

Correlation versus prediction in children's word learning: Cross-linguistic evidence and simulations

ELIANA COLUNGA, LINDA B. SMITH, AND MICHAEL GASSER*

*University of Colorado at Boulder
Indiana University*

Abstract

The ontological distinction between discrete individuated objects and continuous substances, and the way this distinction is expressed in different languages has been a fertile area for examining the relation between language and thought. In this paper we combine simulations and a cross-linguistic word learning task as a way to gain insight into the nature of the learning mechanisms involved in word learning. First, we look at the effect of the different correlational structures on novel generalizations with two kinds of learning tasks implemented in neural networks—prediction and correlation. Second, we look at English- and Spanish-speaking 2-3-year-olds' novel noun generalizations, and find that count/mass syntax has a stronger effect on Spanish- than on English-speaking children's novel noun generalizations, consistent with the predicting networks. The results suggest that it is not just the correlational structure of different linguistic cues that will determine how they are learned, but the specific learning mechanism and task in which they are involved.

Keywords

word learning, crosslinguistic, mass/count syntax, neural networks, prediction/correlation

* (Correspondence addresses: Eliana Colunga, Department of Psychology and Neuroscience, University of Colorado, Boulder, Colorado, 80309-0345 USA; Linda B. Smith, Department of Psychological and Brain Sciences, Indiana University, Bloomington, Indiana 47405-7007, USA; Michael Gasser, Computer Science Department, Indiana University, Bloomington, Indiana 47405-7104, USA. E-mails: colunga@psych.colorado.edu; smith4@indiana.edu; gasser@indiana.edu.

1. Introduction

1.1. *Correlation versus prediction in children's word learning: Cross-linguistic evidence and simulations*

When we look at the world, we seem to understand it as made up of bounded and unbounded entities, an ontological distinction between discrete bounded things and continuous quantitites. The origin of these notions and their relation to quantitative terms in the world's languages have been widely studied (e.g. Quine 1969; Gordon 1985; Soja 1992; Smith et al. 2003). Different languages express this ontological distinction in different ways making this a fertile area for examining the relation between language and thought. Most past research has been concerned with whether a language's devices with respect to individuation do or do not affect a speaker's ideas about object and substance categories. As such, these studies examine performance differences and make inferences about underlying concepts. They do not consider the question of the learning mechanisms that might bring about those cross-linguistic effects. Here we use cross-linguistic differences as a natural experiment not about whether language affects thought but as a natural experiment for examining the nature of the learning mechanism.

In general, the mechanistic question concerns syntactic bootstrapping. This refers broadly to the idea that children use correlations between syntactic regularities and categories of meaning to determine the likely meaning of a newly encountered word (Gleitman 1990). The case of interest here concerns children's use of count-mass syntactic cues to make inferences about whether a noun refers to the shape or material of the named thing. Count nouns take the definite article, number, and the plural (a dog, two cups) and as such refer to bounded and shaped entities. Mass nouns are not pluralized and take continuous quantifiers (some sand, a lot of water) and as such refer to substances. Children learning English exploit the linguistic markers of count and mass status in learning new nouns. For example, in artificial word learning tasks, a count frame ("*a* *zup*") leads children to interpret the noun as referring to a category of similarly shaped things whereas a mass frame ("*some* *zup*") leads children to interpret the noun as referring to things of the same material (Soja 1992; Gathercole 1997). Children's use of count-mass frames in this way exploits a real regularity in the English language; common count nouns such as "*cup*" and "*ball*" systematically refer to shape-based categories whereas mass nouns such as "*milk*" and "*soap*" refer to material-based categories (Samuelson and Smith 1999). Thus, it seems possible that children learn the relation between syntactic cues and category structure as

they learn early nouns. The research reported here tests two contrasting hypotheses about the nature of the learner and the learning task. The hypotheses and methodological approach build on two facts: (1) cross-linguistic differences in the correlations of count/mass syntax with category organization and (2) two different kinds of learning tasks—one in which the learning mechanism gleans bi-directional associations and one in which the learning mechanism makes *directional* predictions. Cross-linguistic differences between English and Spanish provide a natural test of which of these two learning mechanisms underlies syntactic bootstrapping.

1.2. Two different patterns of correlations

Count nouns such as “*cup*” and “*study*” refer to entities conceptualized as discrete countable units. Mass nouns such as “*water*” and “*research*” refer to entities conceptualized as continuous masses. English and Spanish both make this count-mass distinction but in different ways. In English, with relatively few exceptions, nouns are classified as either count or mass. Thus, the language forces the speaker to view the referent of the noun as either an individuated object or a continuous substance. In contrast, in Spanish, nouns are not strictly classified as count or mass. In principle any noun can co-occur with both count and mass syntax, depending on how the speaker views the entity (Gathercole 1997). Comparisons of mass and count syntax in English and Spanish show that the distinction between mass and count is much less sharply defined in Spanish than in English. Spanish has fewer formal features distinguishing mass from count, there are more Spanish nouns than English nouns that can function as both mass and count nouns, and although in both English and Spanish a count noun can be changed into a mass noun and vice versa (e.g. *more car*, *a fear*), the practice is more frequent and productive in Spanish than in English (Iannucci 1952).

These differences are illustrated in Figure 1. An English-speaker talking about the wooden block in the picture could say “*a block*” if they construe it as an object or as “*some wood*” if they construe it as a substance. Furthermore, one cannot say “*some block*” or “*a wood*” because “*block*” is a count noun and “*wood*” is a mass noun. Thus, both the syntax used (*a* vs. *some*) and the noun itself (*block* vs. *wood*) indicate whether the speaker is referring to an object or a substance.

