CLS 38: *The Panels* (2002); 235-245

A1

The associative origin of words

Jesse Harris and Terry Regier
University of Chicago

1.0 Overview

Among the very earliest stages of language acquisition is the development of a lexicon, by which a child associates the form of a word with its meaning. It is not yet clear exactly how the child accomplishes this. It has been observed that the type of learning shifts from slow and errorful acquisition during the first period of word learning to rapid and efficient acquisition usually at 2 – 3 years. One proposal has characterized this transition as a shift from associative learning to specialized symbolic learning (McShane 1979), such that the child gains a referential "symbolic insight", which enables the rapid learning of novel words. In contrast, we present an account of this transition in a fully associative model. We argue that the improvement in learning may result from increased attention to those aspects of the world that carry communicative content, and that this increased attention reduces memory interference between words. Thus, there may be no need to posit a specialized symbolic insight in order to account for the improvement in word learning – linguistic symbols may emerge from associative origins.

In this paper, we first present empirical evidence that children in their second year of life exhibit four parallel transitions in word learning ability - suggesting an underlying shift in learning mechanism. We then present an associative model, first in intuitive and then in formal terms, and demonstrate that the model accounts for each of these four transitions.

2.0 Ease of Learning

In the earliest stages of lexical acquisition, at 13 – 15 months of age, children require 9 - 12 trials before they successfully map a word to its referent (Schafer and Plunket 1998; Woodward et al 1994). Older children, in contrast, at 2 – 3 years of age require only 1 – 3 training trials ('fast mapping') (Behrend et al., 2001; Carey 1978). Observational studies reveal a similar improvement in word learning (Dromi 1987): children's word-learning accelerates – sometimes sharply – at this age ("the word spurt"). In order to account for this transition, a shift in learning mechanism has been posited: sometime during the second year the child may acquire a conceptual insight into the symbolic, referential character of words (McShane 1979).

2.1 The Honing of Linguistic Form

As children encounter and learn more words, they begin to build expectations about what forms words may take. Infants at 18 months will learn to attach meaning to a broad array of potential word forms: gestures, pictograms, nonlinguistic 'beeps' from a noisemaker, and spoken words. Older children, in contrast, will learn to attach meanings only to potential word forms of the type used in the child's native language – typically, spoken words (Namy & Waxman, 1998; Namy, 2001; Woodward & Hoyne, 1999).

There is further evidence that children's conception of what form a word may take becomes refined with age. Infants at 14 months cannot associate similar word forms such as *bih* and *dih* with different referents, but have little trouble associating globally dissimilar forms such as *lif* and *neem* with different items (Stager and Werker 1997). Thus, while developing preferences for spoken words over other potential word forms, children also start to make fine grained distinctions *within* the class of spoken words, eventually becoming sensitive to minor phonemic differences.

2.2 Generalization and the Shape Bias

The manner in which children generalize newly-learned words may also change at approximately this age. In one study, 13-month-olds were taught a novel name for an object, and then tested to see whether they would extend it to other objects differing from the original only in color, but with the same shape as the original. These children sometimes extended the name to such exemplars, but sometimes not (Woodward et al., 1994). Older children, in contrast, reliably extend newly-learned object nouns largely on the basis of object shape, rather than other characteristics such as color or material (Landau, Smith, and Jones, 1988; Smith et al, 2002; but see also Booth and Waxman, 2002). This aspect of word-learning has been explained in associative terms as a reflection of selective attention for object shape: when multiple words are learned, all of which are well-characterized by some dimension, such as object shape, attention shifts to that dimension, thus enhancing the acquisition of yet more words defined along that dimension (Smith et al., 2002).

2.3 Learning synonyms: the Mutual Exclusivity assumption

Again at about the same age, children improve in their ability to learn synonyms. In one word-learning study, 16-month-olds were found to be capable of learning a new word if it was a word for an as-yet-unnamed object – but not if the intended referent already had a name, that is, if the word was a synonym. Older children, at 24 months of age, could learn both synonyms and non-synonyms (Liittschwager and Markman, 1994). There is still, however, a persisting preference for non-synonymous mappings (Markman, 1987).

