HOKKAIDO UNIVERSITY

Title	Computational Modeling of Word Learning Biases by using Known Words Meanings
Author(s)	Kurosaki, K.; Omori, T.
Citation	Artificial Intelligence and Soft Computing ~ASC 2005~ (9/12/2005 - 9/14/2005 Benidorm, Spain), 201-206
Issue Date	2005
Doc URL	https://hdl.handle.net/2115/17140
Type	journal article
File Information	ArtificialI&SC.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

COMPUTATIONAL MODELING OF WORD LEARNING BIASES BY USING KNOWN WORDS MEANINGS

Kousuke Kurosaki
Graduate School of Information Science and Technology
Hokkaido University
email: kurosaki@complex.eng.hokudai.ac.jp

Takashi Omori
Graduate School of Information Science and Technology
Hokkaido University
email: omori@complex.eng.hokudai.ac.jp

ABSTRACT
In the acquisition of their early nouns, it is well-known that young children have a tendency to understand the meaning of novel nouns based on the similarity of shape. This phenomenon is called ”shape bias.” Though this bias is remarkable in solid objects, it is reported that children overgeneralize and misapply the bias to non-solid objects. For this phenomenon, learning models using distributed representation are proposed. But the computational mechanism behind such children's behavior has not been clarified. In this paper we aim to clarify the more detailed computational mechanisms of these biases. Therefore, we explicitly define word meanings by a ”word category neuron model” and propose a ”nearest neighbor hypothesis” that represents a plausible mechanism for children's cognitive processes. Then, from a computer simulation based on the novel noun generalization task of developmental psychology, we show that the proposed hypotheses can better explain the emergence of word learning bias and deflection in children's word learning.

KEY WORDS
connectionist model, cognitive process, word learning bias, novel noun generalization task, vocabulary spurt

1 Introduction

When we meet a novel word and estimate its meaning, there are an infinite number of logically possible meanings. Nevertheless, children as well as adults can estimate the meaning of the words relatively well. Although children have little knowledge of the world, their learning is very fast, actually by only one presentation (fast mapping [1]). Such ”fast mapping” can't be explained by a machine learning algorithm based on trials and errors. However, children don't acquire this ability so early in their development. They need to hear a word repeatedly until they can produce it on their own and often make mistakes using those words [2]. Their meanings often differ from that of adults (e.g., overextension of meanings). Their early nouns include many names representing solid objects [3]. And the pace with which they acquire the meanings is slow, and the knowledge learned is unlikely to be fixed and surprisingly likely to fade away. Considering these phenomena, even though children are eagerly struggling with this difficult problem, the learning does not always progress smoothly and the efficiently. But after children's productive noun vocabulary exceeds 50 words, around eighteen months, their vocabulary begins to grow rapidly compared to previous learning speeds. This is called the ”vocabulary spurt.” To explain these phenomena, developmental psychologists have suggested some constraints or ”word learning biases.” They explained that the constraints narrow the infinite number of logically possible correspondences between a word and objects, and that the constraints enable the children to estimate a word's meaning more accurately. They revealed the existence of such ”biases” with which children focus on certain features when applying a novel name to objects [4] [5]. The problem is that these biases are just phenomenological explanations that don't explain why they exist or how they are realized in the human brain. In this paper we deal with just two biases: ”shape bias” [5] and ”material bias” [6]. Recently, the mechanisms of shape bias have begun to be investigated. Smith *et al.* [7] have suggested a mechanism based on psychological experiments, but didn't validate the mechanism's evidence. Samuelson [2] and Hidaka [8] validated it by a neural network model and a computer simulation. But their results ware still rough because of the stochastic model properties of the stochastic models they used.
Therefore, in this paper, we explicitly define a ”word meanings” model and then suggest a hypothesis for a more detailed mechanism of novel noun generalization that is also plausible as a brain processing model. We show that our hypothesis can better explain the phenomena of these biases by computer simulation.

2 Word Learning Biases and NNG Task

To study word learning biases, a novel noun generalization task (NNG task) is typically used. In this task, an experimenter names a novel exemplar object with a new nonsense name (e.g., ”dax”) in front of a subject. Next, the subject is presented novel test objects that aren't identical to the exemplar as a whole but are identical to particular features of the exemplar (only shape or material, etc.). Then the subject is asked which of the test objects can be given the same name as the exemplar. This task reveals the strategies used by children when naming novel nouns, that is, the features with which children generalize novel nouns.