A Spanish-speaker, talking about the same wooden block, could say “*un bloque*” (a block) or “*algo de madera*” (some wood), but also “*una madera*” (a wood). Although “*bloque*” does refer to the object construal (block-shaped entity) and “*madera*” does refer to the substance construal

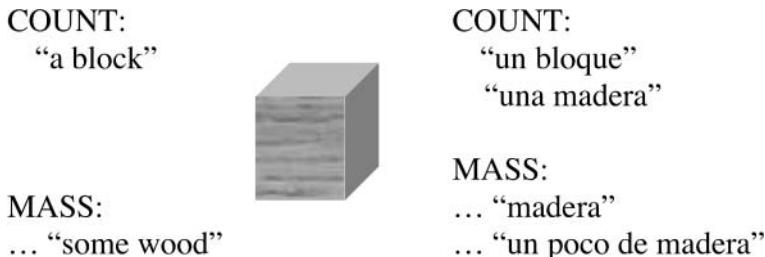


Figure 1. *A block of wood can be construed as an object (block) or as a substance (wood). In English all blocks are objects and the word block is always a count noun that refers to a block-shaped entity; all wood is a substance and the word wood is always a mass noun referring to wood-material. In Spanish, a wooden block could be called un bloque (a block) or algo de madera (some wood), but it could also be called una madera (a wood)*

(wood-stuff), *madera* can be used in both count and mass syntactic frames.¹ Thus, the noun *madera* cannot be said to be either a count noun or a mass noun, as it can occur with either count or mass syntax. Similarly, in English one would call one spider *a spider* and a big pile of spiders *many spiders*, but in Spanish, given enough spiders, one can talk about the pile of spiders as being *muchas arañas* (much spider). Note that although English has some nouns that work like *madera* in Spanish, (e.g. *a muffin* would imply muffin-shaped object; *some muffin* would refer to muffin-stuff), in Spanish, in principle, all nouns work this way, and in everyday speech many more nouns are used in both count and mass frames than in English (Gathercole 1997; Gathercole et al. 2000; Iannucci 1952).

These differences between English and Spanish mean that nouns, count-mass syntax, and category structure correlate differently in the two languages, as shown in Figure 2. In English, syntax and the noun are correlated and redundantly predict shape-based versus material-based categorization. In Spanish, syntax and noun are less well correlated and it is syntax, rather than the noun, that best predicts shape-based versus material-based categorization. Thus in both languages count/mass syntax is predictive of category organization, but the lexical category is more predictive in English than in Spanish. Does this difference between the way features correlate in the two languages have an effect on children's attention to count/mass cues as predictors of category structure? These differences in correlational structure between Spanish and English provide a

1. English can also use mass nouns in a count syntax to denote a kind. For example, “a wood” or “many woods” referring to wood kinds, like oak, maple, Mahogany. This use is also found in Spanish, and is in fact more frequent in Spanish (Iannucci 1952).

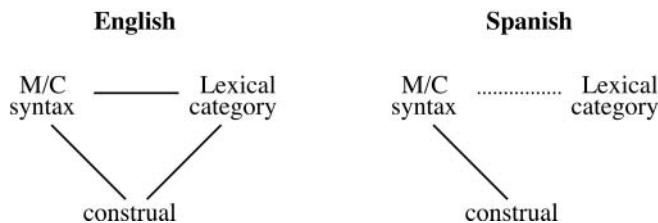


Figure 2. *An illustration of the differing correlations among nouns, count/mass syntax, and category structure in English and Spanish*

way to distinguish the two plausible but fundamentally different learning mechanisms.

1.3. Two ways of structuring the learning task

There are two ways to think about learning the relation between linguistic entities and the kinds of categories to which they refer. One is that the child treats determiners, nouns, and category structure equivalently and learns the *bi-directional* co-occurrences among them. Colunga and Smith (2005) showed that this kind of mechanism could learn and generalize the count-mass distinction in English. However, another possibility is that words and their referents do not have equal status but that instead learning is *directional*, from the word to the predicted referent (Regier 2005). This difference between learning bi-directional correlations and learning directional cues that predict outcomes is a fundamental one in learning theory that is often talked about in terms of unsupervised and supervised learning. In unsupervised tasks, the learner records co-occurrences between the elements, apprehending the correlational structure of the input, and shifting attention to elements that have been more predictive. In supervised tasks, the learner is given the task of predicting a target, given some information. Both of these learning setups have been used by models of category learning without much systematic comparison in the context of real-world learning (Landauer and Dumais 1997; Miikkulainen 1997; Rogers and McClelland 2004; Regier 2005; Li Zhao and MacWhinney 2007). Importantly, these two ways of structuring the task and thinking about the learning mechanism make different predictions about the relative potency of count-mass syntactic cues for learners of English and Spanish.

The key property of the first kind of learning, passive registration of co-occurrences, is that it forms *bidirectional* associations. These are known

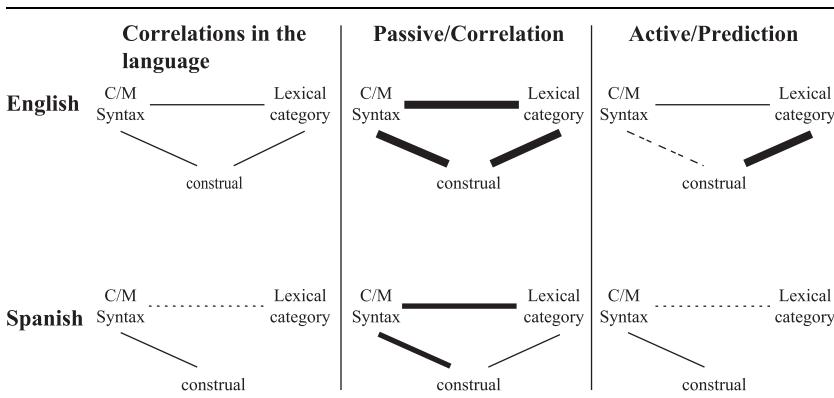
to lead—both in simulation studies and in experimental studies to self-reinforcing connections (Billman and Knutson 1996; Love et al. 2004; Yoshida and Smith 2003a, 2003b). A link between two associates is stronger in the context of a third redundant associate that is connected to each of them. By this learning mechanism, count-mass syntactic cues should be more potent in English than in Spanish, because the association between syntactic cue and category structure is redundant to an association between noun and syntactic cue and between noun and category structure—a redundancy that is not so strongly present in Spanish.

