ELSEVIER

Cognition 73 (1999) 1–33

COGNITION

www.elsevier.com/locate/cognit

Early noun vocabularies: do ontology, category structure and syntax correspond?

Larissa K. Samuelson*, Linda B. Smith

Department of Psychology and Program in Cognitive Science, Indiana University, Bloomington, IN 47405, USA

Received 3 November 1998; received in revised form 19 March 1999; accepted 21 May 1999

Abstract

This paper examines children's early noun vocabularies and their interpretations of names for solid and non-solid things. Previous research in this area assumes that ontology, category organization and syntax correspond in the nouns children learn early such that categories of solid things are organized by shape similarity and named with count nouns and categories of non-solid things are organized by material similarity and named with mass nouns. In Experiment 1 we examine the validity of this assumption in a corpus of early-learned nouns and conclude that one side of the solidity-syntax-category organization mapping is favored. In our second experiment we examine the relation between early noun vocabulary development and novel word generalization. We find that children between 17 and 33 months of age do not systematically generalize names for solid things by shape similarity until they already know many nouns, and do not systematically generalize names for non-solid substances by material similarity. The implications for children's acquisition of the ontological distinction, count/mass syntax, and novel nouns are discussed.

Keywords: Ontology; Categories; Syntax; Vocabulary development; Nouns

1. Introduction

Children are smart noun learners. They are so smart that they seem to learn the whole category to which a novel noun applies from hearing a single instance named (Landau, Smith & Jones et al., 1988; Markman, 1989; Soja, Carey & Spelke, 1991;

* Corresponding author. Fax: +1-812-855-4691.
E-mail address: samuelso@indiana.edu (L.K. Samuelson)

0010-0277/99/$ - see front matter
PII: S0010-0277(99)00034-7

Markman, 1992; Waxman & Hall, 1993; Imai, Gentner & Uchida, 1994; Waxman & Markow, 1995; Smith, Jones & Landau, 1996; Imai & Gentner, 1997). This is particularly remarkable given the many different kinds of entities there are to be named – objects, substances, people, animals and places.

This paper is specifically concerned with young children's understanding of the different category organizations of objects versus substances. Interest in how children categorize and name solid objects and non-solid substances has generated much recent interest because of what such findings can tell us about children's ontologies and concepts of individuation. As Soja et al. (1991) noted 'For human infants, solid objects are bodies that are cohesive, bounded, spatio-temporally continuous and solid or substantial; they move as connected wholes independently of one another, on connected paths through unoccupied space…Non-solid substances are spatio-temporally continuous and substantial, but not cohesive or bounded; they do not retain either their internal connectedness or their external boundaries as they move and contact one another.' (p. 183) Concrete objects and substances thus present children with perceptual evidence pertinent to a fundamental distinction between a discrete object and continuous mass. That children are sensitive to the perceptual cues that indicate a discrete solid object has been well documented in studies of infants (Gibson, 1979; Gibson & Walker, 1984; Streri & Spelke, 1989; Baillargeon & DeVos, 1991; Xu & Carey, 1996; Needham, Baillargeon & Kaufman, 1997). The concern of this paper is specifically how the perceptual differences between solid and non-solid things may be mapped onto language and category organization.

One task that experimenters have used to probe these issues is a novel word learning task (Landau et al., 1988; Soja et al., 1991; Soja, 1992; Smith et al., 1996; Gathercole & Min, 1997; Imai & Gentner, 1997; Jones & Smith, 1999). In these tasks, the child is presented with a novel exemplar and is told its name, e.g. 'this is the mel'. The child is then shown test items that match the exemplar on one perceptual property or another and is asked which of these test items can be called by the same name as the exemplar. Numerous studies using this task have demonstrated that when the exemplar is a solid object, 2- and 3-year-olds robustly generalize the novel name to new instances by shape (Landau et al., 1988; Soja et al., 1991; Imai & Gentner, 1997). However, when the exemplar is formed from non-solid stuff, children typically do not generalize the name by shape. Instead they are more likely to generalize the name to new instances that are made from the same material as the exemplar (Dickinson, 1988; Soja et al., 1991; Soja, 1992; Imai & Gentner, 1997).

These findings have potentially important implications for semantic and syntactic interactions in early word learning. On the semantic side, children's differing generalizations of names for solid and non-solid exemplars have been interpreted in terms of an ontological distinction between objects and substances (Soja et al., 1991; Soja, 1992). The idea is that children construe solid things to be bounded and individuated objects and non-solid stuff to be non-discrete and continuous and moreover they know that different properties define object and substance categories. On the syntactic side, the count/mass distinction has been interpreted s being fundamentally about individuation (Pelletier, 1979a; Gordon, 1988; Soja et al., 1991; Bloom, 1994;

Gathercole, Cramer, Somerville & Haar, 1995; Imai & Gentner, 1997). Count nouns are nouns that can take the plural (e.g. cookies and studies); mass nouns cannot be pluralized (e.g. applesauce, research). Conceptually, then, count nouns refer to entities (concrete and abstract) that are discrete whereas mass nouns refer to entities (concrete and abstract) that are continuous (see Pelletier, 1979a) for arguments for and against this view). Critically, young children learning English know something about the semantic force of the count/mass distinction at least in terms of how it relates to solid objects and non-solid substances. In novel noun learning tasks, young children are more likely to generalize a novel name for a solid exemplar by shape when it is presented in a count noun syntactic frame (e.g. 'this is a mel') and more likely to generalize a novel name for a non-solid exemplar by material when it is presented in a mass noun syntactic frame (e.g. 'this is some mel,'; Soja, 1992).

The key issue that has motivated the previous developmental research is the relation between emerging semantic and syntactic knowledge. Quine (1960) hypothesized that the ontological categories of object and substance emerged as a consequence of language learning, the product of contingencies between language and perception. Soja et al. (1991) offered an alternative hypothesis: that children approach the task of language learning with a pre-existing set of ontological categories and more specifically with the knowledge that solid whole objects are organized into kinds by their bounded shapes but that non-solid substances are organized into kinds by their material.

On the whole the evidence pertinent to these two hypotheses is mixed and complicated. Children learning English begin to show sensitivity to count/mass syntax in their third year (Gordon, 1988) but may not have an adult-like understanding of the distinction until many years later (Gathercole, 1983; Gathercole, 1985). Two-year-olds learning English consistently generalize novel names for solid things by shape, regardless of whether that name is embedded in count, mass or neutral syntactic frames. However, across experiments, 2-year-olds generalizations of novel substance names by material is less robust and is aided by mass syntax (Soja et al., 1991; Soja, 1992; Imai & Gentner, 1997). Moreover, children learning Japanese, a language without count/mass syntax, also generalize names for solid things in different ways than they generalize names for non-solid things (though in ways that do not correspond perfectly to their peers learning English; Imai & Gentner, 1997).

We seek insight into these previous findings by asking what we take to be a logically prior empirical question: How well do solidity, syntax, and category organization correspond? Do count nouns name solid things in shape-based categories and do mass nouns name non-solid substances in material-based categories? The answer is clearly 'no' in adult language given abstract nouns such as *research* and *study*. It is also clear that they do not correspond perfectly even for concrete nouns; for example, wood is a mass noun that names a solid substance and bubble is a count noun that names a non-solid object. However, solidity, syntax, and category organization may correspond well enough among the nouns that children learn early to be developmentally consequential – that is, if Quine is right, to help children discover the ontological distinction or if Soja et al., are right, to help children discover the count/mass syntactic distinction.

In brief, the two hypotheses that motivated previous research – that semantics leads syntax versus syntax guides semantic development – both presume an early correspondence among the perceptible properties that distinguish objects and substances, on the one hand, and syntax and category organization on the other. Moreover, a distinction between objects and substances and/or knowledge of count/mass syntax can explain children's rapid learning of nouns only if there is – at least statistically and at least among the words and categories children learn early – a correspondence between the perceptible properties of objects and substances and shape-based versus material-based category organization and/or between syntax and shape-based or material-based category organization.

Accordingly, in Experiment 1 we seek evidence for the hypothesized correspondences in a corpus of early learned nouns. In Experiment 2 we examine the developmental relation between the kinds of nouns that individual children know and their novel noun generalizations when the named entity is solid or non-solid.

2. Experiment 1

The goal of this first study is to see whether the correlates of an object-substance distinction proposed in the past really do exist among the nouns that children learn early. Following the lead of Soja et al. (1991) and subsequent developmental studies concerning this distinction (Gordon, 1985; Gathercole et al., 1995; Hall, 1996; Gathercole, 1997; Imai & Gentner, 1997), we use solidity as an index of the consortium of perceptible properties pertinent to the object-substances distinction. That is, in the experiences of young children, solid things have cohesive bounded shapes that do not change when touched whereas non-solid substances have variable shapes and disperse into disconnected portions when touched.

We ask two questions about how solid and non-solid entities are named in early vocabularies. First, we ask whether solidity is correlated with syntax. If solid things with bounded shapes are conceptualized as individuals then they should be named by count nouns; if non-solid substances are conceptualized as continuous masses, then they should be named by mass nouns. Second, we ask whether solidity is correlated with category organization. If solid things are conceptualized as unitary whole objects with a fixed shape, than the shape of the whole should be defining of category membership. Analogously, if non-solid substances are conceptualized as a continuous mass without fixed boundaries, then the material and not the shape of the mass should be defining of category membership.

To address these issues, we studied a corpus of 312 nouns taken from the toddler form of the MacArthur Communicative Development Inventory (Fenson, Dale, Reznick, Bates, Thal & Pethick, 1994). This form of the MCDI is a parental checklist of words designed to measure the productive vocabulary of children between 16 and 30 months of age. Given that the MCDI was developed from extensive studies of parents reports of the words children produce and laboratory measures of children's vocabulary (Fenson et al., 1994), these nouns serve as a reasonable proxy for the nouns children learn early.

We determined the syntactic category of each noun by asking the intuitions of adults. The everyday use of nouns is fluid and there are contexts in which many nouns can be used in both count and mass syntactic frames (Pelletier, 1979a). Consider for example, 'I'll have two waters please' and 'I'll have two glasses of water, please'. Therefore we asked naive native speakers to indicate the use of these nouns in everyday discourse. We wanted a consensus judgment of the everyday usage and thus set a conservative criterion of 85% agreement among adult judges to designate any individual noun in the corpus as count or mass.

We also asked adults to tell us the solidity/non-solidity of items named by each noun and about the similarities in shape, material and color of the instances named by each noun. Although adult intuitions of the properties common among instances of a category are likely to be imperfect, they are also likely to be correlated with the real similarities that do exist among members of the same category. Moreover, such judgments have been shown to predict adults' performances in various kinds of categorization tasks (Rosch, Mervis, Gray, Johnson & Boyes-Braem, et al., 1976; Malt, 1994). Because we are interested in the properties that are sufficiently true about the category of things named by each noun to matter in children's noun learning, we again set a conservative criterion of 85% agreement among adults in order to designate any individual noun as referring to entities of a particular kind.