”Shape bias” in NNG tasks is the tendency of children tend to generalize novel nouns by attending to the shape when an exemplar object is made of solid and hard material. This bias appears robustly. Meanwhile, in NNG tasks ”material bias” is the tendency of slightly older children tend to generalize novel nouns by concentrating on the material when a novel exemplar is made of non-solid material. When testing shape and material bias, an experimenter prepares two stimuli for each test object: shape-match test stimuli are identical to the exemplar in shape feature but different in other features and material-match test stimuli are only identical to the exemplar in the material feature. The experimenter asks children to grab the object which can be given the same name as the exemplar. There are two remarkable points in these biases. One, material bias is often less robust than the shape bias. Some studies reported that children generalized novel names for non-solid exemplars by material better than chance. But other studies reported that children didn’t systematically do so (see [2]). The other is that children sometimes overgeneralize the novel nouns by ”shape” when they see the substance (e.g., water, gel) to be named, unlike slightly older children and adults who tend to generalize by ”material”. We call this the ”overgeneralized shape bias” tendency.

3 Previous Works and the Problems

Smith *et al.* [7] have suggested a developmental mechanism of shape bias based on psychological experiments. They proposed the following four-step model for the development of shape bias. Here, we call this hypothesis ”Smith’s hypothesis.”
Step 1: Mapping name to objects. Children map individual names to individual objects one to one. Step 2: Learning first-order generalization. Children acquire knowledge about individual categories with individual names by generalizing the knowledge learned in Step 1. Step3: Learning higher-order generalizations. Children acquire higher-order knowledge about relationships between categories by generalizing the knowledge learned in Step 2. Step4: Rapidly acquiring novel names by using the knowledge of higher-order generalization. Children learn novel words faster by using the knowledge acquired in Step3, that is, biased attention to particular properties.
The knowledge acquired in Step 3 defines children’s tendency to generalize novel nouns. The number of word categories organized by shape similarity (called ”shape-based category” here) are significantly larger than those organized by material similarity (”material-based category”) in children’s noun vocabularies [3]. When we follow Smith’s hypothesis, such higher-order generalizations as ”__ are shaped” will occur in this case, and the knowledge will lead to the shape bias and the overgeneralized shape bias in NNG tasks. Smith *et al.* simply explained the bias phenomena from a viewpoint of learning without built-in constraints. But their hypothesis is too abstract to verify the detailed mechanism of how this higher-order knowledge is acquired and applied.
Samuelson studied the regularity of children’s vocabulary [3] and suggested that statistical regularities cause shape bias and the overgeneralized shape bias [2], which she called a ”statistical regularities hypothesis”. ”Statistical regularities” consist of two parts. One is the shape-dominance, which is the property where the number of shape-based categories is significantly larger than the material-based categories in an child’s vocabulary. The other is the high correlation between solidity, syntax, and category organization. She validated her hypothesis by computer simulations and showed that the emergence of these biases is largely dependent on shape dominance.
We support this ”statistical regularities hypothesis.” In particular, the necessity of shape dominance for the emergence of these biases seems important when we consider ”overfitting” in machine learning. But she just indicated the relationship between deflection in an child’s vocabulary and these attentional biases without clarifying how deflection leads to them. It is because she focused on the effects of statistical regularities and didn’t deal with ”word meanings.”
Hidaka & Saiki [8] tried to explicitly model Smith’s hypothesis by extending the category learning model ALCOVE [9]. But their explanation also remained qualitative for two reasons: one, though exemplar nodes have attention to the input layer, the values of the attention are shared in all exemplar nodes; two, the position vectors of all exemplar nodes aren’t updated. These two factors cause the difficulty of analyzing what was learned, though ALCOVE adopts a kind of local representation for the category and learns the attention to particular features.

4 Proposed Model and Rationale

In this study, we don’t use distributed representation for word meaning, and try to explicitly describe it with a ”word category neuron model” that we can easily analyze what is learned by the network and suggest a more detailed scope for novel noun generalization and resulting word learning biases. Though explicit modeling isn’t necessarily needed, it is an efficient approach to the more detailed mechanisms of word meaning acquisition.