The key property of the second kind of learning, prediction, is that connections are not bidirectional. Instead there are cues that predict and outcomes that are predicted. Experiments and simulations of this kind of learning suggest that learners attend preferentially to the most predictive cues, perhaps even to the extreme of ignoring redundant cues that are less predictive (Rescorla and Wagner 1972; Kruschke 2001; Kruschke and Blair 2000; Bott et al. 2007). This possibility suggests that the redundancy in the correlations in English may not benefit the learner. Indeed, if the correlation between count-mass syntax and category structure is weaker in English than is the correlation between the noun itself and category structure, one might even predict that the redundancy in English would cause learners of English to pay less attention to count-mass syntax than learners of Spanish, because in Spanish the most predictive cue of intended meaning is the syntactic frame and not the noun.

1.4. *Rationale*

Table 1 summarizes the two different patterns of attention to count/mass syntax cues in English versus Spanish as they might be predicted by the two different characterizations of the learner—bidirectional associations versus prediction. If children go about learning connections between words and categories in a bi-directional associative way, one might expect the redundant correlations of mass and count syntax, lexical category, and category structure in English to result in better learning for each of these cues in English than in Spanish. However, if children in a way that is more similar to a prediction, from words to categories, the less redundant mass/count syntax in Spanish should cause count/mass syntax to receive more attention in Spanish than in English. We test this prediction in two steps. First we verify our analysis of the different effects of redundancy on a bi-directional learner of correlations versus a directional learner in which cues predict categories. We present networks with idealized versions of the English and Spanish correlational structures and train them using the two different training regimes—the passive storage of cor-

Table 1. *The different correlational structures of the two languages yield different predictions for the relative attention to Count/Mass syntax in the two languages depending on the nature of the learner. A passive/correlational learner predicts that count/mass syntax will be more strongly attended by English speakers than by Spanish speakers; an active/predictive learner predicts the opposite, that count/mass syntax will be given more attention by Spanish speakers than by English speakers.*



relations among linguistic forms and referents and the prediction of the referent from the linguistic forms. Second, in a behavioral experiment, we examine young English- and Spanish-speaking children's sensitivity to count-mass cues in an artificial noun learning task. Are the effects of these cues greater in English or in Spanish?

2. Experiment 1

The goal of this simulation is to demonstrate that the implications of the different correlational structures presented by the count/mass systems of English and Spanish depend on the nature of the learning system that is assumed. By our analysis, very different cross-linguistic differences should be obtained if one assumes a system that learns bi-directional associations among all three cues (the syntactic frame, the specific noun, the category structure) versus if one assumes a system that attempts to predict category structure from the two linguistic cues (syntactic frame and specific noun). We demonstrate the validity of our analysis by presenting artificial learning systems—one designed to learn bi-directional correlations and the other designed to predict the referent from the linguistic input—with a set of training stimuli that mimic either the structure of the English count/mass system or the structure of the Spanish count/mass system.

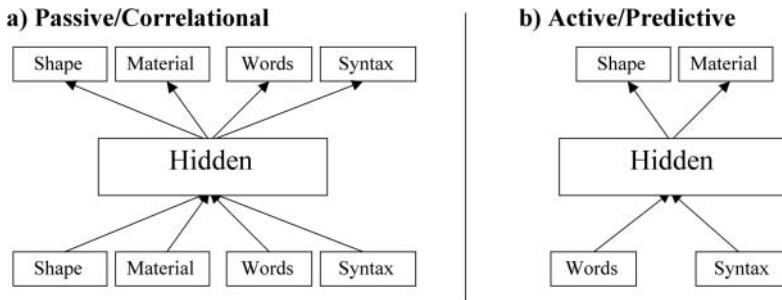


Figure 3. *Network architecture for the correlational and predictive networks. Each block indicates a set of nodes*

2.1. Method

2.1.1. Architecture The predictive networks were designed to model a comprehension task, given a noun and count/mass syntax, the network's task was to predict the shape and material of the object referent. The architecture is shown in Figure 3a. The connectivity is feed-forward; each unit in the lower level is connected, feeds activation to, each unit in the next level. The input layers consist of a Word layer which contains 12 units, one for each word to be taught to the network, and a Syntax layer, consisting of two units, one for count syntax and one for mass syntax. The output layers consist of banks of units in which the material and a shape of the object were represented using distributed representations. The Material and Shape layers consisted of eight units each. There was also a Hidden layer through which input and output layers were connected.

The correlational networks modeled the task of associating the elements of form and meaning with one another. Unlike the predictive networks, these had no directionality to the associations learned. This algorithm was implemented as auto-association, that is, the network was trained to copy a pattern appearing on its input to its output via a hidden layer.

As shown in Figure 3a, the correlational networks had the same layers as the predictive networks, but all of the input and output layers—Word, Syntax, Shape, and Material—appeared as layers in both the input and the output of the auto-associative network. These layers had the same number of units as in the supervised network, and they were joined through a hidden layer.