2.1. Method

2.1.1. Participants

Twenty-six undergraduate students at Indiana University participated for course credit. The participants were recruited from two sections of an advanced laboratory course. The participants were randomly assigned to one of two groups with thirteen adults in each group. One group of adults was asked to judge the syntactic class of each noun, the other group of adults was asked to judge both whether items in the labeled category were solid or non-solid and whether they were alike in shape, color or material.

2.1.2. Materials

The studied corpus consisted of nouns from nine noun sections of the MCDI; 'animal names,' 'vehicles,' 'toys,' 'food and drink,' 'clothing,' 'body parts,' 'small household items,' 'furniture and rooms,' and 'outside things'. Thus, the corpus of words studied consisted of 312 nouns commonly found in the productive vocabularies of children between 18- and 30-months-of-age. Booklets containing the words from the nine noun sections were constructed. Each booklet contained instructions and the list of words. In the syntax condition, the instructions explained the distinction between count and mass nouns and gave several examples of the appropriate syntactic category of familiar nouns. The adults were encouraged to consider how they would use each word in the context of everyday discourse and use these examples as the basis of their classifications. Response sheets allowed participants to indicate that the noun was a count noun, mass noun, or neither, i.e. unclear. The instructions for the category organization task asked participants to indicate for each

word which perceptible properties were characteristic across instances of the named category. Participants made independent judgments of within category similarity on shape, color or material and thus could judge a noun as referring to objects that were similar on any one of these dimensions, any combination of these dimensions or none of these dimensions. The instructions for the solidity task asked participants to indicate for each word whether instances of the named category were solid, non-solid, or neither clearly solid nor non-solid.

2.1.3. Procedure

Participants were tested in groups. The instructions were presented verbally as well as in writing and all the participants' questions were answered prior to the judgments. Judgments were made on scantron sheets for later scoring by computer.

2.2. Results

Each noun in the corpus was categorized as a count noun, mass noun, or neither a count or mass noun if 85% of adults (11/13) agreed upon that designation (binominal probability, $P < 0.05$ for each noun). A noun was designated as ambiguous in syntactic class if there was not at least 85% agreement among adults. Likewise each noun was categorized as referring to solid, non-solid, or neither solid or non-solid things if 85% of adults agreed (binomial probability, $P < 0.05$, for each noun designation). If they did not, the noun was classified as ambiguous with respect to solidity. There was only one word (owie/boo boo) for which 85% of adults chose the 'neither solid or non-solid' option. Due to the low frequency of the 'neither' choice by raters, we collapsed this category with 'ambiguous' as defined. For the category organization judgment, each word was categorized as referring to a category of things that shared similarity in shape (and/or color and/or material) if 85% of adults made that designation (binominal probability, $P < 0.05$ for each classification of each noun), and as ambiguous with respect to category organization if agreement among adults did not meet this criterion for at least one of these three dimensions.

The number and percentage of the 312 nouns that were classified as count and mass nouns, or as ambiguous with respect to syntax are given in Table 1. Many more nouns were judged to be count nouns ($n = 232$) than mass nouns ($n = 31$). A substantial number of nouns ($n = 49$) did not meet the agreement criterion. These syntactically ambiguous items were primarily items like 'Coke' and 'cake' that are used in both syntactic frames (e.g. 'I'd like four Cokes please,' or 'Can I please have some Coke').

The number and percentage of nouns that were classified as names for solid or non-solid entities or for things ambiguous with respect to solidity are also given in Table 1. Adult judgments indicate that many more nouns on the MCDI named solid things ($n = 197$) than non-solid substances ($n = 14$). However, a substantial number of nouns in this corpus are ambiguous with respect to solidity ($n = 101$). This is less surprising than it might at first seem. The 'ambiguous' items include words like 'egg' that refer to entities that change state or are malleable as well as more abstract items such as 'story' or 'game'. Other examples of these nouns are listed in Table 2.

Table 1
Total number and percentage of the 312 nouns that were classified as count, mass or ambiguous in syntax; as referring to a category of solid objects, non solid substances or ambiguous with respect to solidity; and as referring to a category of things that share similarity in shape, color or material or for which similarity in perceptual properties among instances is ambiguous[a]

Syntax		Count	Mass	Ambiguous
Total		232	31	49
Percent		74%	10%	16%
Solidity		Solid	Non-solid	Ambiguous
Total		197	14	101
Percent		63%	4%	32%
Category organization	Shape	Color	Material	Ambiguous
Total	150	11	51	122
Percent	48%	4%	16%	39%

[a] Nouns could be classified referring to a category whose members share similarity on more than one perceptual property.

It is not unreasonable to assume that the items that give adults pause when thinking about solidity/non-solidity also do not present clear perceptual evidence to children.

Table 1 also provides the number and percentage of nouns classified as referring to categories of objects organized by shape, color, or material, as well the number and percentage of 'ambiguous' categories that that did not meet our 85% agreement criteria for within-category similarity. Note that individual nouns could be classed as referring to a category of objects similar in more than one dimension. Thus, by our

Table 2
Examples of words classified as ambiguous with respect to solidity that were also judged to be count nouns, mass nouns, or ambiguous in syntax[a]

Syntactic class	Count	Mass	Ambiguous
Total in class	232	31	49
Total also classified as of ambiguous solidity	65	16	20
Examples	Pancake	Butter	Tuna
	Hose	Cereal	Cake
	Cheek	Gum	Ice cream
	Pillow	Jell-O	Light
	Towel	Peanut butter	Soap
	Diaper	Play dough	Tissue
	Sock	Snow	Shorts
	Flower	Grass	Cloud

[a] Also given are the number of nouns classified as each syntactic class and the number of those that were also classified as ambiguous with respect to solidity.

criteria *crayon* is both a shape-based and a material-based category whereas *ball* is a shape-based category. As can be seen in the table, by adult judgments, this corpus of early nouns contains many names for categories organized by shape ($n = 150$) and very few for categories organized by color ($n = 11$) or material ($n = 51$). However, many categories were ambiguously organized by this measure ($n = 122$). For some of these words, such as 'light', 'kitchen' and 'wind' the ambiguous classification likely reflects the fact that shape, color and material are perceptual dimensions that can not be readily applied to instances of the category. For other words, such as 'animal' and 'food', the disagreement among adults may reflect the superordinate nature of the category; neither all animals nor all foods are highly similar along any of the three perceptual dimensions queried.

Overall then, if the MCDI checklist is a good proxy for the nouns children learn early, and if the adult judgments are a good proxy for the syntactic category of these nouns and the perceptual properties within the named categories, then the most dominate segments of children's early noun vocabularies consist of count nouns, names for solid entities, and names for categories organized by shape. Early noun vocabularies do not include many mass nouns, names for non-solid entities, or names for categories organized by material.

How frequently do syntactic category and the solidity of the named entities correspond in this corpus? Note that this question can be asked in two directions; we can ask both if count nouns tend to label categories of solid things and if the names for categories of solid things tend to be count nouns. Fig. 1 presents the data in the syntax to solidity direction. The percentage of nouns judged to be count nouns, mass nouns, or ambiguous in syntax that were also judged to refer to solid and non-solid things is shown. As can be seen in the figure, 71% of the nouns on the MCDI that adults judged to be count nouns were also judged by adults to refer to a category of solid things. Only 35% of the nouns on the MCDI that adults judged to be mass nouns were also judged by adults to refer to categories of non-solid things. Further, none of the nouns that adults judged to be count nouns were judged to refer to categories of non-solid things, and only 13% of the nouns judged to be mass nouns were also judged to refer to categories of solid things. In addition, the percentage of nouns classified as ambiguous with respect to solidity is much larger for the mass nouns (52%) compared to the count nouns (28%). In brief, the count nouns that children learn early refer to solid things, but the mass nouns they learn early refer to a broader range of non-solid entities and entities of more ambiguous solidity.

Fig. 2 presents the data in the solidity to syntax direction. As can be seen in the figure, 85% of the nouns that adults judged to refer to categories of solid things were also judged to be count nouns. And, 79% of the nouns adults said referred to categories of non-solid things were judged to be mass nouns. None of the nouns that adults judged to be labels for categories of non-solid entities were judged to be count nouns. Among the names children learn early, solid things are named by count nouns and non-solid things are named by mass nouns.

Considering the two directions jointly, solidity is not distributed equally across syntactic categories, $\chi^2(4) = 99.30$, $P < 0.001$. However, the relation is not symmetric. Given a count noun in this corpus, one can be fairly certain that the

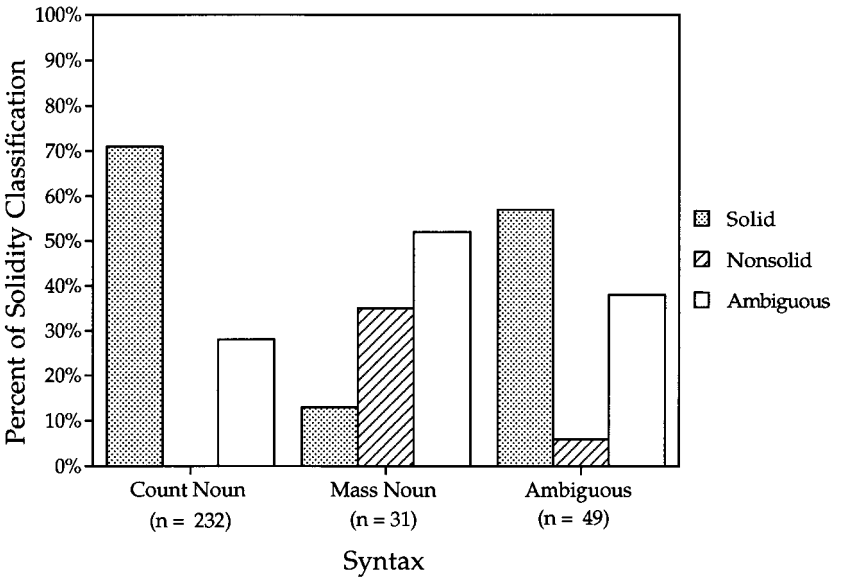

Fig. 1. Percent of nouns classified as count nouns, mass nouns or ambiguous with respect to syntax that were also classified as referring to categories of solid things, non-solid things, or categories ambiguous with respect to solidity.

referent is solid; however, given a mass noun one can not be *so* sure the referent is non-solid. On the other hand, given a solid thing one can be quite certain it will be labeled by a count noun and given a non-solid thing one can be quite certain it will be labeled by a mass noun.