4.1 Word Category Neuron Model

We assume a broad information space of multidimensional psychological input that involves multimodal information processed in the brain. Here we assume that inputs for a particular category/name form a distribution in this space. For example, though ”apple” is sensitive to the ”circularity” feature in shape dimensions, it is not necessarily restricted to only perfect circles but also permits a certain range of deformation. On the other hand, the word is undiscriminated in other dimensions such as ”time”, by suppressing

the effect of unnecessary inputs.
So, we define the "meaning" for a particular category with a particular name as multidimensional normal distribution in multidimensional psychological space that represents distribution as one neuron (Fig.1). We call it "word category neuron." However, for simplicity, we assume its covariance matrix to be diagonal. This means that there is no correlation between input dimensions. It is known that estimating the parameters of full covariance normal distribution requires many learning data and computation. Since non-correlated input is unlikely to occur in the real world, we need to improve this simplification.
$\boldsymbol{x} \in \Re^M, \boldsymbol{x} = (x_0 \ x_1 \ \ldots \ x_{M-1})^T$ is the input vector in the M-dimensional psychological space, and j is the ID number of a word category neuron, and $\boldsymbol{\mu}_j \in \Re^M, \boldsymbol{\mu}_j = (\mu_{j0} \ \mu_{j1} \ \ldots \ \mu_{j(M-1)})^T$ is the mean vector of the activations of input units, and $\mu_{ji} \in \Re$ is the mean value of activation of input unit i, and $\Sigma_j \in \Re^M \times \Re^M$ is a diagonal variance matrix, and $\sigma_{ji} \in \Re$ is the standard deviation of the activation of input unit i. Then, likelihood $p_j(\boldsymbol{x})$ of word category neuron j is calculated as below:

$$p_j(\boldsymbol{x}) = \frac{1}{(2\pi)^{\frac{M}{2}}|\Sigma_j|^{\frac{1}{2}}} \exp\Big(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_j)^T\Sigma_j^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_j)\Big) \tag{1}$$

Word category neurons are learned by the following algorithm.
(i) In the initial state, each word category neuron matches no particular word. Each word is also matched to no particular neuron.
(ii) Given input vector $\boldsymbol{x}$ and a word, all neurons calculate likelihood $p_j(\boldsymbol{x})$, and the "winning neuron c" that has highest likelihood is chosen.
(iiia) If a winning neuron hasn't been matched to any words and a given word hasn't been matched to any neurons, then they are matched with each other and the matching is fixed. In that case or if the winning neuron is the correct one matched to the given word, the winning neuron learns to increase its likelihood to the input. The loss function $\varepsilon_c(\boldsymbol{x})$ and update rules for each parameter are defined as below:

$$\varepsilon_c(\boldsymbol{x}) = -\log\left(p_c(\boldsymbol{x})\right) \tag{2}$$

$$\Delta\mu_{ci} = -\alpha(t)\frac{\partial\varepsilon_c(\boldsymbol{x})}{\partial\mu_{ci}} = -\alpha(t)\frac{\mu_{ci}-x_i}{{\sigma_{ci}}^2} \tag{3}$$

$$\Delta\sigma_{ci} = -\beta(t)\frac{\partial\varepsilon_c(\boldsymbol{x})}{\partial\sigma_{ci}} = -\beta(t)\Big(\frac{1}{\sigma_{ci}} - \frac{(\mu_{ci}-x_i)^2}{{\sigma_{ci}}^3}\Big) \tag{4}$$

$$\alpha(t) = \alpha_{\text{initial}}\left(1.0 - t/\text{EPOCH_MAX}\right) \tag{5}$$

$$\beta(t) = \beta_{\text{initial}}\left(1.0 - t/\text{EPOCH_MAX}\right) \tag{6}$$

$\alpha(t)$, $\beta(t)$ are the learning rate that declines with time to zero at the end of learning. EPOCH_MAX is the total loop count.
(iiib) If the winning neuron is an incorrect one matched to another word or if the given word is matched to another neuron, then the neuron learns to decrease its likelihood. Its update rule corresponds to the opposite one of (iiia). And if the given word is matched to another neuron, the neuron learns by the update rule of (iiia).
To retain learning stability, we set the lower limit of σ_{ji} to 0.1 and the range of μ_{ji} to $[-1, 1]$. The above algorithm corresponds to the one that extends LVQ [10].

