The Emergence of Words

Terry Regier (regier@psych.uchicago.edu)
Bryce Corrigan (b-corrigan@uchicago.edu)
Rachael Cabasaan (rrcabasa@uchicago.edu)
Amanda Woodward (alw1@psych.uchicago.edu)
Department of Psychology, University of Chicago
Chicago, IL 60637 USA

Michael Gasser (gasser@indiana.edu)
Linda Smith (smith4@indiana.edu)
Cognitive Science Program, Indiana University
Bloomington, IN 47405 USA

Abstract

Children change in their word-learning abilities sometime during the second year of life. The nature of this behavioral change has been taken to suggest an underlying change in mechanism, from associative learning to a more purely symbolic form of learning. We present a simple associative computational model that accounts for these developmental shifts without any underlying change in mechanism. Thus, there may be no need to posit a qualitative mechanistic change in the word-learning of young children. More generally, words, as symbols, may emerge from associative beginnings.

Overview

Word-learning is likely to rely heavily on associative learning, such that the child comes to associate the sound “dog” with dogs, the sound “cat” with cats, and so on. However, children’s word-learning abilities change significantly during the second year of life, and some have proposed that this behavioral change reflects an underlying mechanistic shift away from a purely associative base. In particular, it has been proposed that sometime during the child’s second year, a conceptual insight into the symbolic, referential nature of words occurs (McShane, 1979). This insight then supports a more purely symbolic form of learning, in contrast with the simple associative learning that preceded it.

A number of changes in word-learning occur at around this age. When viewed as a totality, this array of behavioral changes does suggest a mechanistic change of some sort. We shall argue, however, that these changes may be accounted for without recourse to any posited conceptual insight, or any qualitative mechanistic change in the nature of word-learning. Instead, they flow naturally from a purely associative mechanism, operating over similarity-based representations. As these representations gradually become more peaked and finely differentiated, the child’s linguistic behavior becomes more recognizably “symbolic”. We argue this point by presenting an associative computational model, and demonstrating that it matches the developmental shifts of 1- to 2-year-old children. Thus words, as discrete arbitrary symbols, may emerge from fundamentally associative, similarity-based mental material.

We are not the first to propose this general idea, nor to present a computational model supporting it (Cottrell and Plunkett, 1994; Elman et al., 1996; Merriman, 1999; Plunkett et al., 1992). However, the specific cluster of behavioral issues we address have not, to our knowledge, yet been accounted for computationally.

We begin by briefly outlining the empirical evidence for a change in word-learning during the child’s second year of life. We then present an associative computational model, and demonstrate that it accounts for this change. We also highlight predictions made by the model, and some preliminary evidence in favor of them. We conclude with a discussion of the ramifications of this account.

Empirical Evidence

During the second year of life, the child’s word-learning behavior changes in at least four respects: ease of learning, honing of linguistic form, honing of linguistic meaning, and the learning of synonyms.

Ease of learning

As children first begin to produce words, their acquisition of new words is slow and errorful. New words are added at the rate of 1 or 2 every few weeks (Gershkoff-Stowe and Smith, 1997). Then between 18 and 22 months (when the child has about 50 words in productive vocabulary), the rate of new word acquisition accelerates dramatically, with reports from detailed diary studies of children learning as many as 36 words in a single week (Dromi, 1987). Experimental studies replicate this shift in the laboratory. At the beginning of word learning, 13- to 16-month-olds can acquire a word-object linkage in comprehension based on 4-8 training trials (Bird and Chapman, 1998). By the time children are 2 to 3 years of age, a single learning trial is sufficient for word learning in comprehension and production, and for generalization to an appropriate range of referents. Thus, children appear to shift from learning as a gradual process to the sort of all-or-none process that often characterizes symbolic learning.

Honing of linguistic form

Infants must learn what counts as a word in the language they are learning, and what does not. The developmental evidence suggests that in the beginning, words function as ordinary members of the open set of possible associates. Later, however, the range of acceptable word forms becomes narrower. For example, Namy and Waxman (1998) found that 18-month-olds readily accepted a hand gesture as a word form — in that they learned to associate the gesture with a referent. Older children however, 26-month-olds, did not learn the association. This developmental trend has been replicated using other non-phonological "words" (Woodward and Hoyne, 1999).

