Brief article

Early noun lexicons in English and Japanese

Hanako Yoshida[a], Linda B. Smith[a,b,*]

[a]*Department of Psychology, Indiana University, Bloomington, IN, USA*
[b]*Program in Cognitive Science, Indiana University, Bloomington, IN, USA*

Abstract

Previous research suggests that children learning a variety of languages acquire similar early noun vocabularies and do so by similar and universal processes. We report here results from two studies that show differences in the early noun learning of English- and Japanese-speaking children. Experiment 1 examined the relative numbers of animal names and object names in vocabularies of English-speaking and Japanese-speaking children. English-speaking children's vocabularies were heavily lopsided with many more object than animal names whereas Japanese-speaking children's vocabularies were more evenly balanced. Experiment 2 used a novel noun extension task to examine what young children know about the different organizations of animal and artifact categories. The results suggest that early learners of English but not Japanese over-generalize what they know about object categories to animal categories. The role of culture, input and linguistic structure in early noun acquisitions is discussed. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Early noun lexicons; English; Japanese

1. Introduction

Languages differ in the meanings they lexicalize and in their syntactic structures. Yet, by most accounts, they do not differ greatly in their common nouns (Tardif, Gelman & Xu, 1999; Gentner, 1982; Sandhofer, Smith, & Luo, 2000). Further, by most developmental accounts, children learning all languages acquire names for concrete things easily and do so in pretty much the same way (e.g. Gentner,

1982; Markman, 1989). This makes sense, if as Gentner (1982) (see also Gentner & Boroditsky, 2001) proposed, the common nouns in a language as well as children's category learning are primarily determined by the perceptual coherence of concrete objects. However, the perceptual coherence of categories may not be the only relevant factor for lexical learning. Early lexicons are constructed in a linguistic as well as perceptual world. We present evidence of differences in the nouns known by English- and Japanese-speaking children that appear related to the two languages' different systems of individuation.

Linguistic individuation concerns the kinds of entities a language treats as discrete and countable. English treats animals and objects as countable, but not substances. Thus, "cat" and "cup" are both count nouns in English and obligatorily take the plural given multiple instances. In contrast, "mud" is a mass noun and is not pluralized. Lucy (1996) has argued that this system of individuation has an "object focus" and that it objectifies animates because it treats animate and inanimate objects in the same way, as countable discrete individuals.

Japanese presents a very different system of individuation. First, Japanese nouns that refer to multiple entities are not obligatorily pluralized. Nouns referring to multiple humans or young animals are optionally pluralized with the suffix *-tachi*. Second, when Japanese speakers do need to count discrete entities, they use a system of classifiers much as English speakers count "loaves of bread" or "panes of glass". Japanese classifiers used for animates do not overlap with those used for inanimates. Thus, in contrast to English, Japanese privileges animates, individuating animates differently from inanimate objects and substances.

2. Experiment 1

Experiment 1 asked whether the greater object focus in English and the greater animacy focus in Japanese are also evident in early noun vocabularies. We examined the numbers of animal and object names known by young Japanese-speaking and English-speaking children. Parents were presented lists of nouns commonly known by young children learning that language and were asked to indicate the nouns known by their child. We specifically used the American English and Japanese versions of the MacArthur Communicative Development Inventory (MCDI). These vocabulary checklists have been independently developed and normalized for many first languages including American English (Fenson et al., 1993) and Japanese (Ogura & Watamaki, 1997; see also Ogura, Yamashita, Murase, & Dale, 1993). Because the normalization process selects words for inclusion on the checklists that are known by at least 50% of 30-month-old children *learning that particular language*, the English and Japanese checklists differ in ways that are relevant to the hypothesis being tested. The English checklist contains 312 nouns and the Japanese checklist contains 353 nouns. But the Japanese checklist has absolutely more and proportionally more animal names (52 versus 43), absolutely and proportionally more people names (35 versus 29), and absolutely and proportionally more

names for body parts (33 versus 27). The two lists of typically learned early nouns contain similar numbers of names for inanimate objects.