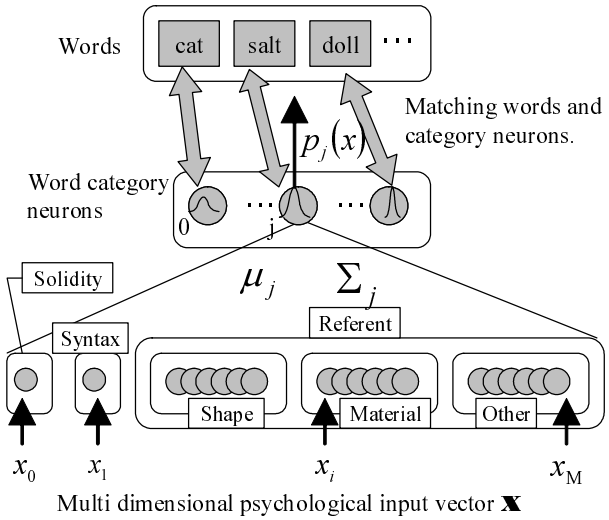

Figure 1. Word category neuron model

4.2 Nearest Neighbor Hypothesis

Consider that the input for a particular novel noun is just a point in multidimensional space. Nonetheless, children who have acquired fast mapping ability can smartly estimate correct distribution for the word category from the point. The estimated distributions contain certain "bias" that force them to pay selective attention to particular dimensions. This is the key point for considering such bias. Smith *et al.* [7] have suggested that children acquire simple but efficient higher-order generalized knowledge by just attending to shape properties which causes shape bias. But in that case, children should show stronger and more stable attention to shape features in NNG task. However, as Samuelson & Smith [3] reported, the rate of shape choice in their experiments was at most 70%. This rate is not so high enough to guarantee the acquisition of such effective knowledge as Smith *et al.* suggested.
So in this paper, we suggest the following hypothesis. Given multidimensional input and particular word, the child searches the nearest neuron c to the input in terms of Mahalanobis distance and applies the distribution of the winning neuron to estimate the distribution of the novel word. We call this the "nearest neighbor hypothesis." Concretely, Σ_c is assigned to Σ_{new}, and input vector $\boldsymbol{x}$ is assigned to $\boldsymbol{\mu}_{new}$. (Fig.2)

$$c = \arg\max_j \ p_j(\boldsymbol{x}) \tag{7}$$

$$\Sigma_{new} = \Sigma_c \tag{8}$$

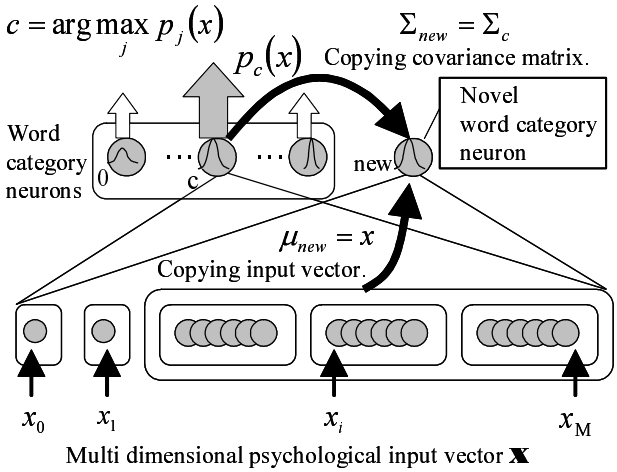

Figure 2. Mechanism of novel noun generalization

$$\boldsymbol{\mu}_{new} = \boldsymbol{x} \tag{9}$$

This estimation mechanism for novel noun distribution may sound too simple. But it's impossible to estimate the true distribution from one experience. This strategy seems better for children since they actually learn words fast from few experiences and little knowledge about the world.

When applying this hypothesis, shape choice and material choice will occur depending on the structural ratio in the child's vocabulary as below. If shape-based category words dominate the child's vocabulary, the category's neurons for the category become more likely to be chosen as the winner of the NNG task, and shape-based naming tends to take place. If material-based category words dominate, just the opposite tendency take place. However, since the number of material-based categories is much lower than shape-based categories, such neurons of shape-based category tend to win, even when naming a material-based exemplar; that is, the mechanism for the emergence of overgeneralized shape-bias.

