basic-level category, because, as mentioned above, the basic-level category is the most efficient for semantic tasks (Rosch et al., 1976). According to Rosch, cue validity, the validity of a given cue x as a predictor of a given category y, is maximized in the basic-level category. The overlap measure defined in Equation 2 and cue validity might be considered to share the similar principle qualitatively, because less category overlap organization has a larger conditional likelihood given the cue (i.e., a subregion of feature space). More specific formulation is required to describe the mathematical relationship between cue validity and packing.

Acknowledgments

This work was supported by grants from Grants-in-Aid for Scientific Research from JMEXT (No. 15650046), the 21st Century COE Program (D-2 to Kyoto University), MEXT, Japan , and the National Institutes of Mental Heath, R01 MH60200-06.

References

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59* (5, Serial No. 242) Chicago: University of Chicago Press.

Gelman, R., & Williams, E. (1997). Enabling constraints on cognitive development. In D. Kuhn & R. S. Siegler (Ed&.). Cognition, perception and language. Vol. 2. Handbook of child development. (5th ed.) (pp. 575-630). (W. Damon, Ed.) New York: Wiky.

Hidaka, S. & Saiki, J. (2004). A mechanism of ontological boundary shifting , *Proceedings of the Twenty Sixth Annual Conference of the Cognitive Science Society* (pp. 565–570). Chicago, IL.

Jones, S.S. (2003). Late talkers show no shape bias in object naming. *Developmental Science, 6(5)*, 477–483.

Murphy, G.L. & Medin, D.L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92*, 289–316.

Rogers, T. T. & McClelland, J. L. (2004). *Semantic Cognition: a parallel distributed processing approach.* Cambridge, Mass. : MIT Press.

Rosch, E., Mervis, C. B., Gray,W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382–439.

Samuelson, L. & Smith, L. (1999). Early noun vocabularies: do ontology, category structure and syntax correspond?, *Cognition, 73*, 1–33.

Soja, N. N., Carey, S. & Spelke, E. S. (1991). Ontological categories guide young children's inductions of word meanings: object terms and substance terms., *Cognition, 38*, 179–211.