2.1.2. Simulating English and Spanish Networks of both types were trained on either a set of twelve “English” patterns or a set of twelve

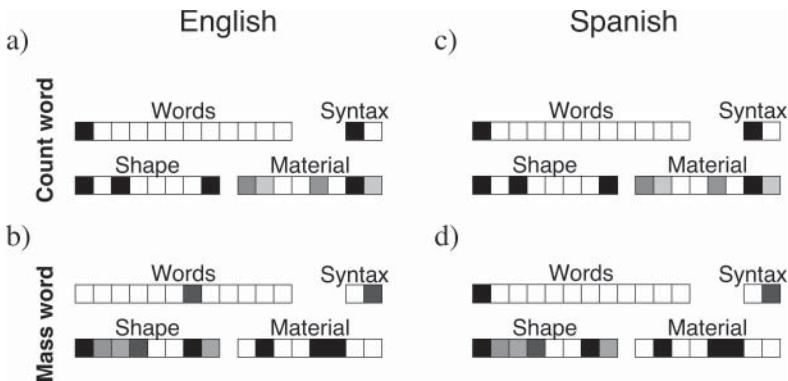


Figure 4. These figures show examples of the activation patterns on the four layers (shape, material, words and syntax) used on the training trials for (a) English count nouns (b) English mass nouns (c) a Spanish noun in the context of count syntax and (d) the same Spanish noun in the context of mass syntax. See text for further clarification

“Spanish” patterns Figures 4a and b illustrate two sample English patterns—one for a thing labeled with a count noun and one for a thing labeled with a mass noun. Here we show the individual units in each layer along with its activation level (black indicates high activation, white low activation). In this way we show a possible pattern of activation on each layer in a training trial. For English, each training word always appeared with the same syntactic pattern, either count or mass. Each count noun also had a randomly generated—but for that word, fixed—Shape pattern associated with it. On each training trial, each count word with its fixed shaped pattern was presented with a variable material pattern generated randomly on each trial. Likewise, each mass word had a fixed Material pattern associated with it and a Shape pattern that varied randomly across training trials. Thus for the “English” patterns, the word specified either a particular shape pattern or a particular material pattern, and the syntax cues redundantly predicted whether it was the shape or the material pattern that was relevant. For example, a word like block would be associated with the count pattern in the syntax layer and with a specific pattern of activation in the Shape layer (block-shape), but could go with any pattern in the Material layer (e.g. wood, rubber, concrete). In contrast, a word like wood would be associated with the mass pattern in the syntax layer and a specific Material pattern (wood-stuff) and could co-occur with any Shape pattern (e.g. table-shape, ball-shape, column-shape).

The Spanish patterns are illustrated in Figures 4c and d. Here each input word could occur with either of the two syntactic patterns, and each word was assigned both a fixed Shape pattern and a fixed Material pattern. However, the associated syntactic cue determined on each trial whether the associated fixed Shape *or* Material pattern was presented together with the word; the other layer was assigned a random pattern generated during training. Given a count pattern on the syntax layer, the word was paired with the fixed shape pattern (and a randomly generated and variable material pattern); given a mass pattern on the syntax layer, the word was paired with the fixed material pattern (and a randomly generated and variable shape pattern). That is, all words were like “muffin”, associated to both a shape (muffin-shape) and a material (muffin-stuff). If the Count unit was on, the muffin-shape pattern was represented in the Shape layer and a randomly generated material in the material layer (simulating, for example, a toy muffin made out of rubber); if the Mass unit was on, the muffin-stuff pattern was activated on the Material layer and a randomly generated shape was represented on the Shape layer (simulating, for example, a torn piece of muffin).

2.1.3. Training Training for both the predictive and the correlational networks used the back-propagation learning algorithm (Rumelhart et al. 1986). For the predictive networks, during each training event a word and a syntactic pattern were presented to the input layers, and activation was passed through the network, yielding a pattern of activation on the output layers. This pattern was compared to the target output pattern associated with the input word and syntactic pattern. The error derived from that comparison was propagated back through the network, and the connection weights were adjusted accordingly (see Appendix). For the correlational networks, the target pattern was identical to the input pattern. Again, back-propagation was used to adjust the weights. For each training event, a word, syntactic pattern, shape, and material pattern were presented to the input layers, and activation was passed through the network, yielding a pattern of activation on the output layers. This pattern was compared to the target output pattern, which was identical to the input pattern, the error was propagated back through the network, and the connection weights were adjusted accordingly.

2.1.4. Testing The key question concerns what the networks have learned about syntactic cues. Do count cues—even given novel things and a novel noun—push attention to shape? Do mass cues—even given novel things and a novel noun—push attention to material? To address this question, the networks were presented, after training, with a novel

word in two syntactic contexts, one specifying a count noun and one specifying a mass noun².

The resulting patterns of internal representations on the hidden layer were then examined for the same input pattern in each syntactic context. If a network is relying relatively heavily on syntax, then the very same word should lead to different patterns of hidden layer activation in the two different syntactic contexts. Thus our dependent measure, Syntax Effect, was the Euclidean distance between the internal representations for the same input pattern when the count versus the mass syntax unit was turned on. If the syntactic context matters more for one training set ("Spanish" or "English") than for the other, then the Syntax Effect, the difference between the internal representations given the two syntactic contexts, should be greater for that training set than for the other. We trained and tested 20 predictive networks and 20 correlational networks, 10 on the "English" and 10 on the "Spanish" patterns for each type of network. Each had randomly generated starting weights. During training each network was taught 12 words. For the English set, half were count nouns and half were mass nouns; for the Spanish set, each word was associated with count syntax (and shape) on half the training trials and with mass syntax (and material) on the other half. Networks were trained for 100 repetitions of the 24 training patterns (two for each input word), and tested on 12 novel testing patterns, each presented in both a count and mass syntactic context.

2.1.5. Results and discussion The measure of the magnitude of the Syntax Effect for the networks is the Euclidean distance between representations of the same novel words given count versus mass syntax. Figure 5 shows this measure for the predictive and the correlational networks trained on the "English" and "Spanish" training sets. As is evident, whether the correlations in the "English" versus the "Spanish" sets lead to greater or weaker Syntax Effects depends on the network.