There is other evidence that the process is one of "honing" or narrowing the set of possible forms. Although infants readily discriminate between individual phonemes in perceptual tasks, they do not exploit this level of detail in their initial representations of words (Stager & Werker, 1997; see also Bird & Chapman, 1998). Specifically, at the beginning of word learning, at 14 months, babies cannot learn that *bih* and *dih* refer to different items, although they can learn this for globally different forms such as *lif* and *neem*. Thus children seem to move from a state in which they are sensitive primarily to overall similarity or difference between word forms to one in which they are acutely sensitive to minor differences.

Honing of meaning

Just as forms become progressively restricted with development, so do meanings. Early in word learning, 13- and 18- month old children generalize a newly learned object name to new referents by overall similarity across all dimensions (Smith, Jones, Gershkoff-Stowe & Samuelson, 1999). But older children systematically generalize novel names for artifact-like things using the specific dimension of *shape* (Smith et al., 1999). Thus, children come to pay attention to particular dimensions of referents and disregard others - much as they do with word forms. (A distinction however is that this "shape bias" holds for object names and not other sorts of names.)

Synonyms

Children assume that two different forms carry two different meanings. This has been termed the *mutual exclusivity assumption*. One specific manifestation of this assumption is that young children tend to resist learning synonyms. Liittschwager and Markman (1994) found that 16-month-olds, who can learn a new word for an as-yet-unnamed object, have trouble learning a new word for an already-named object (i.e. a synonym). However, 24-month-olds learn novel names and synonyms equally well - they do not exhibit a particular resistance to learning synonyms. Thus, there is a shift in the ease of learning synonyms, one that occurs at about the same age as the other changes in word-learning outlined above.

These roughly simultaneous changes, in ease of learning, form-honing, meaning-honing, and synonym-learning, may suggest an underlying change of mechanism sometime near the second birthday. However, we shall argue that no qualitative change in mechanism is necessary to account for these parallel developmental trajectories. They emerge naturally from a single fundamentally associative mechanism.

Foundational Assumptions

As we have seen, children do not enter the world with a clear sense of what counts as an acceptable word form. But if this is the case, what differentiates word forms from meanings in the first place, in the mind of an infant? Both are experiences of events or objects in the world. We assume that the answer lies in the child's awareness of her interlocutor's stance as a social other. Specifically, we assume that the child will take the object of the interlocutor's *attention* as a potential referent (Baldwin et al., 1996). Further, we assume that those intentional actions of the interlocutor to which the interlocutor is *not* attending are taken as potential forms – this will include verbal utterances, gestures, and any other unattended action. It is known that pre-linguistic infants are sensitive to the object of attention of another person (Corkum and Moore, 1998). Thus, the deployment of the interlocutor's attention serves as a plausible starting-point for the development of the form/meaning distinction.

The Model

The model, shown in Figure 1, builds on these social assumptions. It accepts as input a potential form and a potential referent, which are assumed to have been determined by the interlocutor's deployment of attention. These inputs are each represented by a bank of nodes, corresponding to features of experience. Form and referent are associated in the top layer of the network, which holds a localist lexicon - in which each node stands for a distinct pairing of form and meaning. The form and meaning for a given lexicon node are encoded on its incoming weights. New nodes are added to this lexicon as new form-meaning pairings are encountered (Carpenter and Grossberg, 1988).

The central concept of the model is that different dimensions of experience acquire different degrees of *communicative significance*, or selective attention (Nosofsky, 1986). This is true of both form and meaning, which are represented in the same psychological space. At the beginning of learning, all dimensions are equally, and weakly, weighted, and the model responds in a graded fashion, on the basis of overall similarity. Later in learning, however, some dimensions become very significant, and others insignificant. The model then responds categorically, in the all-or-none fashion characteristic of symbolic representations. It is this transition that we suggest underlies the emergence of words, as symbols.