These four developments – increased ease of learning, refinement of linguistic form, generalization by shape, and ability to learn synonyms – all occur during the second year of life. These parallel behavioral shifts thus suggest an underlying mechanistic change at that age – consistent with the proposed symbolic insight, and a move away from purely associative learning. However, as we shall see, these four shifts can all be accounted for in associative terms.

3.0 The Lexicon as Associated Exemplars (LAE)

We propose that the parallel behavioral transitions sketched above may spring from shifts in *selective attention*: toward dimensions of form – and of meaning – that are relevant for communication, and away from other dimensions. These shifts in attention are both the result of associative learning, and the enhancer of future associative learning. We pursue these ideas through the Lexicon-as-Associated-Exemplars (LAE) computational model (Regier, under review). This model generalizes an existing associative model of category-learning in adults (Kruschke, 1992) – thus grounding an account of word-learning in associative principles of category-learning.

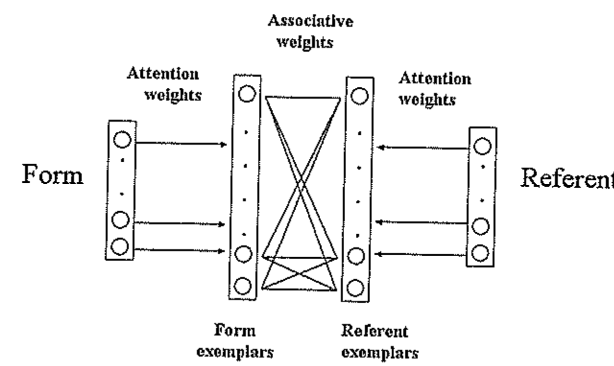

Figure 1: The Lexicon as Associated Exemplars (LAE).

There are 3 central features of the model (see Figure 1). First, the model contains representations of *exemplars*: particular instances of form and referents that have been encountered in training. Similar exemplars will tend to be easily confused in memory, while dissimilar exemplars will be less easily confused. Second, exemplars of form are linked one-to-one to exemplars of meaning (i.e. referents) through bidirectional associative connections. Third, the exemplars reside in a multi-dimensional psychological space (either form space or referent space), and the dimensions of that space stretch and contract (Nosofsky, 1986) to facilitate learning. When a dimension of space receives selective attention, space is effectively stretched along that dimension – such that exemplars differing along that dimension become more psychologically distant, and thus less similar. For instance, if voicing gradually receives more attention in the course of learning – which we would expect, as it carries communicative content – the voicing dimension of form space will be stretched, such that the phonological distinction between word-forms such as "bit" and "pit" will become more salient. A similar process operates over meaning space, highlighting those aspects of the referent that are predictive of naming; for object names, this will tend to be object shape.

Not all dimensions are communicatively significant. For instance, the color of a ball is not predictive of the choice of object noun used to refer to it ("ball"), and the pitch at which the word "ball" is spoken is not predictive of what is intended by the utterance. The LAE model, in the course of learning, comes to pay less attention to such insignificant dimensions, effectively shrinking these dimensions. In the extreme, when no attention is paid to a dimension, exemplars that differ along that dimension collapse together into a cluster. The overall picture that emerges is one in which exemplars of form are connected one-to-one to exemplars of meaning; both form and meaning exemplars then gradually coalesce into clusters that correspond to a word's form and its meaning.

3.1 Formal exposition

The LAE is provided with a form input vector, representing an exemplar of form, specified over a set of form dimensions, and a referent input, similarly specifying an exemplar of meaning. Without loss of generality, we consider the case of predicting meaning from form; the model operates analogously in the other direction. These equations are adapted from (Kruschke, 1992).

Given a form input vector, LAE computes the distance between that input and the position in form space of each form exemplar node i. This distance d_i is weighted by the amount of attention currently allocated to each dimension of form space:

$$d_i = \sqrt{\sum_j \alpha_j (input_j - e_{ij})^2}$$

Here, α_j denotes the amount of attention currently paid to dimension j of form; $input_j$ denotes the value (0 or 1) of the form input on dimension j; and e_{ij} denotes the value of form exemplar e along dimension j. Given this, the activation a_i of exemplar i is computed as:

$$a_i = \exp(-d_i^2)$$

The activation of the exemplar is inversely related to its attentionally-weighted distance from the input. Form exemplars then activate associated referents:

$$net_i = \sum_j w_{ij} a_j$$

net_i is the activation of referent (meaning exemplar) i, and w_{ij} is the weight on the associative link between form exemplar j and meaning exemplar i. These associatively-induced activations then yield a probability distribution p_i over meaning exemplars i:

$$p_i = \frac{net_i}{\sum_j net_j + noise}$$

where *noise* is a constant set to 0.1

Finally, an error quantity $E_{f\to r}$ (error in predicting referent from form) is calculated. This is a measure of the match between what we expect the form to refer to, and what it does in fact refer to.