The relation between syntax and the organizing perceptual features of categories is shown in Fig. 3. As can be seen in the figure, shape is the most pervasive within-category similarity for count nouns, $\chi^2(4) = 101.99$, $P < 0.001$, whereas material is the most pervasive within-category similarity for mass nouns, $\chi^2(4) = 86.49$, $P < 0.001$: 59% of nouns judged to be count nouns name categories in which instances share similarity in shape and 58% of nouns judged to be mass nouns name categories in which instances share similarity in material substance. However, the mapping between count and mass nouns and the organizing perceptual features of the named categories is far from perfectly regular. By adult judgments 38% of count nouns and 42% of mass nouns name categories in which within-category similarity is ambiguous on the queried dimensions. Thus, within this corpus of nouns, one would be best off generalizing a count noun on the basis of shape and a mass noun on the basis of material, but even these best options would produce erroneous generalizations some of the time.

The relation between solidity and within-category similarity is pictured in Fig. 4. As can be seen, nouns that adults judged to refer to categories of solid things were most often judged to also refer to categories organized by shape, $\chi^2(4) = 55.34$, $P < 0.001$. In contrast, nouns that adults judged to refer to categories of non-solid

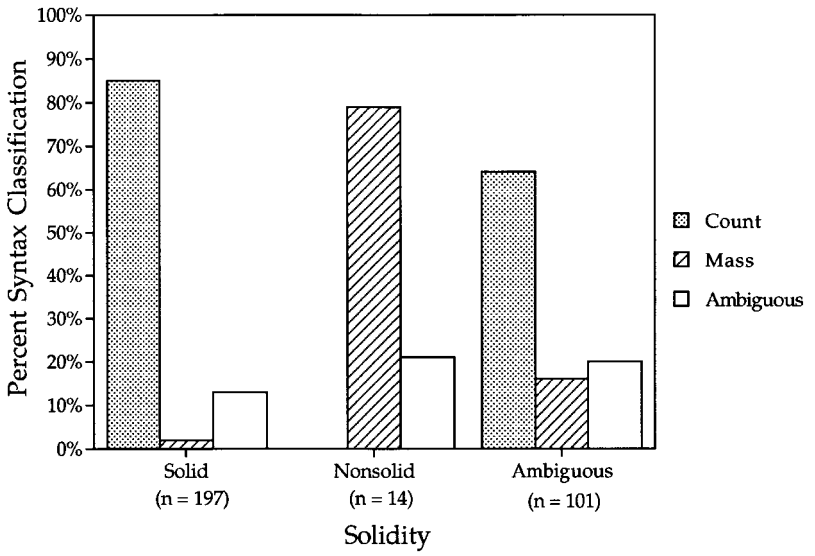

Fig. 2. Percent of nouns classified as referring to categories solid things, non-solid things or to categories ambiguous with respect to solidity that were also classified as count nouns, mass nouns, or nouns ambiguous with respect to syntax.

things were judged most often to also refer to categories organized by material substance, $\chi^2(4) = 44.54$, $P < 0.001$. The mapping between solidity and shape is somewhat more systematic than that between non-solidity and material substance:

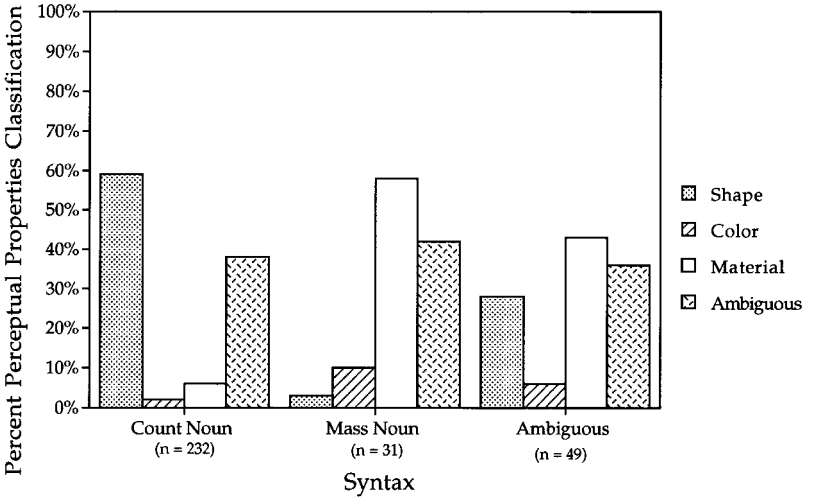

Fig. 3. Percent of nouns classified as count nouns, mass nouns or ambiguous with respect to syntax that were also classified as referring to categories organized by similarities in shape, color, or material, or to ambiguously organized categories.

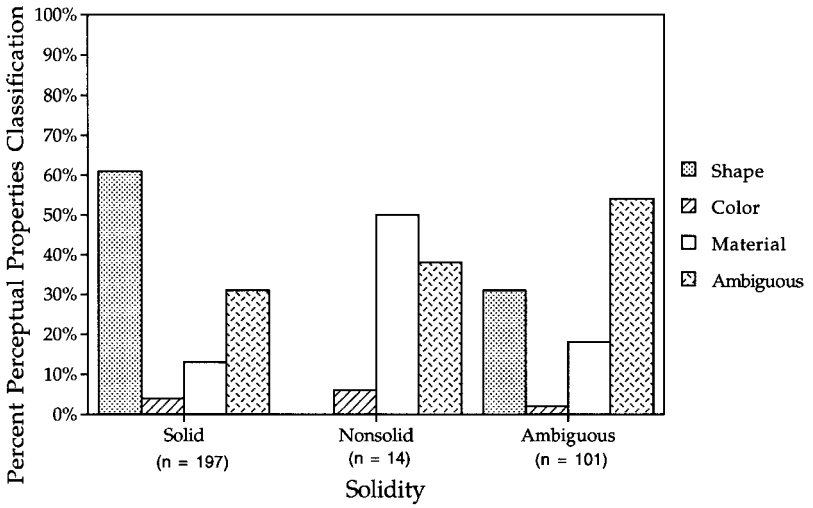

Fig. 4. Percent of nouns classified as referring to categories solid things, non-solid things or to categories ambiguous with respect to solidity that were also classified as referring to categories organized by similarities in shape, color, or material, or to ambiguously organized categories.

61% of the words that adults judged to refer to categories of solid things were also judged to refer to categories organized by shape and 31% did not meet our criteria for unambiguous category organization; however, 50% of the words that adults judged to refer to categories of non-solid things were judged to refer to categories organized by material substance and 38% did not meet our criteria for unambiguous category organization.

Fig. 5 presents a different view of the regularities in this corpus – one that emphasizes both the number of nouns contributing to the correspondences among syntax, solidity and category organization and the lopsidedness of those correspon-

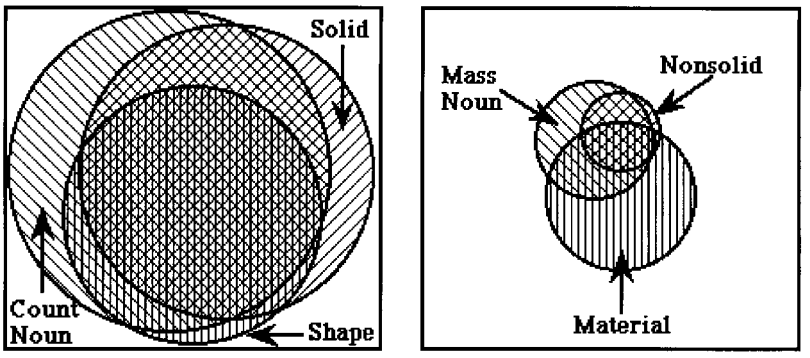

Fig. 5. Graphical representation of the relationship between solidity, syntax and category organization in the studied corpus of 312 nouns. See text for details.

dences. The outer squares each represent the universe of the 312 nouns. On the left, the inner circles depict the relative proportion of count nouns, names for categories of solid entities, and names for categories organized by shape, as well as the amount of overlap among these categories. On the right, the inner circles represent the relative proportion of mass nouns, names for categories of non-solid entities, and names for categories organized by material substance, as well as the amount of overlap among these categories. As is evident, most of the nouns children learn early are count nouns, names for categories of solid things or names for categories organized by shape. Moreover, the correspondences among these are high in all directions. The number of nouns that are mass nouns, names for categories of non-solid entities, or names for categories organized by material substance are much fewer. And, the correspondences are weaker and more asymmetric. Many non-solid things are named by mass nouns. However, many mass nouns also name things of ambiguous solidity, but in *absolute* numbers more things of ambiguous solidity are named by count nouns ($n = 65$) than by mass nouns ($n = 16$). Moreover, neither mass syntax nor non-solidity line up well with material-based categories.

How much do these conclusions depend on our use of an 85% criterion to classify individual lexical items? We examined this issue in a second analysis of the corpus with a 60% agreement criterion. That is, in this new analysis, for any individual noun to be classified as referring to objects similar in shape, for example, eight of 13 adult judges had to agree. This weaker criterion resulted in more coherence on the count-solid-shape side. Indeed, using the 60% criterion, 194 nouns (as compared to 110 nouns by the 85% criterion) lie in the intersection of solid, shape-based, and named by a count noun. Further, by this criterion 204 (98%) of the 209 total shape-based categories are categories of solid things named by count nouns. On the mass-non-solid-material side, 20 nouns (as compared to 6 nouns by the 85% criterion) lie in the intersection of non-solid, material based, and named by a mass noun, and this lesser criterion led to many more nouns being designated as referring to things in material-based categories (164). However, by the 60% criterion 127 of these material-based categories (77%) are neither named by a mass noun nor composed of non-solid things. Overall, then, the results are globally the same by both the 60% and the 85% agreement criteria: This early corpus contains many count nouns that name solid things by shape, very few mass nouns that name non-solid stuff, and mass noun syntax, non-solidity, and category similarity in material do co-occur but only weakly.

2.3. Discussion

This pattern of results is consistent with either the syntax-teaches-ontology view or the ontology-precedes-syntax view. Overall, however, the data fit the idea that an ontological distinction between objects and substances precedes or exists independently of the count/mass distinction better than it fits the idea that count/mass syntax teaches that distinction. Simply, the solidity of the named object is a somewhat better predictor of syntax and of category structure, in this early corpus, than is syntax a predictor of solidity or category structure.

However, the data in Fig. 5 also suggest a modified version of this ontology-first conclusion. The lopsidedness of this early lexicon suggests that children may start with just one half of the object-substance distinction; they may know only that solid objects are named by shape but not know anything systematic about the how categories of non-solid entities are organized. This is suggested both by the greater number of nouns naming solid than non-solid things in this early noun corpus, and by the more consistent category organization of solid as compared to non-solid things. This solids-first hypothesis fits past findings from novel noun generalization tasks showing that names for solid things are consistently generalized by shape but names for non-solid things are more inconsistently generalized by material. This solids-first hypothesis is also consistent with one interpretation of Gentner's natural-partitions hypothesis (Gentner & Boroditsky 1999). By this hypothesis children learn object names earliest that refer to easily individuated entities, that is, discrete things with bounded shapes.