Formal presentation: Given any input, the model computes the distance of the current input from the weight vector of each lexicon node:

$$d_i = \sqrt{\sum_j s_j (i_j - w_{ij})^2}$$

Here s_j is the communicative significance of dimension j, i_j is the current value (+/-1) of input dimension j, and w_{ij} is the weight on the connection from input node j to lexicon node i. Note that distance is computed over both

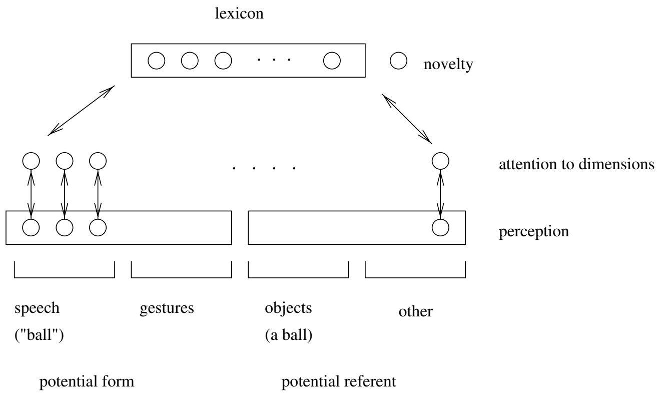


Figure 1: A model of early word learning.

form and meaning dimensions, together. The activation of the lexicon node is then an exponential function of this psychological distance (Shepard, 1987):

$$a_i = exp(-d_i)$$

Thus, the activation for a lexicon node i will be at its maximum (1.0) when d_i is zero - which will occur when the input and weight vectors are identical along significant dimensions. There is also a "novelty node", a novelty detector that is activated to the extent that no existing lexicon node is activated:

$$a_{novelty} = 1 - max_i(a_i)$$

Given these activations, we can compute a probability distribution over the lexicon, including the novelty node, using the Luce choice rule:

$$p_i = \frac{a_i}{\sum_j a_j}$$

A new lexicon node is then created with probability $p_{novelty}$ – the probability associated with the novelty node. If this node-creation does occur, the newly-created node k is incorporated into the lexicon, and its weights are set by the current input values, multiplied by $p_{novelty}$ (Hebb, 1949). Thus, the more clearly novel the input is, the stronger the weights on the newly allocated nodes will be.[1]

$$w_{kj} = i_j \times p_{novelty}$$

If a new node is not created, the most highly activated node k in the lexicon is selected for training. The model is then trained under gradient descent, with a target value of 1.0 for p_k, and target values of 0.0 for all other $p_i, i = k$. This will reinforce node k's representation of the form-meaning pairing currently presented as input, moving its weights w_k closer to the current input, and those of competing nodes farther away. Similarly, the s values (significance weights) will be adjusted to help discriminate lexicon nodes from each other. We then train again with only the form as input, and with the same target outputs. And finally, we train again with only the referent as input, and with the same target outputs. This three-step training causes the selected node k to act as a category node for both form and referent - and thereby to link the two.

The equations above, with the exception of the novelty node activation and creation, are adapted from existing models of categorization (Kruschke, 1992; Nosofsky, 1986). Merriman (1999) has shown that a similar formalism can account for the mutual exclusivity bias and shape bias in word-learning – the present model builds on this work. The combination of form and meaning in a single representation echoes the linguistic production model of Dell et al. (1997). This links theoretically central aspects of our model to existing models of related processes, models that have already received empirical support in their own right.

Testing: Once a set of words has been learned, one may test the model on either comprehension or production of a learned word. In a comprehension test, the word form is supplied to the network, but no referent is supplied. A winning lexicon node k is selected, as above, but from a competition based on the form dimensions only. The referent dimensions of node k's weight vector w_k are then projected down to the referent inputs:

$$i_j = w_{kj}$$

This reconstructed referent constitutes the model's response to the word form supplied as input. Production tests proceed analogously, but with the referent supplied as input, and the form produced as output.