The distance measures (syntax effect) predicted by the network were submitted to an analysis of variance for a $2(\text{Language}) \times 2(\text{LearningTask})$ design. The analysis yielded main effects of Language $F(1,16) = 75, p < .001$ and LearningTask $F(1,16) = 200, p < .001$, and also a reliable interaction between Language and LearningTask $F(1,16) = 138, p < .001$.

2. Presenting the network with a novel word was accomplished by randomizing the weights between the Hidden Layer and the Word Layer after training and before testing. Because these are feedforward networks, and because words were represented using a localist representation, this is equivalent to having 12 extra Word units that were never trained on during training.

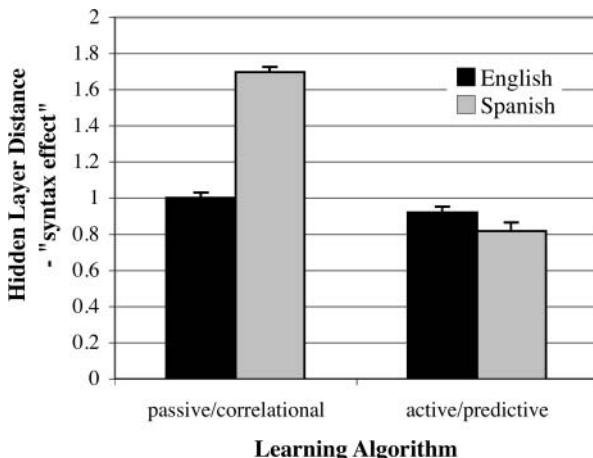


Figure 5. The distance of the hidden layer representations for the same shape and material input patterns in the context of count versus mass syntax. This measure of the magnitude of the effect of the syntax cues on hidden layer activation is shown for Supervised and Unsupervised networks given training on the “English” and “Spanish” correlations

Post-hoc analyses were conducted within each training task type comparing the syntax effect in the “Spanish”- and “English”-trained conditions (Tukeys, alpha = .05). Both comparisons were reliable. That is, for the networks trained using the predictive task, the syntax effect is much greater for the “Spanish” than the “English” networks. For the networks trained with the correlational task, the differences are smaller but still reliable, and opposite to those for the predictive network; that is, given a correlational, bi-directional learner, the syntax effect is greater for the “English” than the “Spanish” networks.

These results tell us that the correlations that characterize the English count/mass system and those that characterize the Spanish count/mass system may well have consequences for just how readily children use the information in the syntactic frame to constrain the possible meaning of a novel noun. But just what those consequences are depends on the learning mechanism. *Which set of correlations—English-like or Spanish-like—leads to the greater potency of the syntactic frame depends on the nature of the learning system that is assumed.* If one assumes a system that learns bi-directional associations among syntactic frames, individual nouns, and shape- or material-based categories, then English with its greater redundancy among these various cues, leads to the greater potency of the syntactic frame. If, however, one assumes a

mechanism that partitions words and referents into cues and outcomes, then the correlations in Spanish, rather than those in English, should lead to the greater potency of the syntactic frame. This predicted pattern occurs because in Spanish, the syntactic frame is the better predictor of category structure because many individual nouns can refer to either object or substance categories whereas in English the specific noun is the better predictor because individual nouns refer (for the most part) to either object or substance categories and thus are the most reliable indicator (over specific linguistic frames that may or may not co-occur with the noun).

Thus the simulations affirm our analysis that the different correlations presented by the count-mass system in English versus that in Spanish. In this context, it is important to note that this is the point of the simulations; they are not offered as contender accounts of just how children learn count or mass syntax in either language. The "English" and "Spanish" training patterns used in these simulations are extreme versions of the more nuanced differences between English and Spanish. We used these more extreme versions of the cross-linguistic differences in the two patterns of correlations in order to test our analysis about how these patterns of correlations would interact with the two different kinds of learning systems. Further, we implemented the two learning systems in comparable networks to examine our analysis of the different cross-linguistic effects as a function of the learning system. The results show that just how these two cross-linguistic differences matter depends on the kind of learner.

In sum, if children simply register co-occurrences among syntactic frames, nouns, and category structure then we should expect that English-speaking children relative to Spanish speaking children should show a stronger (earlier) effect of syntactic frames on their extension of a novel noun because the overlapping correlations among these frames, nouns, and category structure should reinforce each other. If, however, children *predict* expected category structure from words, then syntactic frames should be more potent than the specific noun in Spanish but not in English and Spanish-speaking children should therefore show a stronger effect of syntactic frame than English-speaking children.

3. Experiment 2

Unlike the idealized training sets of the networks, in the real world-learning environment, many different kinds of cues are correlated—to varying overlapping degrees—with category structure. There is the specific noun: *water* refers to a material-based category, *cup* to a shape-based category. There is the syntactic frame: "some *singular noun*" refers to a

material based category, “*a singular noun*” refers to a shape-based category. And, there are the perceptual properties of the things referred to: nonsolid things such as water are typically in material based categories; whereas solid things like cups are typically in shape based categories (see Samuelson and Smith 1999; Colunga and Smith 2005). The relation between solidity and category structure is somewhat lopsided; at least among the categories typically known by young children (Samuelson and Smith 1999), solid things are typically in shape-based categories but sometimes in material based categories (e.g. block versus wood). Non-solid things are rarely in shaped based categories. For these reasons, we test the effect of syntactic frame on children’s categorizations of solid things. The expectation is that children will be strongly biased to extend the names of these solids things by shape (a bias that can only be reinforced by a count syntactic frame). The question, then, is the effect of mass syntax. Will it be potent enough to cause children to extend the name of a solid object by material? And, most critically, will it be more potent for children learning English or for children learning Spanish? A cross-linguistic difference such that English-speaking children are more sensitive than Spanish-speaking children would support the idea that children learn bi-directional correlations; greater sensitivity of the Spanish-speaking children to syntactic frame would support the idea that children use words (syntactic frame + noun) to *predict* category structure.