By Gentner's natural-partitions hypothesis, the link between solid individuated things and categories organized by shape predates language learning and resides in the correlation between solidity, connected movement, and constant shape (see also Xu, 1997). The observed correspondence between solidity and the shape-based categories in Fig. 5 is certainly consistent with this view. These ideas lead to the prediction that children should be biased to attend to shape when naming solid things at the *start* of noun learning.

An alternative view is also tenable, however. The correspondence between solidity and shape-based categories is so strong that it could teach children to attend to shape when a novel solid thing is named. In other words, whereas syntax may not teach a distinction between objects and substances, children's attention to shape when naming solid things could nonetheless be learned in the course of early word learning. Specifically, a child's bias to attend to shape when naming a novel solid thing could be a generalization from the regularities among the nouns already known by that child. If this is so, children should not show a bias to attend to shape when learning the name of a novel solid thing (or a bias to attend to material when naming a non-solid thing) until after they already know many such names. We examine these issues in Experiment 2.

3. Experiment 2

This experiment focuses on the origin of the link between solidity and category organization by shape. It is this link that comprises the evidence for an ontological distinction in novel noun generalization tasks. We seek evidence on two specific issues: (1) the developmental priority of knowledge that solid things are named by shape over knowledge that non-solid things are named by material; and (2) the developmental relation between children's novel noun generalizations and the category structure of the nouns they already know. We address these two issues by examining the compositional structure of children's vocabularies at different points

in noun learning and by examining the relation between the words children know and their novel noun generalizations.

3.1. The developmental trajectory of noun acquisitions

If children begin noun learning with pre-lingual biases then they should learn nouns that fit with these pre-existing biases first. For example, children should learn names for solid things in categories that are also organized by shape before those that are in ambiguously organized categories or in categories organized by material. And, if children have pre-linguistic assumptions about the category organizations of non-solid substances, then one might expect that they would learn names for non-solid things in categories organized by material before names for non-solid things in categories that are ambiguously organized. Further, if a pre-linguistic understanding of ontological categories guides the acquisition of syntax children might also learn count nouns that label solid objects before other count nouns and they might learn mass nouns that label non-solid substances before other mass nouns.

In contrast, if children do not begin the noun learning process with already established biases, then the particular nouns they learn early must reflect their individual experiences. However, as children begin to learn some nouns, they should learn about the most prevalent statistical and perceptual regularities first. Thus, if the regularities found in Experiment 1 reflect those in the nouns individual children learn, this hypothesis suggests that the correspondence between being solid and being in a member of a category organized by shape might teach children to attend to shape when naming a solid thing.

Thus the hypotheses that children begin with only pre-existing knowledge about how solid things are named, and the hypotheses that children learn (and use to learn new lexical items) the statistical regularities most prevalent, differ in their suggestions of how children should enter the noun learning task. However, these two hypotheses make the same general prediction concerning the composition of individual children's early noun lexicon (at least once learning has progressed) – names for solid things should dominate.

3.2. The assumed meaning of novel nouns

Children's novel noun generalizations provide a measure of their assumptions about the potential meaning of nouns. If children begin noun learning with a conceptual understanding of the difference between solid things and non-solid things, they should distinguish between novel names for solid objects and non-solid substances even before they know many names for solid objects and non-solid substances. They should generalize a novel name for a solid thing to other solid things of the same shape and a novel name for a non-solid thing to other non-solid things of the same material from the beginning of noun vocabulary development. If they begin with only an understanding that solid things are named by shape, then, at the start of noun learning they should show only this knowledge in novel noun generalization tasks.

If, however, the statistical regularities among early learned nouns teach children how nouns are likely to map to categories, then children should *not* distinguish

between solid and non-solid things at the start of noun learning. Rather, this knowledge should develop as children acquire nouns. Given the results of Experiment 1, this means that as children acquire count nouns, names for solid objects and names for shape based categories they should learn that solid things are named by shape because this correspondence is well represented in the early-learned nouns. In contrast, children should not learn much about the category structure of non-solid entities, since the correspondence between non-solidity and material-based categories is not well represented among the nouns they know.

3.3. Experimental design

Experiment 2 examined these issues by measuring the number of nouns from the MCDI known by children between 17 and 32 months of age with each noun classified according to solidity, category organization and also syntax using the 85% criterion of Experiment 1. Children were also tested in a novel noun generalization task with three kinds of stimuli; solid, non-solid, and solid but non-rigid (e.g. sponge). We included a solid but non-rigid stimulus set to capture the ambiguous solidity reflected in adult judgments of items such as 'towel' and names for articles of clothing. In this task, novel names were presented in a neutral syntactic frame (i.e. 'This is my lom') so that we could measure the strength of the link between the perceptible properties of solid and non-solid things and category organization unaffected by children's emerging knowledge of count/mass syntax. In other words, this experiment is principally concerned with the relation between solidity and category organization, and therefore syntactic knowledge is neither manipulated nor tested directly in this experiment.

3.4. Methods

3.4.1. Participants

Seventy children between the ages of 17 months, 16 days and 33 months 18 days, participated. Children were recruited from birth announcements in local newspapers. The thirty-six males and thirty-four females were all from middle-class, English-speaking families. Twelve total children were excluded from the final sample, seven because they did not complete the task, two because they picked more than one object on an experimental trial, and three due to experimenter error. The mean age of the children who did not finish the task was 25 months 7 days with a range of 18 months 8 days to 31 months 25 days. One of the oldest children from this group was scared by one of the stimulus items, two pieces of silver worm-like plastic glued together. This item was subsequently replaced for the remaining children. The final sample included fifty-eight children with equal numbers of males and females.

3.4.2. Stimuli

The stimuli consisted of three sets of novel objects (Fig. 6). Each stimulus set contained an exemplar and four test objects. Two test objects in each set matched the exemplar in shape but not material. The other two test objects in each set matched

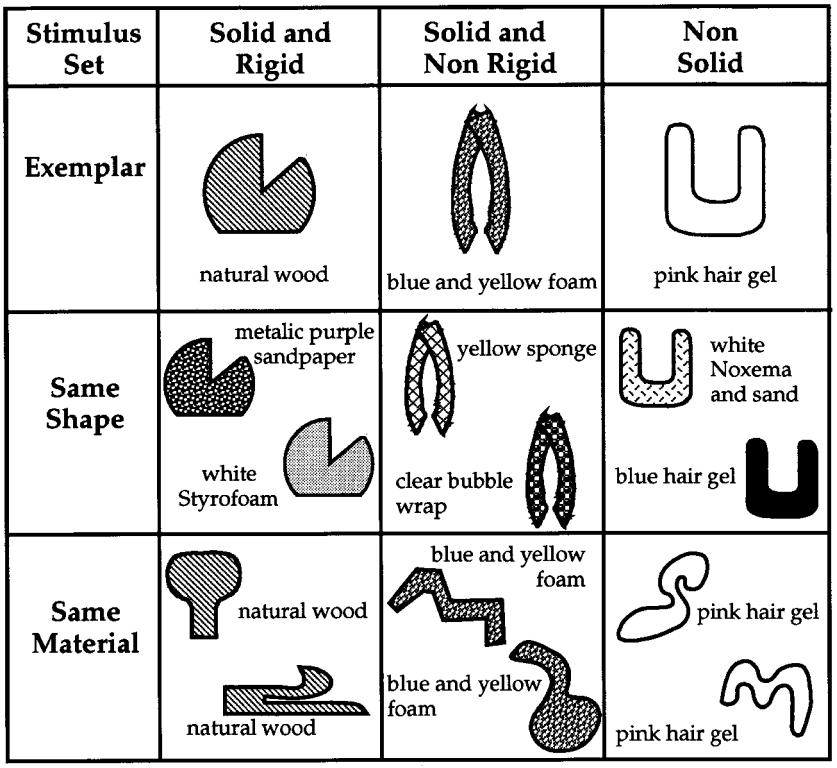

Fig. 6. Stimuli used in the novel word generalization task of Experiment 2.

the exemplar in material but not shape. The exemplar for the Solid and Rigid set was a 6.4 cm diameter circle made of 1 cm thick wood with a flattened bottom and a wedge cut out. One of the same-shape test objects for this set had a sandpaper surface painted metallic purple. The other same-shape test object was made of white Styrofoam. For most of the children the same-material objects for this set were a 6.0×4.5 cm wooden knob and a 9.0×2.0 cm irregularly-shaped piece of wood. However, for several children the same-material test objects were a thin, flat square piece of wood glued to a long thin rectangular piece of wood and another irregularly-shaped piece of wood. These items were broken by a child early in the experiment and were replaced.

The exemplar for the Solid and Non Rigid set was a 13.3×5.0 cm upside-down 'V' shape made of variegated blue and yellow foam carpet padding. For most of the children, the same-shape test objects for this set were made of yellow sponge and clear bubble wrap. However, for several of the first children tested, one of the same-shape test objects was made of two pieces of silver worm-like plastic glued together. The same-material test objects for all the children were a 13.0×7.6 cm zigzag shape and a 12.0'6.4 cm roughly musical-note shape.

The exemplar for the Non Solid set was a 7.6'6.2 cm 'U' shape made of pink hair

gel with pink glitter. The same-shape test objects in this set were made of Noxema mixed with sand, and blue hair gel. The same material test objects for this set were a 10.8×5.0 cm elongated 'S' shape and a 10.8×7.6 cm flattened 'M' shape. All stimuli in this set were presented on 15 cm white paper plates and were remade for experimental sessions as needed.

3.4.3. Procedure

After parents read and signed an informed consent statement they were asked to mark the words their child was currently producing on a list of nouns taken from the toddler form of the Mac Arthur Communicative Development Inventory. The list contained only nouns from the noun sections of the MCDI studied in Experiment 1. When this was completed the parent, child and experimenter moved to a quiet testing room in the laboratory for the novel word generalization task.

The task began with a series of training trials to familiarize the child with the procedure. The stimuli for these trials were all familiar objects such as two purple plastic eggs, one of which served as the exemplar, a white plastic flower, a small brown basket, a white doll hat, and a green triangle block. Each trial began with the experimenter giving the child the two plastic eggs and one contrast item (e.g. the flower) to play with for about 15 s. The contrast item differed from the exemplar in color, shape, and material. The experimenter then retrieved the three objects from the child, placed one of the eggs and the contrast item on a tray, held up the exemplar egg and said, 'See this? This is my egg.' The experimenter then pushed the tray towards the child while saying 'Show me your egg.' She then waited for the child to respond. If the child did not respond the experimenter repeated the question. If the child picked the egg he or she was enthusiastically praised. If the child did not pick the egg the experimenter said 'That's not your egg, This is your egg,' while pointing to or picking up the egg on the tray. This procedure was then repeated with the two eggs and other test items until the child correctly responded to three training trials in a row.