Because of the differences in the two checklists, we conducted the experiment in two ways. First, we gave one group of mothers the appropriate checklists for their language, checklists that were thus different for the Japanese-speaking and English-speaking mothers. However, we selected specific lexical categories to query (animals, vehicles, toys, and small household items) such that the two lists contained proportionally similar numbers of animate and inanimate object names. Second, we gave another sample of Japanese-speaking and English-speaking mothers an abbreviated checklist that included only category names that were common to the two language-specific checklists. In these two ways, then, we control for the possibility that the number of items and their relative proportions on the language-specific checklists might influence mothers' judgments.

2.1. Method

2.1.1. Participants

The participants were 92 mothers of monolingual English-speaking children residing in Bloomington, IN and 92 mothers of monolingual Japanese-speaking children residing in Niigata, Japan. The children in both samples ranged in age from 15 to 44 months (means are 29.45 and 29.39 months for the English-speaking and Japanese-speaking children, respectively).

2.1.2. Language-specific checklists

The nouns on each checklist contained all nouns from four corresponding sections of the English and Japanese MCDI 'animal names', 'vehicles', 'toys', and 'small household items'. The latter three sections of the MCDI contain most of the object terms on the checklists. These four sections of the English MCDI contain 125 total nouns, 43 animal terms and 82 object terms. These four sections of the Japanese MCDI contain 143 total nouns, 52 animal terms and 91 object terms. Thus, by selecting these four sets of nouns, Japanese mothers had the opportunity of checking both more *animal terms* and *object terms*, a fact we hoped would moderate the inherent difference in opportunities to indicate known animal names on the two checklists. These language-specific checklists were given to 38 of the mothers of English-speaking children and 38 of the mothers of Japanese-speaking children.

2.1.3. Common-categories checklist

The common-categories checklist was constructed from the language-specific checklists described above such that the English and Japanese checklists contained nouns that named the same categories. Specifically, we made a checklist containing the 31 animal terms and 67 object terms that are listed on both the Japanese and English versions of the MCDI. The 98 nouns in each language are listed in Appendix A; they label categories that are known by 50% of Japanese- *and* English-speaking children at 30 months. The common-categories checklist was given to 54 of the English-speaking and 54 of the Japanese-speaking mothers.

2.1.4. Procedure

Participants were instructed to read each word on the presented list and indicate whether their child uses this word. Parents were asked to indicate the words in their children's *productive* rather than *receptive* vocabulary because the MCDI is a valid and reliable measure of production but not of comprehension (see Tomasello & Mervis, 1994, for a discussion of this issue).

2.2. Results

Fig. 1 presents the mean number of animal and object terms reported by mothers to be in their children's productive vocabulary. The data were analyzed by a 2 (Language) $\times$ 2 (Checklist) $\times$ 2 (Term: Animal/Object) analysis of variance for a mixed design. Japanese-speaking mothers reported that their children produced fewer nouns overall than did the English-speaking mothers ($F(1, 179) = 7.52$, $P < 0.01$). This difference may reflect cultural differences in mothers' criteria for saying that a child produces a word or it could reflect real differences in the size of early noun vocabularies (see Choi & Gopnik, 1995; Tardif, Gelman, & Xu, 1999). Overall, children were reported to know more names for objects than for animals ($F(1, 179) = 764.67$, $P < 0.01$). However, the "object" advantage is much greater in the vocabularies of English-speaking than Japanese-speaking children as indicated by a highly significant interaction between Language and Term ($F(1, 179) = 34.17$, $P < 0.001$). Post-hoc analyses (Tukey, $\alpha = 0.05$) indicated
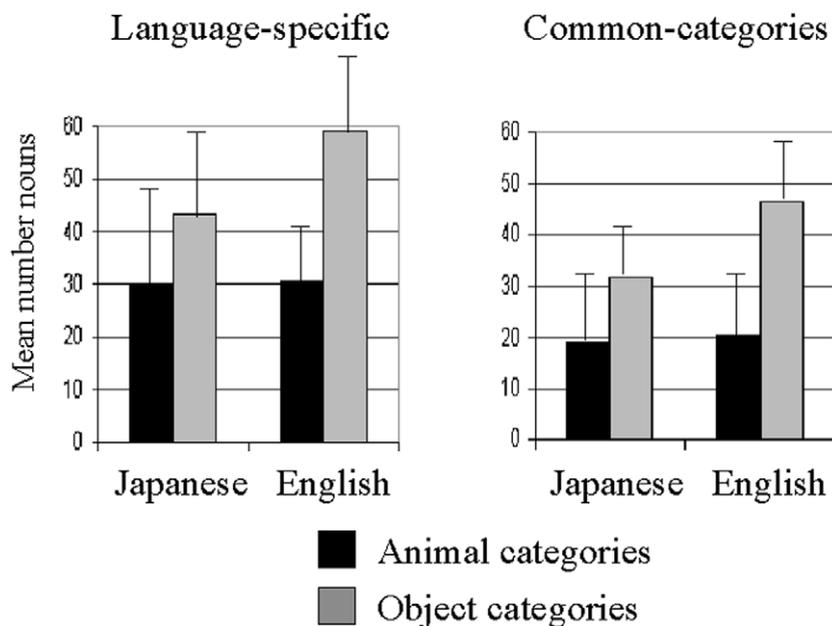
Fig. 1. Mean number of names of animals and objects in English-speaking and Japanese-speaking children's productive vocabulary by parent report for the two checklists.