3.1. *Method*

3.1.1. *Participants* Thirty-two monolingual English-speaking 2-3 year-olds (range 2.13,3.84 M = 3.05) were tested in Bloomington, IN; 32 monolingual Spanish-speaking 2-3-year-olds (range 2.24,3.31 M = 2.91) were tested in Monterrey, NL, Mexico.

3.1.2. *Stimuli* Common objects and substances were used in a familiarization task prior to the main experiment. These included two spoons, two chocolate bars, a lemon, a pencil, a pair of glasses, a biscuit and a slice of bread. The experimental stimuli consisted of the two sets of made-up things, as shown in Figure 6. Each set consists of an exemplar and 8 test objects that matched the exemplar in specific ways as illustrated. We chose to include more test objects examining kinds of material matches than test objects examining shape matches because past research indicates that the syntactic context modulates attention to material more than to shape (Soja 1992; Soja et al. 1991; Gathercole 1997.) The novel nouns used were Dugo and Zup in English and Dugo and Mepa in Spanish.

Exemplar	ID	Shape	Material Whole	Material Piece	Col+Mat Whole	Col+Mat Piece	Color	None
blue sponge	blue sponge	white stuff	green sponge	pink sponge	blue sponge	blue sponge	blue feather	orange scrub
red clay	red clay	gray granite	yellow clay	blue clay	red clay	red clay	red leather	blonde wood

Figure 6. Stimuli for experiment 2. Each set consisted of an exemplar and eight test items: (1) an identity match, (2) a shape match, (3) a color match, (4) an object that differed from the exemplar on all properties, and four different kinds of material matches—(5) a piece of material match in which the substance was presented in a shape that appeared broken off from some object, (6) a whole match in which the substance was presented in a constructed and regularly shaped object, (7) a piece color + material match, again in an accidental shape and (8) a whole color + material match in a constructed shape. We chose to include more material than shape matches in the test because past research suggests that both syntactic contexts more strongly modulate attention (or inattention) to material than to shape

3.1.3. *Design* Children in each Language group were randomly assigned to either the Mass or the Count condition. There were a total of 32 test trials, 16 with each of the exemplar sets shown in Figure 5. Testing on each set was presented in a block and the order of the two blocks was counterbalanced across children. In each block, each unique test object was queried twice.

3.1.4. *Procedure* The experiment began with a series of familiarization trials. A stuffed bear was introduced and the familiarization exemplar was named with the appropriate syntax. In the Count condition the experimenter showed the child the spoon and said the bear “wants more spoons” The child was then shown one of the familiarization test items and asked, “Is this a spoon?” Analogously, in the Mass condition the child was shown one chocolate bar and told the bear “wants more chocolate” and then asked about the training items “Is this some chocolate?” There were a total of 8 familiarization trials; during these trials the children were given feedback and instructed to repeat the correct answer.

The experimental trials, using the novel objects and novel names, followed the same script except no feedback was provided.

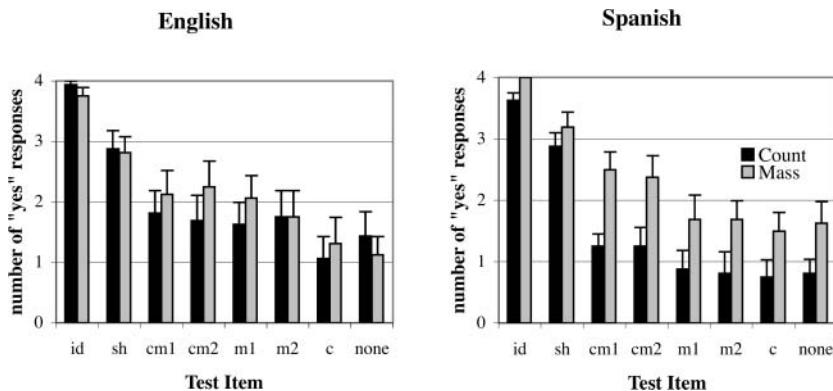


Figure 7. Mean number of "yes" responses for each kind of test item for English- and Spanish-speakers in Experiment 2

3.1.5. Results and discussion The number of "Yes" responses (the name applies) was submitted to an analysis of variance for a 2(Language) \times 2(Syntax) \times 8(Test object) mixed factorial design. The analysis yielded main effects of Syntax, $F(1,52) = 4.7$, $p < .05$, and Test item, $F(7,364) = 72.48$, $p < .001$, and also reliable interactions between Language and Syntax, $F(1,52) = 4.23$, $p < .05$, and between Test item and Syntax, $F(7,364) = 2.06$, $p < .05$. Figure 7 provides the mean number of "yes" responses for each kind of test item and as is obvious there is a much greater difference in responses by Spanish-speaking children in the two syntax conditions than for English-speaking children, a result consistent with the predictions under the assumption of predictive learning.

Post hoc analyses were conducted within each language group comparing the mean number of "yes" responses for each test item in the count and mass conditions (Tukey's, alpha = .05). For the English-speaking children, none of these comparisons were reliable. That is, the English-speaking children said "yes" mainly to items that matched in shape and "no" to those that did not, and they did so to the same degree in both the "count" and "mass" conditions. This result is consistent with Soja's (1992), who found little effect of mass syntactic cues on English-speaking children's extensions of names for solid objects, in spite of robust effects of count syntactic cues on their extensions of names for non-solid substances.