The experimental trials were exactly like the training trials. On each trial the child was allowed to explore the exemplar object, one same-shape test object and one same-material test object for about 15 s. The experimenter then placed the two test objects on the tray, held up the exemplar and said, 'This is my __ (lom, rif, zup). Can you show me your __ (lom, rif, zup),' and pushed the tray forward. The child's response was marked on a response sheet, but was not praised or corrected in any way.

Each of the two same-shape test objects in a stimulus set were presented with each of the two same-material test objects once for a total of four trials per stimulus set. The relative positions of test objects on the tray was randomly determined for each trial. After a child completed all four trials for a given stimulus set the experimenter moved on to the next stimulus set. The order of experimental trials within each stimulus set as well as the order of stimulus sets was counterbalanced across children. The nonsense label used with each stimulus set was also counterbalanced across children.

Throughout the experiment children sat next to their parent or caregiver. Parents

were asked not to label any of the experimental items but to encourage the child to explore the objects and answer the experimenter's questions. Parents also helped the experimenter clean the child's hands if he or she got hair gel or Noxema on them. The experimental sessions were videotaped and the recorded responses of each participant were rechecked for accuracy. The test object that the child touched or picked up first was always taken as the response.

3.5. Results

We present data on vocabulary growth first and performance in the novel noun generalization task, as a function of number and kind of nouns known, second. For some of the analyses under both topics we divide children into four groups based on their total productive noun vocabulary score; under 50, 51–155, 156–250 and more than 250 nouns. These divisions were based on natural breaks in the data and yielded groups with at least 10 children at each level of vocabulary development. Table 3 presents the mean age and range for each of the four groups as well as the means and ranges of total noun vocabulary and noun vocabularies for each syntax, solidity, and category organization. Lexical classifications in this Experiment are based on the adult judgments of Experiment 1 using the 85% agreement criterion. As can be seen in the table, there is a great deal of overlap in the range of ages among children in the four groups. As has been noted by others, children acquire language at variable rates and age is, for this reason, an imperfect index of language development. Also clear is the fact that the age range selected spans a period of dramatic vocabulary growth: noun vocabulary increases nearly 20-fold from the means of the lowest to highest vocabulary groups. Note also that for each vocabulary group count nouns, names for solid things and names for shape-based categories dominate.

3.5.1. Noun vocabulary trajectories

Do children learn nouns that fall in the intersections of solid + shape and non-solid + material first? They should if these are the a priori better examples of objects and substances. The present data provide no support for this idea. Fig. 7 gives the proportion of total nouns that are solid + shape nouns and non-solid + material nouns as a function of number of nouns. Other than declining variability between children as vocabulary increases, there is little change. The proportions of all nouns that are names for solid things in shape-based categories is steady at about 0.40 and names for non-solid substances in material-based categories is steady at about 0.03, at all levels of vocabulary. The lack of developmental change in the proportion of solid + shape nouns was confirmed by an analysis of variance comparing the proportions of solid + shape nouns for the four vocabulary-size groups, $F(3, 54) = 1.41$, $P = 0.25$. A similar ANOVA comparing the proportion of non-solid + material nouns for four vocabulary-size groups revealed a significant difference, $F(3, 54) = 3.00$, $P = 0.03$. However, given the small number of non-solid + material nouns on the MCDI (8 total) the increase in mean number of non-solid + material nouns known is an increase from less than one in the lowest vocabulary group to seven in the highest vocabulary group. Without direct evidence on the input

Table 3
Mean age, total noun vocabulary, and noun vocabularies for each syntax, solidity and perceptual properties classifications for the four vocabulary groups[a]

Group	Age	Total noun vocabulary	Syntax		Solidity		Perceptual properties	
			Count	Mass	Solid	Non-solid	Shape	Material
Below 50	20 months 4 days	18.08	14.75	1.17	12.33	0.83	10.83	1.33
($n = 12$)	(17,26–25,23)	(0–49)	(0–37)	(0–4)	(0–32)	(0–4)	(0–31)	(0–2)
51–155	21 months 25 days	101.90	77.3	10.00	66.30	4.70	54.20	13.80
(n = 10)	(17,16–25,11)	(55–152)	(32–117)	(5–17)	(32–96)	(1–7)	(20–84)	(6 -26)
156–250	27 months 13 days	213.38	158.77	20.69	135.23	9.08	103.46	34.31
($n = 13$)	(21,15–32,29)	(174–248)	(129–178)	(14–26)	(112–153)	(8 –11)	(85 –118)	(22–46)
250 +	29 months 3 days	281.0	210.65	27.74	179.61	12.52	133.00	45.57
($n = 23$)	(25,6–33,18)	(251–309)	(170–229)	(22–31)	(154–195)	(10–14)	(114–150)	(35–0.51)

[a] The ranges of each variable are given in parentheses.

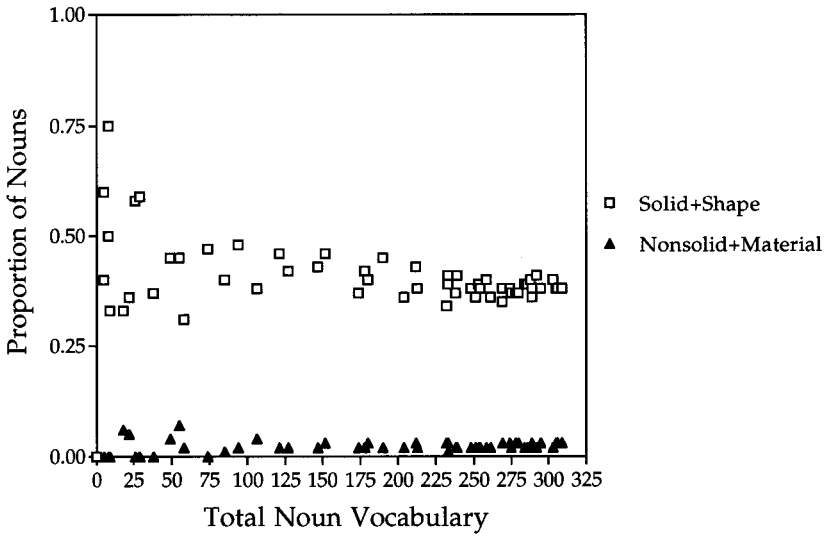

Fig. 7. Proportion of all nouns that refer to solid things in shape-based categories or non-solid substances in material-based categories in the total noun vocabularies of children in Experiment 2.

to children, these results are not unambiguous evidence against pre-linguistic biases; however, they also are not evidence *for* this position. Indeed, the steady proportion of these two kinds of nouns across vocabulary development suggests that their relative frequency in children's vocabularies might reflect the proportions of these kinds of nouns in the input.

We also asked whether children acquire nouns that fall in the intersection of solid + count and non-solid + mass first? They might be expected to if an ontological distinction between solids and non-solids exists pre-linguistically, is highly salient, and gets readily linked to the syntactic distinction. Fig. 8 presents the proportion of all nouns that are solid + count nouns and the proportion of all nouns that are non-solid + mass as a function of the number of nouns in each child's productive vocabulary. Once again there is early variability between children. As noun vocabulary progresses the proportion of nouns that are count nouns naming solid things settles to about 0.54, which is the proportion of these kinds of nouns on the Mac Arthur (167 solid + count out of 312 count nouns). After a brief period of variability, the proportion of all nouns that are mass nouns naming non-solid substances settles to about 0.04 of all nouns. This is the same as the proportion of these nouns on Mac Arthur (11 non-solid + mass nouns out of 312 total nouns). The lack of developmental change in the proportion of all nouns that are solid + count nouns was confirmed by an analysis of variance comparing the proportions of solid + count nouns for the four vocabulary-size groups, $F(3, 54) = 0.124$, $P = 0.95$. A similar ANOVA comparing the proportion of all nouns that are non-solid + mass nouns for four vocabulary-size groups also revealed a no significant difference, $F(3, 54) = 0.219$, $P = 0.88$.

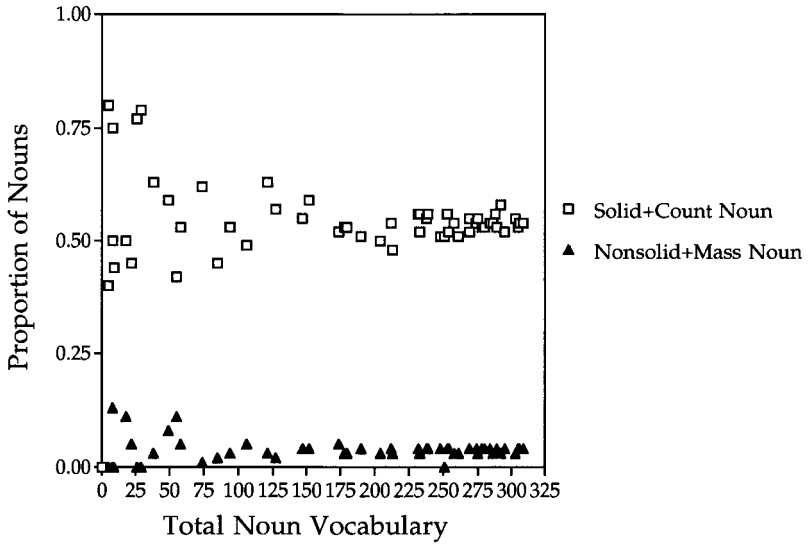

Fig. 8. Proportion of all nouns that refer to solid things and take count noun syntax or to non-solid substances and take mass noun syntax in the total noun vocabularies of children in Experiment 2.

Overall then, the trajectories of noun acquisition indicate that the order and rate of acquisition of the various kinds of nouns is best predicted by their proportion on the vocabulary checklist. Both when children know few nouns and when they know many, the composition of their vocabulary is remarkably stable. This conclusion is supported by the linear relationship between the number of known nouns in each kind of noun vocabulary and the number of total known nouns, as shown in Fig. 9. As can be seen in the figure, there is a strong linear relation between each kind of noun measured and the total nouns in the vocabulary. As children learn more nouns, they learn more of all the kinds of nouns on the checklist. Exactly how this fact should be interpreted is unclear. The considerable research on the validity and reliability of the Mac Arthur checklist suggests that it probably measures fairly the nouns children know (Fenson et al., 1994), but it may not measure fairly the nouns children *could* know. That is, there are two possible relations between the input children receive and the words they know. One possibility is that children learn the words they hear – that the Mac Arthur reflects not just the statistical composition of the nouns learned, but also of the nouns heard. If this is so, then count nouns, names for solid things and shape-based categories must dominate in speech to children.