4.3 The Course of Vocabulary Learning

We divide the children's learning process into two stages: a novice period and an expert period. The novice period is where children have slowly learned early 50 words by trial and error, corresponding to Steps 1 and 2 of Smith's hypothesis. Each "word category neuron" is matched with an individual word and updates its parameters by inputs. The algorithm proposed in Section 4.2 isn't used.

In the expert period, children rapidly learn the meanings of new words by using the knowledge learned in the novice period by the algorithm suggested in Section 4.2. This corresponds to the Steps 3 and 4 of Smith's hypothesis. For simplicity, the neurons added at this stage aren't used for following learning, and in our simulation they aren't updated at this stage.

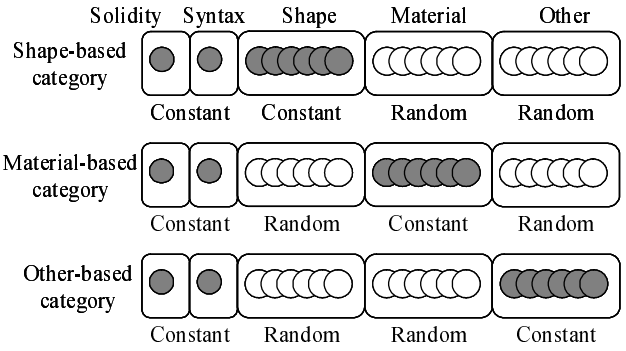

Figure 3. Representations of word categories. Colored circles mean constant values in word category, white circles mean random values in word category.

5 Simulation

In this simulation, we show that the names of solid exemplar objects tend to be generalized to shape-match test stimuli (shape bias), and the names of the non-solid exemplar objects also tend to be generalized to shape-match test stimuli (the overgeneralized shape bias).

5.1 Word Representation

Psychological input is represented as a 20-dimensional vector in which the one dimension denotes the property of "solidity," and another denotes "syntax," and the other eighteen dimensions denote "referent" properties. In the referent properties, six dimensions are used to denote "shape," "material" and "other" properties, respectively. Each dimension is represented as a real number of [-1,1].

Possible features for the solidity property include "solid," "non-solid," and "other" (ambiguous or non-noun). Each of them is assigned a value of 0.95, 0, and -0.95, respectively, and contains the noise between [-0.05,0.05]. Possible features for the syntax property include "count noun," "mass noun," and "other" (ambiguous or non-noun). Each of them is assigned a value of 0.95, 0, and -0.95, respectively, as well as the same noise.

Referent properties have three possible cases(Fig.3). One, the input object belongs to the shape-based category. The shape feature of the object is represented by arbitrary but fixed values. Material and other features are represented by arbitrary random values that change in every input. Two, the input object belongs to the material-based category. Its features consist of fixed material values, random shapes and other values as well. The last is where the input object belongs to the other-organized category whose features consist of the fixed other values, and the random shape and material values. All fixed feature values contain the noise within a range of [-0.05,0.05].

	solid	non-solid	other
solidity	66%	6%	28%
	count noun	mass noun	other
syntax	66%	10%	24%
	shape	material	other
category organization	54%	14%	32%

Table 1. Structural ratio of each property in a 50 vocabulary

cat,giraffe,owl,tiger,bus,motorcycle,train,doll,pencil,salt,cookie
jelly,cereal,bread,potato chip,spaghetti,gloves,underpants,tongue
shirt,hand,camera,glasses,potty,sofa,mop,towel,chair,movie,rock
purse,lawn mower,snow,garage,girl,peekaboo,drive,hug,read,last
swim,clean,night,sticky,you,what,into,more,is,if

Table 2. List of 50 words used in our simulation

5.2 Selection of Vocabulary

For this simulation, the vocabulary used must be appropriate words that children at that age generally tend to hear and produce. We also consider that the overgeneralized shape bias appears at 17-30 months of age [3] and that the vocabulary spurt occurs at about 18 months when the children can produce about 50 words. So we randomly chose 50 words from the "MacArthur Communicative Development Inventories(CDIs)" [11] which includes a typical vocabulary for 16-30 months toddlers. We assumed that each word forms its own category representation, as described above. Though Samuelson & Smith [3] and Hidaka & Saiki [12] evaluated the vocabulary representation by psychological experiments on adults, we gave the value by ourselves in this paper. To retain the structural ratio of children's vocabulary, we adopted the one Samuelson [2] used (Table 1). She used a 65%:35% ratio of noun and non-noun while we used 70%:30%. The vocabulary we chose is shown in Table 2.