Accounting for Existing Data

In the simulations reported here, the model was trained on a dataset of form-meaning pairings, and tested at various points during training. There were 6 dimensions for form: 4 were significant (such that a pattern over these dimensions was predictive of the referent), and 2

[1] Nodes that are created but not revisited within a given number of epochs are pruned from the lexicon. In current simulations, that number of epochs is 1.

were insignificant (not predictive). Similarly, there were 6 dimensions for meaning: 4 significant (such that a pattern over these dimensions was predictive of the form), and 2 insignificant (not predictive). The training set consisted of 75 variants of 5 words; a variant of a word preserved the significant dimensions of the word while altering insignificant dimensions. The words and their variants were represented by either +1 or -1 on each dimension. The specific values were chosen randomly, subject to the constraint that patterns over the significant dimensions of form be predictive of significant dimensions of meaning, and vice versa. The model was trained on this dataset for 100 epochs. The learning rate was 0.05. As expected, the significance weights differentiated significant from insignificant dimensions increasingly clearly over the course of training. At epoch 6, the difference between the average significance weight over dimensions intended to be significant and the average significance weight over dimensions intended to be insignificant was 1.06608. By epoch 60, this difference was 2.74704, reflecting clearer differentiation with training.

During training, tests were performed in order to probe the model's behavior on the four empirical trends noted at the beginning of the paper. In all cases, this involved presenting a new form-meaning pairing for the model to learn, and then determining the probability of correct comprehension or production. We define a "correct" response to be one that is within 0.9 of the target along all significant dimensions (which generally vary from -1 to 1), using the comprehension and production output rules outlined above. We then calculate the summed probability across all lexicon nodes that would produce a correct response: this yields the probability of correct response.

Ease of learning: We first tested how easily the model could learn a novel form for a novel object, and how that ease changed with age. A new word was created that was representationally distant from the existing words in the training set – specifically, each of form and meaning in this new word differed from those for existing words along all significant dimensions. We shall refer to this word henceforth as the "novel" word. We examined the probability of correct comprehension of this new word after simulating one learning trial on the word at two points during the learning of the training set mentioned above: after 6 epochs, and after 60 epochs. The results are displayed in Figure 2(a). As the model's space stretches along significant dimensions, this new word is increasingly easily learned, eventually being reliably correctly comprehended given only 1 training trial. This progression into 1-trial learning reflects the behavior of 1-2 year old children. Importantly, once the model had learned the novel word, it was removed from the lexicon; thus, later "ages" of the model did not have the benefit of earlier training on the novel word – only of an appropriately stretched psychological space, which caused the word to be perceived as distinct, and therefore easily remembered.