that the number of object terms was significantly greater than the number of animal terms in the English-speaking children's vocabularies but *not* in the Japanese-speaking children's vocabularies. The analysis also yielded a significant main effect of checklist ($F(1, 179) = 21.92$, $P < 0.01$), which is expected since the language-specific checklists contain more items to be potentially checked off than do the common-categories checklists.

The main result is this: English-speaking children's noun vocabularies are strongly lopsided and heavy on the object side; the Japanese-speaking children's noun vocabularies, in contrast, are more evenly balanced between animal names and object names, patterns that match the focus of each language's system of individuation. Caution is warranted in making too strong a conclusion from these data, however. After all, these vocabulary differences were reported by mothers, and thus the language differences could be in the mothers' judgments and not in the children's actual vocabularies. Accordingly, in Experiment 2, we sought converging evidence by examining how English-speaking and Japanese-speaking children form new lexical categories.

3. Experiment 2

Previous research indicates that young English-speaking children (2-year-olds) extend newly learned animal and newly learned object names to new instances in the same way, by shape (Jones & Smith, 1998; Jones, Smith, & Landau, 1991). Older English-speaking children, in contrast, find shape similarity sufficient for extending an object name to new instances but insufficient for extending an animal name (Jones & Smith, 1998; Jones et al., 1991; Ward, Becker, Hass, & Vela, 1991). Instead, older children extend animal names conservatively to instances that are simultaneously similar to the original exemplar on multiple properties. We ask: is this developmental pattern specific to learning English, a language in which animate and inanimate objects are individuated in the same way? Alternatively, are Japanese-speaking children more sensitive to the richer multiple similarities that underlie animal categories and do they show that this knowledge more robustly or earlier than do English-speaking children?

3.1. Method

3.1.1. Participants

The participants were 26 monolingual English-speaking children living in Bloomington, IN who were between the ages of 20.4 and 38.0 months of age (mean 27.64 months) and 27 monolingual Japanese-speaking children living in Niigata, Japan who were between the ages of 20.1 and 36.4 months (mean 26.33 months). Children from each language group were randomly assigned to the two stimulus conditions, Animals or Objects, such that approximately equal numbers of boys and girls participated in each stimulus condition in each language group. A 2 (Language) × 2 (Stimulus condition) analysis of variance of the ages of the children yielded no

significant differences in the ages of the children assigned to the four conditions (all *F*s less than 1.00).

3.1.2. Stimuli and procedure

The parents of the participating children were presented the same language-specific noun checklists used in Experiment 1 and asked to indicate the nouns in their children's productive vocabularies.

In the name extension task, the experimenter presented the child with a novel three-dimensional representation of an animal or object and named it with a novel name. Then, with the exemplar still in view, a second object was presented and the child was asked to indicate "yes" or "no" as to whether the name also applies to that second object.