The second possibility is that the 312 nouns on the MCDI reflect the biases *in the learner*, not the input – that although children hear many different kinds of noun categories organized in many different ways, they are biased to learn names for categories of solid objects that are organized by shape and take count noun syntax. A recent study within our laboratory argues against this view. Sandhofer and Smith (1998) found that the 50 most frequent nouns produced by parents in naturalistic

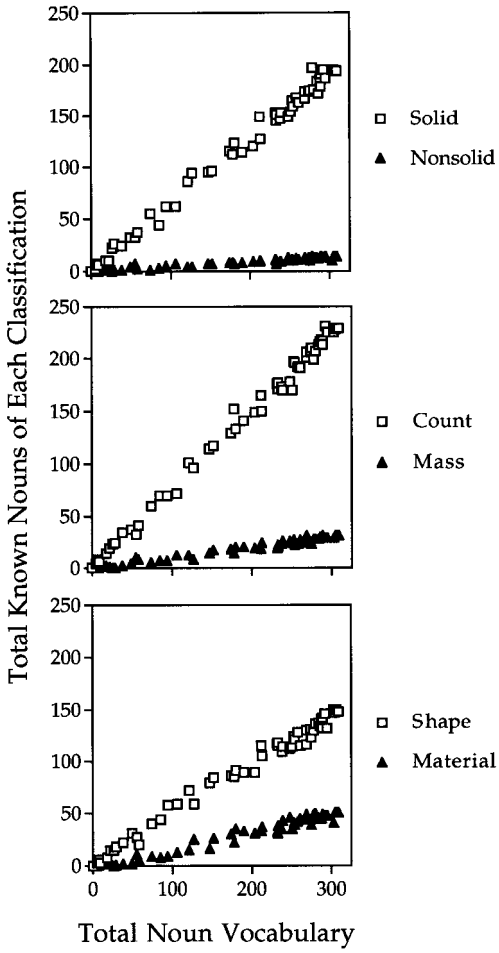

Fig. 9. Total number of names for categories of solid things or non-solid things (top panel), mass or count nouns (middle panel), and names for categories organized by shape or material (bottom panel) by the total noun vocabularies of children in Experiment 2.

play sessions to young children are predominantly count nouns naming solid things in shape-based categories. Using a procedure similar to that reported in this paper, Sandhofer and Smith had adults judge the solidity, basis of category organization and syntax of the 43 of the 50 more frequent nouns in mothers' speech to children (the remaining seven nouns were proper names). They found that, 74% named categories of solid objects, 58% named categories organized by similarities in shape, and 84% were count nouns.

3.5.2. Novel word generalization

The central empirical questions are: (1) do children generalize names for solid,

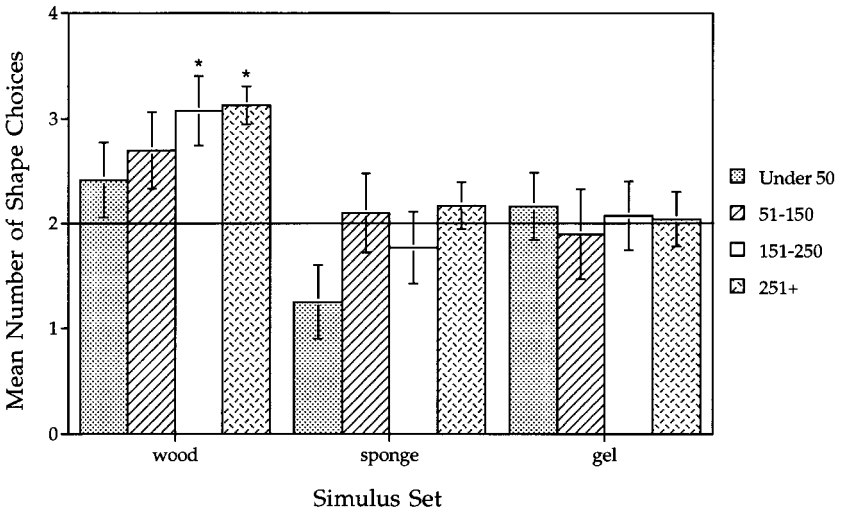

Fig. 10. Mean number of shape choices for each stimulus set and each vocabulary level group in the novel word generalization task of Experiment 2. Error bars represent standard errors. * = $P < 0.01$.

non rigid, and non-solid items differently; (2) is this stimulus effect apparent even when children know few nouns or does it emerge only after they have learned some number of nouns; and (3) do children generalize names for solid things by shape before they generalize names for non-solid things by material? Fig. 10 shows the results. Across age groups, children generalized names by shape more often with solid stimuli than with non rigid or non-solid stimuli, $F(2, 108) = 12.469$, $P = 0.0001$. Analyses of performance at each individual vocabulary group level revealed that only the two groups of children with the most nouns in their vocabulary reliably showed this stimulus effect, $F(2, 24) = 7.349$, $P = 0.0032$ and $F(2, 44) = 6.079$, $P = 0.0029$. That is, at the beginning of noun vocabulary development children did not generalize names for solid, non-rigid, and non-solid items differently. Further, as is evident in the figure, the only consistent pattern of generalization was the name of a solid thing to a new instance by shape, a pattern that became more robust and reliable with development. No systematic bias to generalize names for non-solid substances by material was observed at any level of vocabulary development.

However, it should be noted that there is a hint of a stimulus effect for the children with the fewest number of nouns in their productive vocabulary. There was a marked tendency for the children with the most limited vocabulary to *not* generalize the novel word to the same shape test object when the exemplar was nonrigid; indeed, the stimulus effect at the youngest age level approached conventional levels of statistical significance, $F(2, 22) = 3.09$, $P < 0.07$. This tendency could reflect two possibilities: (1) it could reflect some idiosyncratic aspect of the nonrigid stimulus set that mattered only to the mostly younger children and caused them to reject the shape-matching test objects in the nonrigid case; or (2) it could suggest that there are

early stimulus driven biases that are overridden as children first learn nouns. Nonetheless, the results overall fit the pattern expected if, as a product of learning nouns, children's novel noun generalizations come to reflect the statistical regularities among the nouns they already know. After children have already learned a substantial number of nouns that name solid objects by shape, they know that solid things are named by shape and that non-solid things are not. But at this point they do not know many names for non-solid or nonrigid things and they do not know just what regularities matter for naming those kinds of things.

The above analyses suggest that as children learn nouns they learn the regularities among object properties and category organizations. However, children who know more nouns (and attend to shape when generalizing a name for solid things) are in general older than children who know fewer nouns (and are less likely to attend to shape when generalizing a name for solid things). Is it the number of nouns known by children or their time in the world (and learning about categories in the world) that drives the developmental change? This is a difficult question to answer. On the one hand, and as is evident in Table 3, the age ranges of the children at the four different levels of vocabulary development are broad and there is considerable overlap in ages of individual children at the four levels. On the other hand, age (in days) and number of nouns in productive vocabulary are positively correlated, $r = 0.66$. In order to gain more insight into the relative strength of the association between noun vocabulary and age as predictors of shape classifications, we conducted a stepwise linear regression predicting the number of shape choices on the wood stimulus set by either number of nouns in the productive vocabulary or age in days. The regression equation incorporated productive vocabulary first, and yielded a significant effect of noun vocabulary and shape choices, $F(1, 56) = 4.63$, $P < 0.04$. The addition of the age in the regression equation did not yield a significant increase in the predictive accuracy of the equation. Next, we conducted a linear regression attempting to predict shape choices on the wood set by age alone; by conventional standards, this regression did not yield a reliable relation between age and shape choices, $F(1, 56) = 3.88$, $P = 0.06$. Nouns already known thus appears a better predictor of how children generalize a newly encountered noun than age.

3.6. Discussion

Overall then, the data indicate that between the ages of 17 and 32 months children's productive noun vocabularies contain many more names for categories of solid things, count nouns and names for categories organized by shape. Further, the systematicity of the mapping among these kinds of nouns in early vocabularies is reflected in children's novel word generalizations. The fact that individual children's early noun vocabularies closely parallel the statistical structure of early noun categories, and fact that systematic noun generalization emerges as a function of vocabulary growth, both are consistent with the idea that noun learning promotes children's understanding of how nouns map to ontological kinds.

4. General discussion

The present results make three contributions. First they provide new evidence on the mapping of count/mass syntax to the distinction between objects and substances. Second, the results provide new evidence on the nature and origin of the ontological categories of object and substance. Third, this evidence suggests that children's novel noun interpretations may be generalizations from already learned noun-category mappings. We consider these in turn and then reconsider them as they relate to the issue of individuation.

4.1. Syntax versus semantics

The global regularities observed among the 312 nouns in this early corpus do roughly fit prior assumptions: count nouns do tend to label categories of solid objects and categories organized by shape, and mass nouns do tend to label categories of non-solid entities and categories organized by material substance. However, not all correlations are equally strong and even the best are imperfect. The correspondences between syntactic category and meaning found in this early corpus are decidedly not rule like. Rather, the regularities fit Kelly and Martin's (Kelly & Martin, 1994) more general claim about structure in the world: '[the] environment is not a homogenous soup, although it rarely provides 'sure things.' The basic fact of the matter is that the world is awash with stuff best described as 'tendencies', 'maybes', 'estimates' and 'generally speakings'' (p. 107). And, as Kelly and Martin note, human beings appear particularly constituted to use this probabilistic information.

Thus, even if imperfect, the correspondences among count/mass syntax, solidity, and category organization may matter in children's developing understanding of categories and kinds. Indeed, in a simulation of the development of the shape-bias, Gasser and Smith (reported in Smith, 1995) showed that a neural network would develop a bias to generalize novel count nouns used to name novel entities by shape even when only 50% of the count nouns used in training named objects in categories well organized by shape. Thus, the fact the regularities are imperfect does not contradict the proposal that syntax serves as a bootstrap to semantics nor the opposing proposal that attention to solidity helps children understand count/mass syntax. However, the nature of the specific regularities reported here favor (slightly) the semantic distinction, and not the syntactic one, as the starting point.

Quine (1960) conjectured that syntax might teach an object-substance distinction. If the objects children encounter and first think about as objects are typically solid things with bounded shapes, and the substances they encounter and think about as substances are non-solid continuous masses, then the present results provide weak support, at best, for this idea. Count nouns do tend to refer to solid things. However, according to our adult judgments, the mass nouns children learn early are used to refer to categories that range from liquids such as water, to things that are ambiguous with respect to solidity such as aggregates (i.e. grass), deformables (i.e. playdough) or superordinate categories (i.e. food), and even to solid entities such as wood.

Consequently, syntax provides a poor guide to a semantic distinction based on an opposition of solid and non-solid entities (see Pelletier, 1979b for a similar point).

Syntax maps onto a distinction between shape- and material-based categories better than it does onto solidity. Thus, if the distinction between count and mass nouns is construed as being principally about shape-based versus material-based category structure and not solidity, syntactic bootstrapping seems more plausible. Count nouns consistently name categories by shape. Somewhat more than a third of all mass noun categories are well organized by material. Further, almost no mass nouns refer to categories organized by shape. Thus, syntax does provide information about category organization.