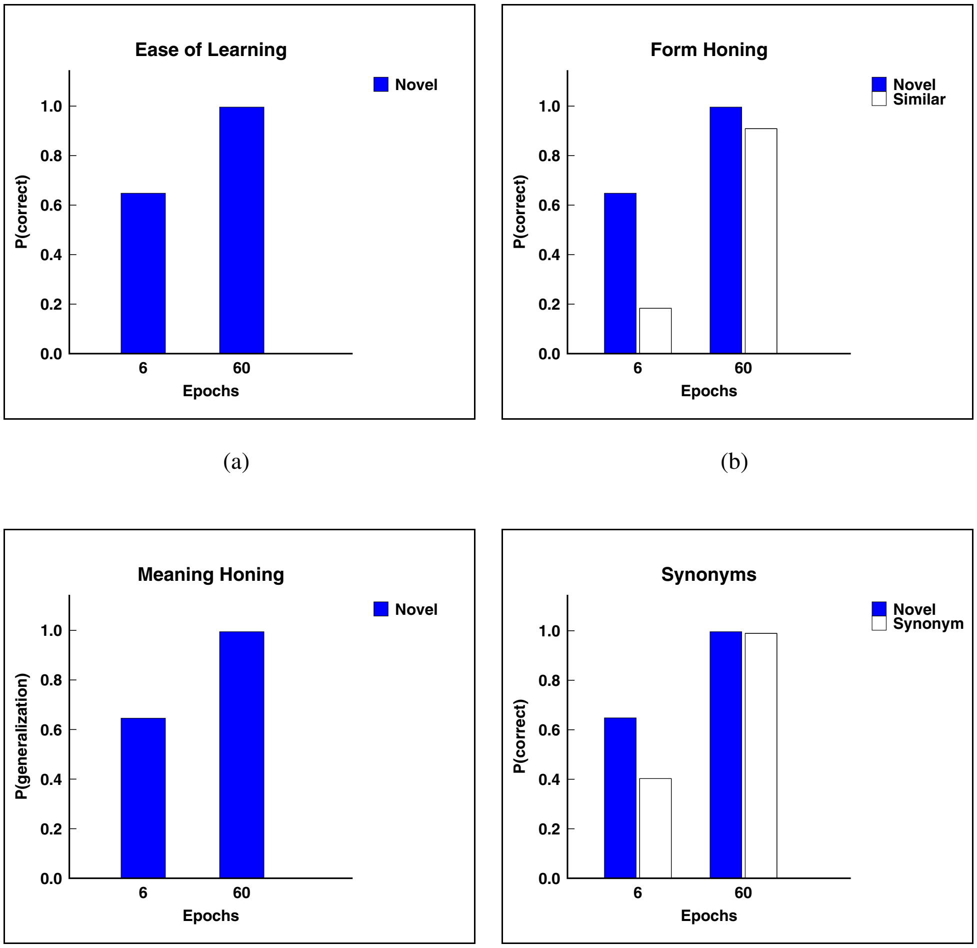
Honing of linguistic form: We next examined the learning of a new word that was *similar in form* to an existing word in the training set. The form of this new word differed from that of an existing word in the lexicon by only 1 significant dimension, while the meaning dimensions differed from other words along all significant dimensions. Thus, this test simulates the potential confusion of learning "bih" and "dih" associated with different sorts of objects (Stager and Werker, 1997). The probability of correct comprehension after one training trial is shown by the crosshatched bars in Figure 2(b). In this figure, the solid bars duplicate the presentation of the model's behavior on the novel (dissimilar) word in (a), for purposes of comparison. As is true of children, similar words are initially somewhat more difficult to learn than are globally dissimilar words. However, eventually these similar words are also successfully learned given one training trial, as the relevant dimensions of space are highlighted, counteracting the confusing similarity. This allows fast learning of minimal pairs such as "bit" and "pit".

Honing of meaning: We were interested in determining whether the model would exhibit a strengthening shape bias, as children do. To test this, as before, we trained the model on the novel word, and then tested the probability of producing the novel word for a different object, which differed from the original in meaning along insignificant (non-shape) dimensions only. This probability of generalization is shown in Figure 2(c). The increasing strength of generalization along the significant dimensions follows directly from the increasing perceived communicative significance of those dimensions (see Merriman (1999) for a similar demonstration). This is analogous to the honing of linguistic form. This account is incomplete however. In actuality the shape bias applies only to object names; thus an additional mechanism would be required to determine whether a given word is an object name, and therefore whether the model's bias should apply.

Synonyms: A synonym for an existing word in the lexicon should be difficult to learn, since it is similar (identical in meaning) to that existing word. Figure 2(d) shows that this is the case in the present model; Merriman (1999) reports similar results with a similar model. The probability of correct comprehension after one training exposure is initially lower for a synonym of an existing word in the lexicon than it is for the (dissimilar) novel word examined above. This matches the findings of Liittschwager and Markman (1994). Eventually however the synonym and the non-synonym are approximately equally likely to be learned – as is eventually true with children. In the model, this is accounted for by the stretching of the underlying psychological space with age, such that even similar lexical entries are kept distinct, and thereby effectively learned.

Predictions

The model makes two predictions. The first is that young children should experience difficulty learning *homonyms* (a single form with multiple meanings, such as the "bank" of a river, and a "bank" as a financial institution). Moreoever, this difficulty should be correlated



(a)

(b)

(c)

(d)

Figure 2: Developmental trends, as exhibited by the model.

with the difficulty of learning synonyms, as the reason for the difficulty is analogous. Half of the model's lexical representation for a homonymous or synonymous word will be identical to that for another word in the lexicon: the identical half is the form for homonyms, and the meaning for synonyms. This means that the two lexical entries in question will tend to be nearer each other in psychological space than two non-homonyms or two non-synonyms, and will therefore interfere with each other. Doherty (2000) has found that understanding of homonymy is strongly associated with understanding of synonymy, in 3-4 year old children.[2] Similar tests on younger children would more directly test this prediction.