$$E_{f\to r} = 1.0 - \sum_i g_i p_i$$

Here, g_i is an activation value which is analogous to a_i, but the distance used is that between the referent input (which is the target of our prediction) and referent exemplar i in memory. This distance is modulated by referent attention weights. Thus error is 0 when the probabilities p_i over referents, induced by the given form, are entirely concentrated on referent exemplars that are maximally activated by the current referent input. Form attention weights α_j are adjusted by gradient descent in $E_{f\to r}$ – thus optimizing "comprehension" of which referent the form refers to.

This entire process, from input to error computation, also operates in the other direction, yielding $E_{r\to f}$, the error in predicting form given the referent. Referent attention weights are adjusted by gradient descent in this quantity, optimizing "production" of an appropriate name, given a referent. Associative weights are adjusted by gradient descent in the sum of these two error terms, optimizing both comprehension and production. (Given a particular form-referent pairing, only one associative weight is updated: that between a form exemplar chosen to represent the input form, and a referent exemplar chosen to represent the current referent.) New form and referent exemplar nodes are allocated as a function of their distinctness from existing exemplars.

3.3 Training

The model was trained on a faux data set, containing 25 form-meaning patterns; there were 50 dimensions of form (25 of predictive of meaning; 25 random), and 50 of meaning (25 predictive of form; 25 random). During learning, the model allocated attention to predictive dimensions of form and meaning, and away from non-predictive dimensions – resulting in the stretching and shrinking of psychological spaces referred to above. Figure 2 shows these shifts in selective attention as a function of epoch, for dimensions of form; dimensions of meaning showed qualitatively the same pattern.

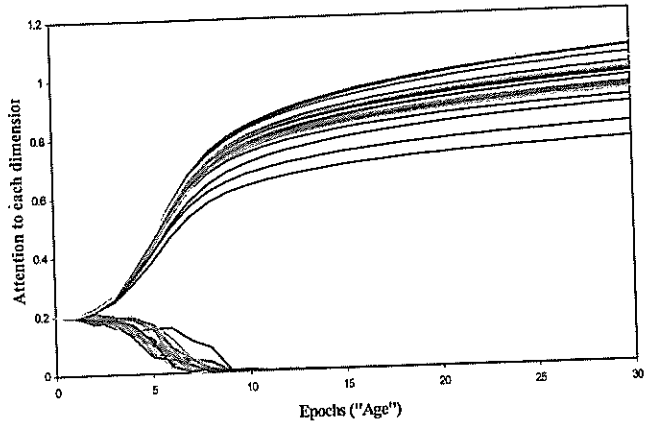

Figure 2: Attention to dimensions of form, as a function of training epoch. Dimensions that are predictive of meaning receive attention, at the expense of other dimensions. This "stretches" the form space along attended dimensions, pulling apart exemplars that vary along these dimensions, and making them less confusable.

Critically, the model was tested to see how well it could learn a novel form-meaning pair (one not in the training set) given only one exposure, at each point in its training. After this single exposure of training, we recorded *p(correct)*, the model's probability of choosing the target referent from among two distractors, upon being provided with the newly-learned form. This probability of correct choice is determined as:

$$p(correct) = \frac{e_t + (residual/3)}{e_t + e_{d1} + e_{d2} + residual}$$

where e_j is the evidence favoring the pairing of this form with referent j. Here, t is the target referent, $d1$ is the first distractor referent, and $d2$ is the second distractor referent. The evidence e_j is defined as:

$$e_j = \sum_{i \in F} g_i p_i + \sum_{i \in M} g_i p_i$$

where F and M are the sets of form and meaning exemplars, respectively, and p_i and g_i are defined as above. This captures the notion of a match between the actual form and the form that would be predicted based on the referent, and the match between the actual referent and the referent that would predicted based on the form. The maximum possible value for e_j is 2.0, so a residual value is also calculated:

$$residual = 2.0 - \max(e_t, e_{d1}, e_{d2})$$

and contributes equally to the evidence for the target and the distractors. After each test, the weight changes from the test were erased, such that they did not affect testing at later epochs.