The Spanish-speaking children, in contrast, modulated their responses as a function of the syntactic cue. These children were equally likely to extend the name to the identical test item and to the shape-matching test item in the two syntactic contexts, but were reliably more likely to extend the noun to all other test items in the mass than in the count condition.

As is evident in Figure 7, these effects are particularly strong for the four kinds of material matching test objects.

In sum, there is a bigger effect of count-mass syntactic cues on Spanish-speaking than English-speaking children's novel noun generalizations, a result consistent with a learning algorithm in which the learner predicts the intended meaning of an utterance and thus learns selectively about the most predictive linguistic cues.

4. Conclusion

This paper makes three contributions: First, it suggests that early language learners are not learning bidirectional associations among count/mass linguistic cues and category structure. Rather, children attempt to predict category structure from those cues and thus they learn about the most predictive linguistic cues. This question of how one should conceptualize associative learning—as the mere counting of co-occurrences or as prediction—is central to understanding the way the learner structures the learning task, and ultimately the learning mechanism (Rescorla and Wagner 1972; Rumelhart et al. 1986; Kruschke 1993; Smith 2000a, 200b). Both kinds of learning are part of the human system and in adults can even be differentially engaged by how one structures the task (Billman 1989; Love 2002; Minda and Ross 2004; Bott et al. 2007). Thus, children could just register bi-directional co-occurrences among syntactic frames, specific nouns, and category structure and generalize from these bi-directional patterns when applying a newly learned name to new things. But apparently they do not. Instead, they structure the task differently by predicting meaning (and category structure) from words. Many developmentalists (e.g. Bloom and Tinker 2001; Lidz et al. 2003) have argued for this conceptualization of the language learner because it implies a more active learner. However, this may not be best characterization of the difference nor the most theoretically significant difference. Bi-directional correlations imply that all sides of the correlation are fundamentally the same. Unidirectional predictions imply, in contrast, that the learning mechanism treats words as having a fundamentally different status from category structures. Words, perhaps as symbols, point to—that is, predict—construals. In the present simulations and behavioral data, we see the computational consequences of this profound difference.

Second, the results show that there are cross-linguistic differences in the potency of count-mass cues in English and Spanish (see also Gathercole 1997 and Imai and Mazuka 2007, for similar results with older children). In the present case, these differences derive from the different *correlational* structures of syntactic frames, specific nouns, and category structure in

the two languages and thus supports the very idea that children are learning correlations. This evidence adds to a growing line of findings that suggest that language learning, and phenomena such as syntactic bootstrapping, may depend critically on the structure of the language being learned. Both English and Spanish have count-mass syntax, but this does not mean similar learning trajectories in the two languages. The simulation results make clear that we cannot simply observe different patterns of correlations in two languages and then make straightforward predictions about what the cross-linguistic differences in language learning should be. The presence of a cue as a reliable predictor is not enough to guarantee that the cue will be developmentally potent. We need to know not just about the reliability of individual cues but also about that cue in relation to the whole *system* of overlapping cues to meaning that the child is simultaneously learning.

The third contribution concerns the importance of specifying the learning mechanism and the task. Knowing the system of cues by itself is also not enough. Overlapping cues could compete or they could reinforce each other, and as the simulations show, which is the case depends on the kind of learning mechanism assumed and the details of how the learning task is structured: the same correlational structure leads to different learned outcomes depending on these assumptions. Cross-linguistic differences and their consequences, then, may be understood only with respect to the learning mechanism assumed (see also Sandhofer et al. 2001). In turn, the study of cross-linguistic differences may be the key to a deeper understanding of underlying mechanisms.

Appendix

In a network trained with back-propagation, an input pattern presented to the input units activates the hidden units in the network and these in turn activate the output units. The activation function for both hidden and output units is a non-linear function of the net input to the unit. We used the usual sigmoidal activation function:

$$a_i = \frac{1}{1 + e^{-\text{input}_i}} \quad (1)$$

For each training input pattern, there is a target output pattern. Once the output units have been activated in response to an input pattern, the activation of each output unit is subtracted from the target activation for that unit, yielding an error for each output unit. This error is used to adjust the weights into the output unit. Next the error at the output layer is propagated back to the hidden layer, much as activation is propagated

forward during the activation phase. This yields an error for each hidden unit, which is then used to adjust the weights into that unit from the input units. Specifically, the change in the weight from an input or hidden unit i to a hidden or output unit j is given by

$$\Delta w_{i\otimes j} = \varepsilon \delta_j a_i, \quad (2)$$

where ε is a learning rate, δ_j is the error associated with unit j , and a_i is the activation of unit i . The error term δ_j for an output unit is given by

$$\delta_j = (t_j - a_j) f'_j(\text{input}_j), \quad (3)$$

where t_j is the target for output unit j and f'_j is the derivative of the activation function f for unit j . The error term for a hidden unit is given by

$$\delta_j = [\sum_k \delta_k w_{j\otimes k}] f'_j(\text{input}_j), \quad (4)$$

where $w_{j\otimes k}$ is the weight from hidden unit j to output unit k . In all of the networks we trained the learning rate ε was 0.05.