3.1.2.1. Pre-training. A pre-training task was used to familiarize children with the procedure. The pre-training exemplar was an orange cone with tufts on top. In the Animal condition, the object had eyes affixed and looked like a "martian". In the Object condition, there were no eyes. Four test objects – all with eyes appropriately affixed in the Animal condition and without eyes in the Object condition – were used. Two of the test objects matched the exemplar in shape, color, and texture and two differed from the exemplar in the same properties. Children in each condition were shown the exemplar, told its name, and then shown each test object individually and asked whether it was also called by the same name. For the English-speaking children, the experimenter introduced the exemplar by saying "This is a mobit" and asked about the test objects by saying "Is this a mobit?" For the Japanese-speaking children, the experimenter introduced the exemplar by saying "*Kore wa mobito*" and queried the pre-training test objects by saying "*Kore wa mobito kana?*" Children were given feedback on each trial and progressed to the experimental trials when they responded correctly (extending the name to the like objects and not extending the name to the different objects) on three successive trials. All children reached this criterion within eight trials.

3.1.2.2. Experimental trials. Children in each condition were tested with two unique exemplars and corresponding test objects. The exemplars and test objects in the Animal condition all had rounded body shapes and eyes; those in the Object condition were angular and without eyes. The two exemplars used in each condition are illustrated in Fig. 2. The exemplars and test objects were made from clay, sponge, cloth, wood, and plastic.

For each exemplar, three kinds of test objects were constructed: (1) Multiple similarities – two test objects were constructed that matched the exemplar in both shape, texture and material; one of these also matched in color and the other differed saliently from the exemplar in color; (2) Shape similarity only – one test object was constructed that matched its exemplar in only shape; (3) Non-shape similarities – two test objects were constructed that differed from the exemplar in shape but matched in either color or texture. Children were queried about each of the test objects for each exemplar twice in a randomly determined order for a total of 20
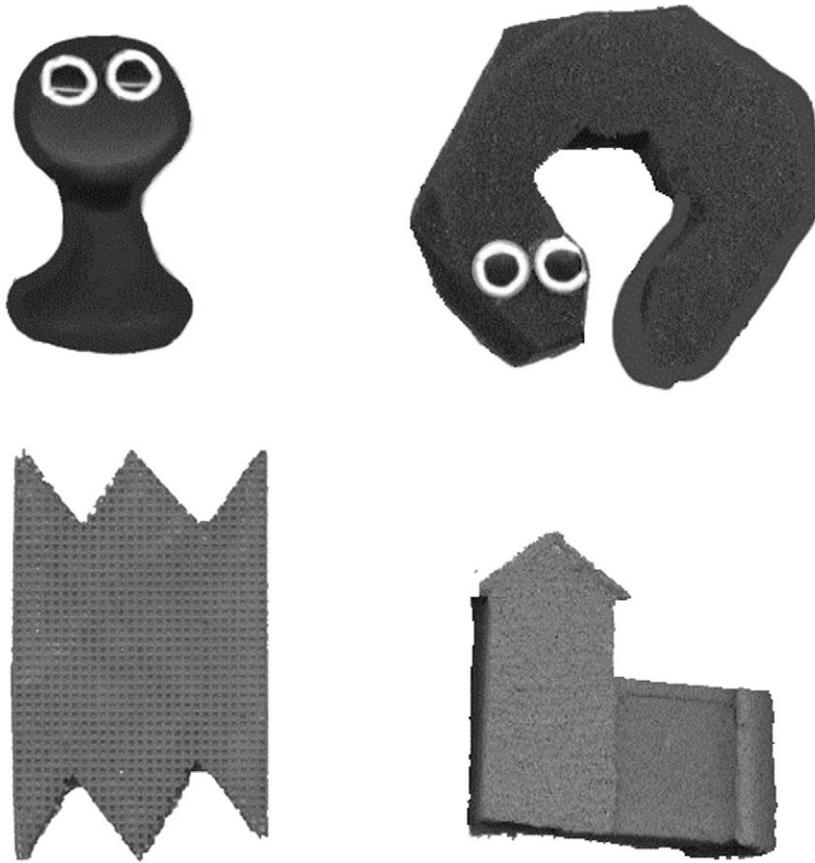
Fig. 2. The exemplars used in the Animal condition (top) and the Object condition (bottom) in Experiment 2.

trials – eight Multiple similarity trials, four Shape-only trials, and eight Non-shape trials. We used fewer Shape-only trials to minimize the imbalance between the expected number of "yes" and "no" responses in the Animal and Object conditions.