The second prediction also concerns the interaction of form and meaning. As we have seen, 14-month-olds have trouble associating similar sounds such as "bih" and "dih" with different referents (Stager and Werker, 1997). On the model's account, this is because the forms are too similar, such that the two lexical representations lie confusingly near each other in psychological space. But since that space contains both form and meaning dimensions, the model predicts that an exaggerated *semantic* difference between the referents should compensate for the confusing formal similarity of "bih" and "dih" in such a task, and should make learning easier. This prediction has not yet been tested.

Discussion

1- to 2-year old children seem to undergo a qualitative change in the manner in which they learn words. It has been suggested that this behavioral change reflects a conceptual insight into the symbolic nature of words. The model we have presented, however, suggests a different, and more parsimonious, account of the same phenomenon. The behavioral change may result not from

[2] These were both also correlated with understanding of false belief.

an abrupt insight, but rather from an associative learner gradually determining which aspects of the world are relevant for communication. In this manner, the symbolic use of words may emerge from an associative base.

Acknowledgments

This work was supported by grant R01-DC03384 from the National Institutes of Health.

References

Baldwin, D., Markman, E., Bill, B., Desjardins, R., and Irwin, J. (1996). Infants' reliance on a social criterion for establishing word-object relations. *Child Development*, 67:3135–3153.

Bird, E. K. and Chapman, R. S. (1998). Partial representation and phonological selectivity in the comprehension of 13- to 16-month-olds. *First Language*, 18:105–127.

Carpenter, G. A. and Grossberg, S. (1988). The ART of adaptive pattern recognition by a self-organizing neural network. *Computer*, pages 77–88.

Corkum, V. and Moore, C. (1998). The origins of joint visual attention in infants. *Developmental Psychology*, 34:28–38.

Cottrell, G. and Plunkett, K. (1994). Acquiring the mapping from meaning to sounds. *Connection Science*, 6(4):379–412.

Dell, G., Schwartz, M., Martin, N., Saffran, E., and Gagnon, D. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, 104(4):801–838.

Doherty, M. J. (2000). Children's understanding of homonymy: Metalinguistic awareness and false belief. *Journal of Child Language*, 27:367–392.

Dromi, E. (1987). *Early Lexical Devlopment*. Cambridge University Press, New York.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., and Plunkett, K. (1996). *Rethinking Innateness: A Connectionist Perspective on Development*. MIT Press, Cambridge, MA.

Gershkoff-Stowe, L. and Smith, L. B. (1997). A curvilinear trend in naming errors as a function of early vocabulary growth. *Cognitive Psychology*, 34:37–71.

Hebb, D. (1949). *The Organization of Behavior*. Wiley.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, 99:22–44.

Liittschwager, J. and Markman, E. (1994). Sixteen and 24-month-olds' use of mutual exclusivity as a default assumption in second label learning. *Developmental Psychology*, 30:955–968.

McShane, J. (1979). The development of naming. *Linguistics*, 17:879–905.

Merriman, W. (1999). Competition, attention, and young children's lexical processing. In MacWhinney, B., editor, *The Emergence of Language*, pages 331–358. Lawrence Erlbaum Associates, Mahwah, NJ.

Namy, L. L. and Waxman, S. R. (1998). Words and gestures: Infants' interpretations of different forms of symbolic reference. *Child Development*, 69:295–308.

Nosofsky, R. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, 115(1):39–57.

Plunkett, K., Sinha, C., Moller, M., and Strandsby, O. (1992). Symbol grounding or the emergence of symbols? vocabulary growth in children and a connectionist net. *Connection Science*, 4:293–312.

Shepard, R. (1987). Toward a universal law of generalization for psychological science. *Science*, 237:1317–1323.

Smith, L. B., Jones, S., Gershkoff-Stowe, L., and Samuelson, S. (1999). The origins of the shape bias. Submitted to the *Monographs of the Society for Research in Child Development*.

Stager, C. L. and Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, 388:381–382.

Woodward, A. L. and Hoyne, K. L. (1999). Infants' learning about words and sounds in relation to objects. *Child Development*, 70:65–77.