5.3 Learning Data

For the novice period, we generated and used 100 input data for each word category based on the above definition that corresponds to the cases where children are presented with an individual word category 100 times. So a data set for 1 epoch consists of 100 data for each word category, 5,000 data in all, and the data set was used repeatedly in this period.

For the expert period, we prepared two data sets: a solid stimuli set and a non-solid stimuli set. Each data set consists of novel exemplars and two types of stimuli: shape-match test stimuli and material-match test stimuli. The novel exemplars for the solid stimuli set have values between [0.9, 1.0] for the solidity dimension and random values for other dimensions. Those for the non-solid stimuli set have the values between [0, 0.1] for solidity dimension and the random values for other dimensions. The shape-match test stimulus has identical fixed values with corresponding exemplar in solidity, syntax, and shape dimensions, and random changing values between [-1,1] in other dimensions. The material-match test stimulus has identical fixed values with corresponding exemplar in solidity, syntax, and material dimensions, and random changing values between [-1,1] in other dimensions.

The corresponding pair of shape-match test stimulus and material-match test stimulus is given for each exemplar, and the distances from each stimulus to the corresponding exemplar is compared. This comparison is done by 4 pairs for 1 exemplar, and we treated them as 1 exemplar set. The solid stimul and the non-solid stimuli sets consist of 121 exemplar sets, respectively.

5.4 Simulations

We defined all fixed parameters and the initial values of the learned parameters below, $\alpha_{\text{initial}} = 0.05$, $\text{EPOCH_MAX} = 1000$, $\beta_{\text{initial}} = 0.05$, initial values of $\mu_{ji} = [-0.001, 0.001]$, $\sigma_{ji} = 0.40$. After the novice period, we confirmed that these parameters were correctly estimated to form the distributions of the corresponding categories.

In the expert period, we didn't give the network the syntax dimension information to compare results with Samuelson [2]. The probability of shape choice for i-th exemplar set is defined below.

$$\text{shapechoice}(i) = \frac{1}{4}\sum_{j=0}^{3}\frac{\exp(\frac{p_s(i,j)}{\tau})}{\exp(\frac{p_s(i,j)}{\tau}) + \exp(\frac{p_m(i,j)}{\tau})} \tag{10}$$

$p_s(i,j)$ is the likelihood of j-th shape-match stimulus for i-th exemplar, $p_m(i,j)$ is the likelihood of j-th material-match stimulus for i-th exemplar, and τ is temperature set to 0.01. We calculated the probabilities of shape choices for 121 solid and 121 non-solid test sets, and compared their mean probabilities to the chance levels by t-test, respectively. The results confirmed that the probability of shape choice with solid stimuli and non-solid stimuli sets were both significantly different from chance levels, $t(120)$=8.64, $p \prec 0.001$ and $t(120)$=2.57, $p \prec 0.01$, respectively (Fig.4). In fact, the shape bias for novel solid objects and the overgeneralized shape bias for novel non-solid substances were generated on the basis of our hypothesis.

6 Discussion

We considered which neurons were used in NNG tasks. The neuron of shape-based category has a variance matrix with small variances for shape properties and large variances for others. These neurons are likely activated by shape-match test stimuli, showing shape-biased behavior. On the other hand, the neuron of material-based cate-

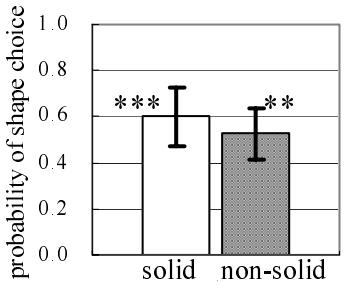

Figure 4. Result of novel noun generalization task. Mean probability of shape choice for solid stimuli set(left) and non-solid stimuli set(right) are shown.

gory has a matrix with small variances for material properties and large variances for others. These neurons are likely to be activated by material-match test stimuli, showing material-biased behavior.