The procedure for the experimental trials was identical to that for the pre-training trials except that no feedback was provided. The novel names "kipple" and "teema" used for the English-speaking children and the names "keppuru" and "tema" were used for the Japanese-speaking children.

3.2. Results

Table 1 provides the means and ranges of the numbers of animal and object terms in the children's productive vocabularies by parent report. A 2 (Language) × 2 (Term: Object/Animal) × 2 (Stimulus condition: Object/Animal) analysis of variance for a mixed design yielded a reliable main effect of Language

Table 1
Means and ranges (in parentheses) of animal names and objects names in children's productive vocabularies (by parent report) in Experiment 2

	English-speaking children		Japanese-speaking children	
	Animal names	Object names	Animal names	Object names
Animal condition	26.7 (12–43)	59.0 (16–82)	28.6 (4–47)	36.8 (8–68)
Object condition	27.8 (2–43)	57.8 (11–100)	25.6 (5–52)	37.6 (5–78)

$(F(1, 49) = 5.95, P < 0.05)$. Again, in comparison to the English-speaking mothers, Japanese-speaking mothers reported that their children know fewer nouns. The analysis also yielded a main effect of Term $(F(1, 49) = 91.6, P < 0.001)$; overall, children are reported to know more object terms than animal terms. However, as in Experiment 1, there is a significant Language × Term interaction $(F(1, 49) = 14.93, P < 0.001)$. Again, the dominance of object terms over animal terms is considerably more pronounced in the English-speaking children's vocabularies.

Fig. 3 shows the results from the name extension task. The data were submitted to a 2 (Language) × 2 (Stimulus condition: Object/Animal) × 3 (Test object: Multiple/Shape/Non-shape) analysis of variance for a mixed design. The analysis revealed
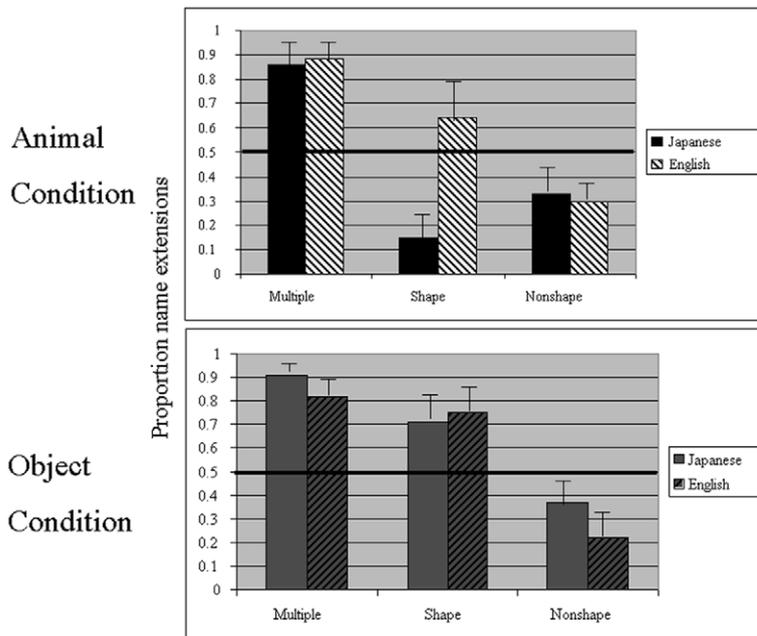
Fig. 3. Mean proportion of times Japanese-speaking and English-speaking children said that the exemplar's name extended to objects that matched the exemplar on multiple properties, on shape alone, or on single non-shape property in the Animal and Object conditions.

significant main effects of Stimulus condition ($F(1, 48) = 7.71$, $P < 0.01$), and Test object ($F(2, 96) = 67.62$, $P < 0.001$), and also significant interactions between Language and Stimulus condition ($F(1, 48) = 4.45$, $P < 0.05$), between Language and Test object ($F(2, 96) = 4.80$, $P < 0.01$), and between Stimulus condition and Test object ($F(2, 96) = 10.48$, $P < 0.001$).