However, although the statistical regularities in the early noun corpus provide weak support for Quine's position there appears to be stronger support for the alternative proposed by Soja et al. (1991). In contrast to Quine, Soja et al. suggested that children may instead learn the syntactic distinction between count and mass nouns via a pre-linguistic understanding of the ontological distinction between solids and non-solids. The data presented in Fig. 2 make it clear that attention to this ontological distinction would be helpful in learning count/mass syntax. Count nouns comprise 74% of the nouns in the corpus as a whole but over 85% of the names for solid things. Mass nouns comprise less than 10% of the nouns in the corpus but more than 80% of the names for non-solid things. Clearly, the solidity of the named stuff provides a good clue to the likely syntactic category (see Hall, 1996 for a related result). Solidity is also a good predictor of category organization; over half of the solid things are named by shape and almost half of the non-solid entities are named by material.

In brief, predictions from solidity to syntax and from solidity to category structure are stronger overall than are predictions from syntax to solidity or from syntax to category structure. Altogether then, the compositional structure of the early noun lexicon in English fits the idea that the ontological distinction between solid and non-solid things precedes the count/mass distinction. This conclusion is also supported by the evidence from Japanese-speaking children and adults. Japanese is a language without the count/mass distinction. Nonetheless, Japanese speakers honor the hypothesized ontological distinction, naming solid things by shape and non-solid ones by material (Imai & Gentner, 1997). Clearly, count/mass syntax is not necessary for an understanding of the differing category organizations of solid and non-solid things.

These conclusions about syntax–semantic relations must be tempered, however, by the lopsided nature of the early noun lexicon. Children could get pretty far in learning this early corpus if they only knew that count nouns name solid things by shape. Table 4 shows how 'correct' children would be (relative to the adult judgments of Experiment 1) if they assumed that all the nouns in this corpus were count nouns, all the objects solid, and all the categories shaped-based. Overall, performance would be pretty good. Performance would be better if children's assumptions about category organization included information about syntax – that count nouns name solid things in shape-based categories and mass nouns name non-solid things in material-based categories. But note again, the best overall performance would

Table 4
Percent correct predictions of solidity, syntax or category organization for nouns in the early noun corpus when predictions are based on: (1) the assumption that all nouns are count nouns naming solid objects in shape-based categories; or (2) on the syntax of novel word; or (3) solidity of named stimuli[a]

Base assumption	Predict	Overall correct (%)
1. All are solid, count and shape	Solidity	74
	Syntax	63
	Category organization	48
2. Syntax	Solidity	66
	Category organization	60
3. Solidity	Syntax	84
	Category organization	61

[a] Adult judgments from Experiment 1 were used to calculate the overall proportion 'correct'. Calculations for assumption 1 were over all nouns in the corpus, which puts this assumption at a disadvantage. Only nouns unambiguously count or mass by adult judgments (assumption 2) and nouns unambiguously solid or non-solid (assumption 1) were used in calculations.

result from assuming that solid things are in shape-based categories named by count nouns and non-solid things are in material-based categories named by mass nouns. In sum, a distinction between solid and non-solid kinds is the best starting point for learning these nouns. However knowledge about only one-half of this distinction, the solid half, would also work pretty well and is perhaps the entry-point to learning about how substances differ from objects.

4.2. Origins of the ontological distinction

All the data suggest the primacy of one side of the ontological distinction in early noun learning. There are many more names for solid objects, shape-based categories and count nouns in the corpus of early-learned nouns. Children's early noun vocabularies – from those as small as 18 nouns to those as large as 280 nouns – contain more names for solid objects, shape-based categories, and count nouns than they do names for non-solid things, material-based categories, and mass nouns. And, children systematically generalize a novel name for a solid thing to new instances but do not systematically generalize a novel name for a non-solid substance. This overall pattern of more robust knowledge about the category organization of solid as compared to non-solid things fits the patterns found by others in novel noun learning tasks (Dickinson, 1988; Soja et al., 1991; Soja, 1992; Imai & Gentner, 1997). The only deviation from this general pattern is the study by Soja et al. (1991). As have other researchers, they found more robust generalizations of novel names for solid things by shape than generalizations of novel names for non-solid things by material. However, they also found that 2-year-olds generalized names for non-solid things by material at above chance levels. We suspect that young children's somewhat better performance in Soja et al's study compared to the present study (or that of Imai &

Gentner) is due to Soja et al's use of additional cues highlighting material over shape (e.g. on many trials the test object matching in material had no shape in the usual sense but was multiple bits of irregularly shaped material (see Carey, 1993 for a further discussion). This fact notwithstanding, the extant data suggest that children may not apply a distinction between objects and substances as a *whole* to noun learning. Instead, children may first work out an understanding of how solid things are categorized and named. An understanding of how non-solid things are categorized and named may emerge subsequently out of the prior understanding of solid things.

This idea fits Gentner's natural-partitions hypothesis (Gentner, 1982; Gentner & Boroditsky, 1999). She has suggested that nouns dominate early vocabulary learning because early nouns name objects that are easily individuated by a perceptual system tuned to pick out objects that retain their shape over movement and other transformations. Thus, names for solid things with bounded shapes may be acquired earlier than names for non-solid substances with variable shapes for the same reasons that (by Gentner's hypothesis) object names are acquired earlier than relational terms, that is, because object names refer to perceptually stable wholes. By Gentner's hypothesis, then, the dominance of object names has a pre-lingual basis in perception.

Our results suggest a second possible explanation of the developmental origins of an object-substance distinction in novel word learning – that it may be a consequence of early noun learning. Perhaps children's ontological kinds are the product of learning the bundles of correlations that co-occur across noun categories. A concept of object (as least as young children understand it) may just be the probabilistic associations among solidity, invariant shape, and naming by shape (see Bloom, 1996 for a similar suggestion). Analogously, the concept of substance may just be the probabilistic connections among non-solidity, variable shape, and naming by material. Further, if children learn these associative bundles as they learn words, they would have more evidence from early-learned nouns for a concept of object than for one of substance. The subtler and more complex correlations that cohere to form a concept of substance may not be well represented in the early-learned nouns. Instead, a concept of substance may require vocabulary development beyond that of 300 nouns. This idea fits the conclusions of Gordon (1988) and Gathercole (1997) that the object-substance distinction is not fully worked out until many years later (see also Bloom, 1994).

Does the infant data contradict this idea that the object-substance distinction derives from noun learning? The infant data suggest that infants expect solid objects to be of bounded shape and to move in a connected manner in space (Gibson, Owsley, Walker & Megaw-Nyce, 1979; Gibson & Walker, 1984; Streri & Spelke, 1989; Baillargeon & DeVos, 1991; Xu & Carey, 1996; Needham et al., 1997). There are few studies suggesting that infants have such expectations about substances. Moreover, there are no data suggesting infants categorize solid objects by shape ignoring other properties and non-solid ones by material ignoring other properties. We propose that it is specifically this aspect of the object-substance distinction that noun learning may teach.

Note, however, that the natural-partitions hypothesis and the learned ontology hypothesis are not in direct opposition. The natural-partitions hypothesis may explain why children learn the names for (and why parents talk about) solid objects more than non-solid ones – because these are the objects more easily segregated by the perceptual system. The learned ontology hypothesis adds to this an explanation of how children learn the category organization characteristic of solid things and how they use that knowledge and its relation to count noun syntax to learn new object names.

4.3. Novel noun generalizations

The present results indicate that young children do not generalize novel names for solid things by shape until they already know many names for solid things in shape-based categories. The present results also indicate that young children do not systematically generalize novel names for non-solid substances by material, a fact that fits their limited knowledge of names for non-solid substances in material-based categories. The correspondences between children's novel noun generalizations and the kinds of nouns they already know suggests that children may become smart noun learners by learning the statistical regularities among the noun categories they already know and using this knowledge to learn new categories. That is, children's smart word learning may arise as a generalization over the structure of already learned noun categories.

If children can learn the correspondence between solidity, shape and count noun syntax, they should be able to learn and use other statistical patterns as well. For example, in the case of substances, non-solidity and shape variability may be predicted by surface reflectance and by particular aspects of shape (simple, irregular, rounded). In addition, for a given category the importance of material over shape may be predicted by contexts such as eating or the breaking of objects into parts. With enough experience, children could learn these patterns of correspondences as well. It may be when these higher order relations among properties and contexts and category organizations are learned, in the course of learning nouns, that children achieve a more adult-like understanding of substance and individuation.

Further, if children are statistical learners in this way, it seems reasonable to expect that some patterns of regularities across categories will be easier to learn than others. Certainly strong patterns of correspondences should be easier to learn than partial ones. Interestingly, partial patterns of correlations *added* to strong ones could slow learning of the pattern as a whole. This idea may explain Imai and Gentner's (1997) paradoxical finding. Not only did Imai and Gentner find that Japanese speakers adhered to an ontological distinction between solids and non-solids without the benefit of count/mass syntax, but Japanese speakers' novel noun generalizations actually honored such a distinction *more* than did English speakers' generalizations. Japanese children and adults robustly generalized names for solid things by shape and names for non-solid things by material. English speaking children and adults generalized names for solid things by shape but generalized names for non-solid things in a disorganized manner that did not differ from chance.

Could it be that the partial and asymmetric relations between mass syntax and non-solidity *hurt* English speakers' discovery of the correspondences between non-solidity and category structure?

4.4. Individuation and naming objects and substances

Soja et al. (1991) first contrasted children's generalizations of names for solid objects and non-solid substances because solid rigidly shaped things perceptually realize what seems to be the essence of the abstract idea of object or individual and because non-solid substances with transient shapes that break into bits perceptually realize the abstract idea of a formless and continuous substance. Starting with this idea, there are three possible accounts of the relation between children's understanding of how solid versus non-solid things are named and concepts of individuation. One possibility is that individuation and the concept of *object* are tightly related developmentally, that abstract ideas of an individuated unitary thing grow out of perceiving, acting on, and learning the names for physical objects (see Lakoff & Johnson, 1980). Within this account, our data suggest that objects have developmental priority over substances – young children gain an understanding of how categories of solid objects are organized and named before they understand how categories of non-solid substances are organized and named. And, our data are consistent with the idea that learning the statistical regularities among object properties, category organization, and naming are instrumental in this development.