So we investigated which neurons were actually used in our NNG task. For the solid stimuli set, neurons for the shape-based category tended to be chosen as the nearest one and neurons for the material-based category were rarely chosen. This result can be explained as follows. The solid stimuli set has a ”solid” property at the solidity dimension, and the categories that have ”solid” property are dominant in the set (Table 1). So, the known neurons that have ”solid” property are more likely activated than those with ”non-solid” property. In addition, the correlation between ”solid” property and ”shape” properties is high [3]. As the result, the neurons of the shape-based category are likely to be chosen as the nearest one, and shape choice probability is larger than chance.

On the other hand, in the non-solid stimuli set, neurons for the shape-based category also tended to be chosen as the nearest ones. However, the superiority of shape-based categories decreased compared with the solid stimuli set. This result can be explained as follows. Since the non-solid test set has ”non-solid” property, the known neurons that have ”non-solid” properties are more likely activated than those with ”solid” properties. However, solidity is just one dimension in multidimensional psychological space. And as above, categories with ”solid” properties are dominant in the set (Table 1), and the correlation between ”solid” and ”shape” properties is high [3]. As a result, a shape-based category neuron is likely chosen as the nearest one, and shape choice probability is larger than chance even for the non-solid test set. However, the balance between the ”non-solid” property effects and the dominance of the shape-based category causes the instability of material bias and overgeneralized shape-bias.

7 Conclusion

In this paper, we proposed a ”word category neuron model” based on a simple word meaning model to verify the more concrete mechanism of a typical children’s phenomena: shape bias for novel solid exemplars and a part of material bias for novel non-solid exemplars, that is, overgeneralized shape bias. We also proposed the ”nearest neighbor hypothesis” as the mechanism of children’s novel word generalization by using above model. Our hypothesis doesn’t contradict the hypotheses of Smith *et al.* [7] and Samuelson [2], because we explain the rationale behind their ”statistical regularities hypothesis,” that is, why shape-dominance is required to generate shape bias.

These research merits come from the explicit definition of the word meaning model that allowed the concrete analysis of the word learning process. Our future works includes the clarification of the process of bias appearance to discover a more concrete mechanism of material bias that hasn’t been so clear in experiments on children.

References

[1] S. Carey & E. Bartlett, Acquiring a single new word, *Papers and Reports on Child Language Development, 15*, 1978, 17-29.

[2] L.K. Samuelson, Statistical regularities in vocabulary guide language acquisition in connectionist models and 15-20-month-olds, *Developmental Psychology*, *38*(6), 2002, 1016-1037.

[3] L.K. Samuelson & L.B. Smith, Early noun vocabularies: do ontology, category structure and syntax correspond?, *Cognition*, *35*(3), 1999, 1-33.

[4] E.M. Markman & J.E. Hutchinson, Children’s sensitivity to constraints on word meaning: Taxonomic versus thematic relations, *Cognitive Psychology*, *16*(1), 1984, 1-27.

[5] B. Landau, L.B. Smith & S.S. Jones, The importance of shape in early lexical learning, *Cognitive Development*, *3*(3), 1988, 299-321.

[6] N.N. Soja, S. Carey & E.S. Spelke, Ontological categories guide young children’s inductions of word meaning: Object terms and substance terms, *Cognition*, *38*(2), 1991, 179-211.

[7] L.B. Smith, S.S. Jones, B. Landau, L. Gershkoff-Stowe & L.K. Samuelson, Object Name Learning Provides On-the-Job Training for Attention, *Psychological Science*, *13*(1), 2002, 13-19.

[8] S. Hidaka & J. Saiki, A Connectionist Account of Ontological Boundary Shifting, *Proceedings of the 11th International Conference on Neural Information Processing*, Springer, Calcutta, 2004, 282-287.

[9] J.K. Kruschke, ALCOVE: An Exemplar-Based Connectionist Model of Category Learning,, *Psychological Review*, *99*(1), 1992, 22-44.

[10] T. Kohonen, *Self-Organizing Maps* (Springer-Verlag, New York, 1995).

[11] L. Fenson, P.S. Dale, J.S. Reznick, E. Bates, D.J. Thal & S.J. Pethick, Variability in early communicative development, *Monographs of the Society for Research in Child Development*, *59*(5), The University of Chicago press, Chicago, 1994, 1-173.

[12] S. Hidaka & J. Saiki, A mechanism of ontological boundary shifting, *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*, Erlbaum, Chicago, IL, 2004, 565-570.