The source of these interactions is obvious in Fig. 3. Both Japanese- and English-speaking children formed object categories by shape, extending the name to objects that matched the exemplar in shape alone as well as to objects that matched in shape and other properties. In contrast, in the Animal condition, Japanese-speaking children clearly limited their extensions to new instances that were similar to the exemplar in multiple ways; shape similarity alone was not enough to warrant the use of the same name. The English-speaking children, however, extended the animals' names on the basis of shape alone. More specifically, eight of the 14 English-speaking children in this condition extended the name to the test object that matched the exemplar in shape alone on over 75% of the trials. The differences in name extensions by English-speaking and Japanese-speaking children in the Animal condition line up with vocabulary differences reported by parents and thus provide evidence for the validity of those reports.

Japanese-speaking children in comparison to their English-speaking counterparts know relatively more animal names and they know more about how animal categories are structured differently than object categories. It is specifically the English-speaking children with the most lopsided, object-heavy, noun lexicons that generalize new object and animal names to new instances in the same way, by shape. Japanese-speaking children with more balanced vocabularies seem to know more about the different organizations that characterize animal versus object categories.

4. General discussion

The cross-linguistic differences reported here suggest that early noun learning is not insulated from the larger linguistic context in which that noun learning occurs. Instead, the relative frequencies of different kinds of noun categories in early English and Japanese lexicons appear correlated with the different systems of individuation in the two languages. This suggested correspondence between early noun vocabularies and individuation is reminiscent of the proposal by Whorf (1956) about languages as pattern systems that emphasize singular themes through multiple devices. The idea is that languages have themes such as an object focus or an animacy focus that are pervasively evident in multiple ways, for example in the lexical contrasts in the language, in syntax, in pragmatic rules, and in habitual manners of speaking. The present results suggest a different kind of correspondence between one aspect of language and another, a correspondence between systems of individuation and the *development* of noun lexicons.

The crucial question, however, concerns the causal mechanisms behind this correspondence. Several possibilities merit attention in future work. First, consistent with Whorf's idea of close correspondences between language and culture, the proximal

cause of the differences in the early noun lexicons may be differences in the frequency of object and animal names in the input. Japanese-speaking parents may talk about and name animals more than English-speaking parents and/or English-speaking parents may talk about and name objects more. In this case, the differences in the early noun vocabularies would be a product of parent values, values that may reflect (and be reflected in) language. Second, the vocabulary differences may be a more direct product of the two systems of individuation. Considerable previous research shows that young children use linguistic cues such as count/mass syntax and classifiers to figure out the referents and meanings of new nouns (e.g. Hall & Graham, 1999; Imai & Gentner, 1997; Soja, 1992). If English provides fewer cues to differentiate animal and object terms (or provides cues that combine them in the larger class of "discrete thing"), then it may take children learning English longer to discover the organizational differences that characterize the two kinds of categories. These cross-linguistic differences thus might derive from the broad effects of syntactic bootstrapping, the influence of syntactic devices such as count/mass syntax on the acquisition of noun meanings (see especially, Chierchia, 1994).

Both accounts – parental input and syntactic bootstrapping – are consistent with the possibility that the cross-linguistic differences observed here are *transient* developmental effects that will disappear as children's vocabularies grow. Nonetheless, these early differences in noun vocabularies may be profoundly important for what they can tell us about the developmental process, and how even the early learning of names for concrete things is a part of a larger organic whole in which language, culture, and the structure of the world conspire to create developmental paths.

Acknowledgements

Appendix A

Animals (real or toy)

1 bear	16 duck
2 bee	17 elephant
3 bird	18 fish
4 bug	19 frog
5 bunny	20 giraffe
6 butterfly	21 squirrel
7 cat	22 tiger
8 chicken	23 turtle

(*continued*)

9 cow	24 horse
10 lion	25 alligator
11 monkey	26 ant
12 mouse	27 sheep
13 penguin	28 wolf
14 pig	29 zebra
15 dog	30 animal
	31 deer

Vehicles (real or toy)

1 airplane	7 truck
2 bicycle	8 train
3 bus	9 helicopter
4 car	10 sled
5 fire truck	11 stroller
6 motorcycle	12 tricycle