References

- Billman, D. 1989. Systems of correlations in rule and category learning: Use of structured input in learning syntactic categories. *Language and Cognitive Processing* 4. 127–155.
- Billman, D. & J. Knutson. 1996. Unsupervised concept learning and value systematicity: A complex whole aids learning the parts. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22. 458–475.
- Bloom, L. & E. Tinker. 2001. The intentionality model and language acquisition: Engagement, effort, and the essential tension in development. *Monographs of the Society for Research in Child Development* 66(4). 1–89.
- Bott, L., A. B. Hoffman & G. L. Murphy. 2007. Blocking in category learning. *Journal of Experimental Psychology: General* 136(4). 685–699.
- Colunga, E. & L. B. Smith. 2005. From the lexicon to expectations about kinds: A role for associative learning. *Psychological Review* 112(2). 347–382.
- Gathercole, V. C. M. 1997. The linguistic mass/count distinction as an indicator of referent categorization in monolingual and bilingual children. *Child Development* 68(5). 832–842.
- Gathercole, V. C. M., D. Evans & E. M. Thomas. 2000. What's in a noun? Welsh-, English-, and Spanish-speaking children see it differently. *First Language* 20(58). 55–90.
- Gleitman, L. 1990. The structural sources of verb meanings. *Language Acquisition: A Journal of Developmental Linguistics* 1(1). 3–55.
- Gordon, P. 1985. Evaluating the semantic categories hypothesis: The case of he count/mass distinction. *Cognition* 20. 209–242.
- Iannucci, J. E. 1952. *Lexical number in Spanish nouns with reference to their English equivalents*. Philadelphia: University of Pennsylvania.
- Imai, M. & R. Mazuka. 2007. Language-relative construal of individuation constrained by universal ontology: Revisiting language universals and linguistic relativity. *Cognitive Science* 31(3). 385–413.
- Japkowicz, N. & D. H. Fisher. 2001. Supervised versus unsupervised binary-learning by feedforward neural networks. *Machine Learning* 42(1–2). 97–122.

- Kruschke, J. K. 1993. Human category learning: Implications for backpropagation models. *Connection Science* 5(1). 3–36.
- Kruschke, J. K. & N. J. Blair. 2000. Blocking and backward blocking involve learned inattention. *Psychonomic Bulletin and Review* 7(4). 636–645.
- Kruschke, J. K. 2001. Toward a unified model of attention in associative learning. *Journal of Mathematical Psychology* 45(6). 812–863.
- Landauer, T. K. & S. T. Dumais. 1997. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review* 104(2). 211–240.
- Li, P., B. MacWhinney & X. Zhao. 2007. Dynamic self-organization and early lexical development in children. *Cognitive Science* 31(4). 581–612.
- Lidz, J., H. Gleitman & L. Gleitman. 2003. Understanding how input matters: Verb learning and the footprint of universal grammar. *Cognition* 87(3). 151–178.
- Love, B. C. 2002. Comparing supervised and unsupervised category learning. *Psychonomic Bulletin and Review* 9(4). 829–835.
- Love, B. C., T. M. Gureckis & D. L. Medin. 2004. SUSTAIN: A network of category learning. *Psychological Review* 111(2). 309–332.
- McClelland, J. & D. Rumelhart. 1981. An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review* 88. 375–407.
- Miikkulainen, R. 1997. Dyslexic and category-specific aphasic impairments in a self-organizing feature map model of the lexicon. *Brain and Language* 59(2). 334–366.
- Minda, J. P. & B. H. Ross. 2004. Learning categories by making predictions: An investigation of indirect category learning. *Memory & Cognition* 32(8).
- Quine, Q. V. 1969. *Onological relativity and other essays*. New York: Columbia University Press.
- Regier, T. 2005. The emergence of words: Attentional learning in form and meaning. *Cognitive Science* 29. 819–865.
- Rescorla, R. A. & A. R. Wagner. 1972. A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (eds.), *Classical conditioning. II. Current research and theory*, 64–99. New York: Appleton-Century-Crofts.
- Rogers, T. & J. McClelland. 2004. *Semantic cognition: A parallel distributed processing approach*. Bradford Books.
- Rumelhart, D. E., G. Hinton & R. Williams. 1986. Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (eds.), *Parallel distributed processing* vol. 1, 318–364. Cambridge, MA: MIT Press.
- Samuelson, L. & L. B. Smith. 1999. Early noun vocabularies: Do ontology, category structure and syntax correspond? *Cognition*. 1–33.
- Sandhofer, C. M., J. Luo & L. B. Smith. 2001. Counting nouns and verbs in the input: Differential frequencies, different kind of learning? *Journal of Child Language* 27. 561–585.
- Soja, N. N. 1992. Inferences about the meaning of nouns; the relationship between perception and syntax. *Cognitive Development* 7. 29–45.
- Soja, N. N., S. Carey & E. S. Spelke. 1991. Ontological categories guide young children's inductions of word meanings: Object terms and substance terms. *Cognition* 38. 179–211.
- Smith, L. B. 2000a. Learning how to learn words: An associative crane. In R. M. Golinkoff, N. Akhtar, L. Bloom, K. Hirsh-Pasek, G. Hollich, L. B. Smith, M. Tomasello, A. L. Woodward (eds.), *Becoming a word learner, A debate on lexical acquisition*. New York, NY: Oxford University Press.
- Smith, L. B. 2000b. Avoiding associations when it's behaviorism you really hate. In R. M. Golinkoff, N. Akhtar, L. Bloom, K. Hirsh-Pasek, G. Hollich, L. B. Smith, M. Tomasello,

- A. L. Woodward (eds.), *Becoming a word learner, A debate on lexical acquisition*. New York, NY: Oxford University Press.
- Smith, L. B., E. Colunga & H. Yoshida. 2003. Making an ontology: Cross-linguistic evidence. In L. Oakes & D. Rakison (eds.), *Early category and concept development making sense of the blooming, buzzing confusion*, 275–302. Oxford: Oxford University Press.
- Tang, Z., M. Ishii, H. Tamura & X. Wang. 2003. An algorithm of supervised learning for multilayer neural networks. *Neural Computation* 15(5). 97–122.
- Yoshida, H. & L. B. Smith. 2003a. Shifting ontological boundaries: How Japanese- and English-speaking children generalize names for animals and artifacts. *Developmental Science* 6(1). 1–17.
- Yoshida, H. & L. B. Smith. 2003b. Response: Correlation, concepts and cross-linguistic differences. *Developmental Science* 6(1). 30–34.