A second possibility is that concepts of individuation and the ontological distinction between objects and substance are tightly linked to the perceptible properties of solid and non-solid things but that these concepts are well formed early in infancy and prior to word learning (Gordon, 1985; McPherson, 1991; Bloom, 1994; Hall, 1994; Bloom & Kelemen, 1995; Hall, 1996). Under this idea, our data suggest that children have to learn to apply this knowledge when forming new lexical categories, and that they learn the mapping from ontological kind to lexical category structure for objects first. The solid-shape-count mapping may be first precisely because of the regularity with which solidity, shape-based category organization and count syntax co-occur in early vocabularies. Alternatively, the solid-shape-count mapping may be first because of the fundamental nature of a concept of object as a bounded, coherent, three-dimensional physical thing (Spelke, 1990; Xu, 1997; Gentner & Boroditsky, 1999). Evidence that infants quantify over solid physical objects but not over continuous substances is consistent with this view (Huntley-Fenner & Carey, 1995; Xu, 1997). Either way, the evidence we reported here is specifically about the nouns children learn early and about children's extensions of novel nouns to new instances. Thus, the findings may not reflect children's understanding of objects or substances per se but rather their use of this understanding in learning nouns.

The third possibility is that concepts of individuation and the perceptible properties of concrete solid objects and non-solid substances are not tightly related, not even in early development (see Prasada, 1993). Lucy's study of quantification and

naming in Yucatec Mayan strongly suggest the possibility of such a differentiation (Lucy, 1992). Common nouns in Yucatec Mayan are substance terms; for example, candles are referred to by the word for 'wax' and pots by the word for 'clay'. In brief, common nouns in Yucatec Mayan are not shape based and are semantically unspecified as to essential quantification unit. Instead, quantification requires numerical classifiers that are shape based, (e.g. two candles may be designated as 'two long thin wax'). Notice that although nouns with concrete reference are substance based in Yucatec Mayan and shape based in English, quantification of concrete things is shape based in both. Thus, bounded shape and not solidity may be the perceptual property most closely linked to individuation. Our finding that in children's early vocabularies, count-mass syntax is more strongly linked to a shape-material distinction than a solidity-non-solidity distinction is consistent with this view.

4.5. Conclusions

Clearly, concepts of individuation, ontological kinds and the means by which these abstract ideas are grounded in perceptual experience and realized in early language learning are complex and not resolved by the present results. However, previous questions about the relation between count/mass syntax and a conceptual distinction between objects and substances have often assumed a link between the properties of solid objects and non-solid substances, category organization by shape or material, and count/mass syntax – at least among the common nouns that form the core of the early noun vocabulary. The present results show that the assumed correspondences exist, but imperfectly. Most critically, among the 312 early-learned nouns we studied, there are few mass nouns, names for categories of non-solid substances or names for categories organized by similarity in material. The new insights about the structure of the early noun vocabulary, in conjunction with the evidence on the relationship between the nouns individual children know and their generalizations of novel nouns, suggest a version of Gentner's natural-partitions hypothesis such that solid objects in shape-based categories have priority in development. Further, the present results offer the possibility that children's novel noun generalizations, in particular, are a reflection of what they have learned about the structure of nominal categories, a product of learning nouns.

Acknowledgements

The research presented in this paper was supported in part by NIH grant HD28675 to the second author. We wish to thank the parents and participants for their time and effort. We also thank Mike Young for his assistance in creating Fig. 5. Thanks also go to Kelly Mix and Susan Jones for insightful comments on an earlier draft of this paper and to John Spencer, Eliana Colunga, Elana Berken, Catherine Lynch, and James Denny for invaluable discussions of the data and ideas contained in the paper.

References

Baillargeon, R., & DeVos, J. (1991). Object permanence in young infants: Further evidence. *Child Development*, *62*, 1227–1246.

Bloom, P., & Kelemen, D. (1995). Syntactic cues in the acquisition of collective nouns. *Cognition*, *56*, 1–30.

Bloom, P. (1996). Controversies in language acquisition: word learning and the part of speech. In R. Gelman & T. K. F. Au, *Perceptual and Cognitive Development.*, (pp. 151–183). San Diego, CA: Academic Press.

Bloom, P. (1994). Syntax-semantics mappings as an explanation for some transitions in language development. In Y. Levy, *Other Children, Other Languages: Issues in the Theory of Language Acquisition*, (pp. 41–75). Hillsdale, NJ: LEA.

Carey, S. (1993). Speaking of objects, as such. In G. Harman, *Conceptions of the Human Mind. Essays in Honor of George A. Miller.*, (pp. 139–159). Hillsdale, NJ: LEA.

Dickinson, D. K. (1988). Learning names for material: factors constraining and limiting hypotheses about word meaning. *Cognitive Development*, *3*, 15–35.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D., & Pethick, S. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, *5*, 242.

Gathercole, V. C. M., & Min, H. (1997). Word meaning biases or language-specific effects?. *Evidence from English, Spanish and Korean. First Language*, *17*, 31–56.

Gathercole, V. C. M., Cramer, L., Somerville, S. C., & Haar, M. J. O. D. (1995). Ontological categories and function: acquisition of new names. *Cognitive Development*, *10*, 225–251.

Gathercole, V. C. (1983). *The mass-count distinction: children's uses of morphosyntactic vs. semantic approaches. Papers and Reports on Child Language Development*, *22*, 395–415.

Gathercole, V. C. (1985). 'He has too much hard questions': the acquisition of the linguistic mass-count distinction in much and many. *Journal of Child Language*, *12*, 395–415.

Gathercole, V. C. M. (1997). The linguistic mass/count distinction as an indicator or the referent categorization in monolingual and bilingual children. *Child Development*, *68* (5), 832–842.

Gentner, D., & Boroditsky, L. (1999). Individuation, relativity, and early word learning. In M. Bowerman & S. Levinson, *Language Acquisition and Conceptual Development.*, Cambridge, UK: Cambridge University Press.

Gentner, D. (1982). Why nouns are learned before verbs: linguistic relativity versus natural partitioning. In S.K. II, *Language Development, Vol. 2: Language, Thought and Culture*, (pp. 301–334). Hillsdale, NJ: Erlbaum.

Gibson, E. J., & Walker, A. S. (1984). Development of knowledge of visual-tactual affordances of substance. *Child Development*, *55*, 453–460.

Gibson, E. J., Owsley, C. J., Walker, A., & Megaw-Nyce, J. (1979). Development of the perception of invariants: substance and shape. *Perception*, *8*, 609–619.

Gibson, J.J., (1979). The ecological approach to visual perception. Erlbaum, Hillsdale, NJ.

Gordon, P. (1985). Evaluating the semantic categories hypothesis: the case of the count/mass distinction. *Cognition*, *20*, 209–242.

Gordon, P. (1988). Count/mass category acquisition: distributional distinctions in children's speech. *Journal of Child Language*, (15), 109–128.

Hall, D. G. (1994). How children learn common nouns and proper names: a review of the experimental evidence. In J. Macnamara & G. E. Reyes, *The Logical Foundations of Cognition.*, (pp. 212–240). New York: Oxford University Press.

Hall, D. G. (1996). Naming solids and non-solids: children's default construals. *Cognitive Development*, *11*, 229–264.

Huntley-Fenner, G.N., & Carey, S. (1995). Individuation of objects and non-solid substances: a pattern of success (objects) and failure (non-solid substances). Poster session presented at the Biennial Meeting of the Society for Research in Child Development, Indianapolis, IN.

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

Imai, M., Gentner, D., & Uchida, N. (1994). Children's theories of word meaning: the role of shape similarity in early acquisition. *Cognitive Development*, *9* (1), 45–75.

Jones, S. S., & Smith, L. B. (1998). How children name objects with shoes. *Cognitive Development 3*, 323–334.

Kelly, M. H., & Martin, S. (1994). Domain-general abilities applied to domain-specific tasks: sensitivity to probabilities in perception, cognition, and language. *Lingua*, *92*, 105–140.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by.*, Chicago, IL: University of Chicago Press.

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

Lucy, J. A. (1992). *Grammatical categories and cognition: a case study of the linguistic relativity hypothesis.*, Cambridge, MA: Cambridge University Press.

Malt, B. C. (1994). Water is not H_2O. *Cognitive Psychology*, *27*, 41–70.

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

Markman, E.M., (1992). Constraints on word learning: speculations about their nature, origins, and domain specificity. In: Gunnar, M.R., Maratsos, M., Modularity and Constraints in Language and Cognition, Volume 25: The Minnesota Symposia on Child Psychology. LEA, Hillsdale, NJ, pp. 59–-101.

McPherson, L. M. P. (1991). A little goes a long way: evidence for a perceptual basis of learning for the noun categories count and mass. *Journal of Child Language*, *18*, 315–338.

Needham, A., Baillargeon, R., & Kaufman, L. (1997). Object segregation in infancy. *Advances in Infancy Research*, *11*, 1–44.

Pelletier, F.J., (1979). Mass Terms: Some Philosophical Problems. Reidel Publishing Company, Dordrecht, The Netherlands.

Pelletier, F.J., (1979). Non-singular reference: some preliminaries. In: Pelletier, F.J. Mass Terms: Some Philosophical Problems. Reidel Publishing Company, Dordrecht, The Netherlands, pp. 1–14.

Prasada, S. (1993). Learning names for solid substances: quantifying solid entities in terms of portions. *Cognitive Development*, *8*, 83–104.

Quine, W. V. O. (1960). *Word and Object.*, Cambridge, MA: MIT Press.

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

Sandhofer, C., Smith, L.B. (1998). A note on counting nouns and verbs in the input: differential frequencies, different kinds of learning. Submitted for publication.

Smith, L. B., Jones, S. S., & Landau, B. (1996). Naming in young children: a dumb attentional mechanism?. *Cognition*, *60*, 143–171.

Smith, L.B., 1995. Self-organizing processes in learning to learn words: development is not induction. In: Nelson, C.A., Basic and Applied Perspectives on Learning, Cognition, and Development, Volume 28: The Minnesota Symposia on Child Psychology. Erlbaum, Mahwah, NJ, pp. 1-32.

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

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

Spelke, E. S. (1990). Origins of visual knowledge. In D. N. Osherson, & S. M. Kosslyn, *Visual Cognition and Action: An Invitation to Cognitive Science, Vol. 2*, (pp. 99–127). Cambridge, MA: MIT Press.

Streri, A., & Spelke, E. S. (1989). Effects of motion and figural goodness on haptic object perception in infancy. *Child Development*, *60*, 1111–1125.

Waxman, S. R., & Hall, D. G. (1993). The development of a linkage between count nouns and object categories: evidence from fifteen- to twenty-one-month-old-infants. *Child Development*, *64*, 1224–1241.

Waxman, S. R., & Markow, D. B. (1995). Words as invitations to from categories: Evidence from 12- to 13- month-old infants. *Cognitive Psychology*, *29* (3), 257–302.

Xu, F., & Carey, S. (1996). Infant's metaphysis: the case of numerical identity. *Cognitive Psychology*, *30*, 111–153.

Xu, F. (1997). From Lot's wife to a pillar of salt: evidence that physical object is a sortal concept. *Mind and Language*, *12*, 365–392.