Toys

1 ball	7 pencil	13 glue
2 balloon	8 toy	14 pen
3 block	9 bat	15 present
4 book	10 chalk	16 clay
5 bubbles	11 crayon	17 puzzle
6 doll	12 game	

Small household items

1 blanket	13 dish	25 telephone
2 bottle	14 wallet	26 toothbrush
3 bowl	15 purse	27 towel
4 box	16 fork	28 vacuum
5 broom	17 cup	29 basket
6 brush	18 glasses	30 camera
7 clock	19 keys	31 bucket
8 comb	20 medicine	32 can
9 money	21 radio	33 nail
10 paper	22 scissors	34 knife
11 picture	23 soap	35 tape
12 pillow	24 spoon	36 tissue

References

Chierchia, G. (1994). Syntactic bootstrapping and the acquisition of noun meanings: the mass-count issue. In B. Lust, M. Suner, & J. Whitman (Eds.), *Syntactic theory and first language acquisition: cross-linguistic perspectives*, *Heads, projections, and learnability* (Vol. 1). Hillsdale, NJ: Lawrence Erlbaum Associates.

Choi, S., & Gopnik, A. (1995). Early acquisition of verbs in Korean: a cross-linguistic study. *Journal of Child Language*, *22* (3), 497–529.

Fenson, L., Dale, P., Reznick, J. S., Bates, E., Hartung, J., Pethick, S., & Reilly, J. (1993). *MacArthur Communicative Development Inventories*. San Diego, CA: Singular.

Gentner, D. (1982). Why nouns are learned before verbs: linguistic relativity versus natural partitioning. In S. A. Kuczaj II (Ed.), *Language development, Language, thought and culture* (Vol. 2). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gentner, D., & Boroditsky, L. (2001). Individuation, relativity, and early word learning. In M. Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 215–256). Cambridge: Cambridge University Press.

Hall, D. G., & Graham, S. (1999). Lexical form class information guides word-to-object mapping in preschoolers. *Child Development*, *70* (1), 78–91.

Imai, M., & Gentner, D. (1997). A cross-linguistic study of early word meaning: universal ontology and linguistic influence. *Cognition*, *62*, 169–200.

Jones, S., & Smith, L. B. (1998). How children name objects with shoes. *Cognitive Development*, *13*, 323–334.

Jones, S. S., Smith, L. B., & Landau, B. (1991). Object properties and knowledge in early lexical learning. *Child Development*, *62*, 499–516.

Lucy, J. A. (1996). The scope of linguistic relativity: an analysis and review of empirical research. In J. C. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 37–69). Cambridge: Cambridge University Press.

Markman, E. M. (1989). *Categorization and naming in children: problems of induction*. Cambridge, MA: MIT Press.

Ogura, T., & Watamaki, T. (1997). *Japanese Communicative Developmental Inventories: user's guide and technical manual*. San Diego, CA: Singular.

Ogura, T., Yamashita, Y., Murase, T., & Dale, P. (1993). *Some preliminary findings from the Japanese Early Communicative Inventory*. Paper presented at the 6th International Conference for Child Language, Trieste, Italy.

Sandhofer, C. M., Smith, L. B., & Luo, J. (2000). Counting nouns and verbs in the input: differential frequencies, different kinds of learning? *Journal of Child Language*, *27*, 561–585.

Soja, N. (1992). Inferences about the meanings of nouns: the relationship between perception and syntax. *Cognitive Development*, *7* (1), 29–45.

Tardif, T., Gelman, S., & Xu, F. (1999). Putting the "noun bias" in context: a comparison of English and Mandarin. *Child Development*, *70* (3), 620–635.

Tomasello, M., & Mervis, C. B. (1994). The instrument is great, but measuring comprehension is still a problem. Commentary. *Monographs of the Society for Research in Child Development, 59* (5, Serial No. 242), 174–179.

Ward, T., Becker, A., Hass, S., & Vela, E. (1991). Attribute availability and the shape bias in children's category generalization. *Cognitive Development*, *6* (2), 143–167.

Whorf, B. L. (1956). *Language, thought and reality: selected writings of Benjamin Lee Whorf*. Cambridge, MA: MIT Press.