We next describe a set of such tests, addressing the four behavioral shifts on which this paper is focused.

4.0 Ease of learning, revisited

We trained the model for one exposure on a form-meaning pairing that was distant (in both form and meaning) from all patterns in the training set. This corresponds to teaching a child a novel form such as "toma" for an unfamiliar object. We noted *p(correct)*, the model's probability of correctly choosing the target object from among 2 distractors, upon hearing the newly-learned word. This quantity is shown by the solid line in Figure 3. Chance performance is 1/3.

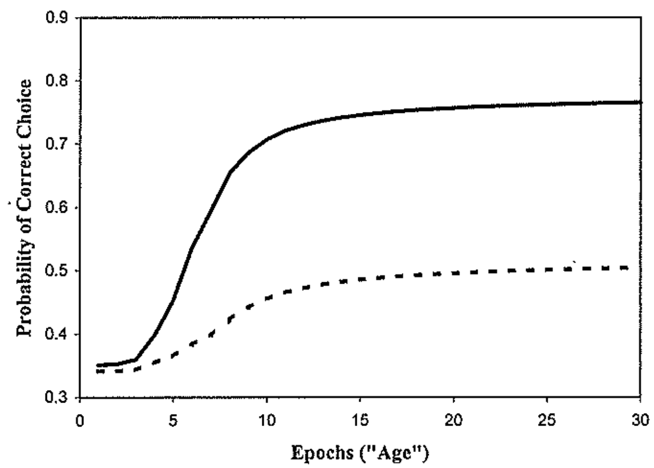

Figure 3: Learning a novel word given one exposure, at different points in training. The solid line shows comprehension of a novel word; the dotted line shows comprehension of a new word that is similar in form to a known word.

The model improves markedly in its ability to learn a word from a single exposure sometime between epochs 5 and 10. Comparison with Figure 2 reveals that this is the time when attention becomes strongly allocated to predictive dimensions, and is taken away from non-predictive dimensions. This attentional allocation improves learning, since it keeps exemplars from different form and meaning categories psychologically distant, such that they are less likely to be confused with each other.

4.1 Honing of linguistic form, revisited

We were interested in determining whether the LAE model, like children (Stager & Werker, 1997) would exhibit worse performance when learning a word that was similar in form to an existing word, than when learning one that was not similar. To that end, we repeated the test above, but this time we used a word-form that was similar (only 1 bit different) from one of the forms in the training set. The vector for the referent was identical to that in the earlier test, and the procedure for the test was also identical. The results may be seen in the dotted line in Figure 3 – the probability of correct comprehension is near chance for longer than in the case of the dissimilar word – in keeping with empirical findings. This poorer behavior stems from the similarity of the forms – there is interference from the existing mapping. As the space stretches, this similarity becomes a less serious problem, eventually enabling the learning of very similar words, such as minimal pairs.

4.2 The Shape Bias, revisited

We also wished to determine whether the LAE model, like children, would exhibit an increase in ability to generalize newly-learned words along significant dimensions – e.g. the dimension of shape for object nouns. An example would be learning that a red ball is named "ball", and then understanding the word to also apply to a green ball. To test this, the model was tested as in the original (ease of learning) test, but with one difference. In the comprehension test, none of the three available referents was the one on which the model had just been trained – however, one of them was identical to it on all significant dimensions, while differing from it along all insignificant dimensions. We recorded *p(correct)*, the probability of selecting the variant of the original referent. This is shown as the dotted line in Figure 4: the probability of correctly generalizing in this manner increased at the same time as the two other transitions already observed. This is due to the increased attention paid to significant dimensions, and decreased attention to others: this attentional shifting causes the variant of the original referent to be psychologically almost identical to the original, such that it comes to be treated (named) like it.

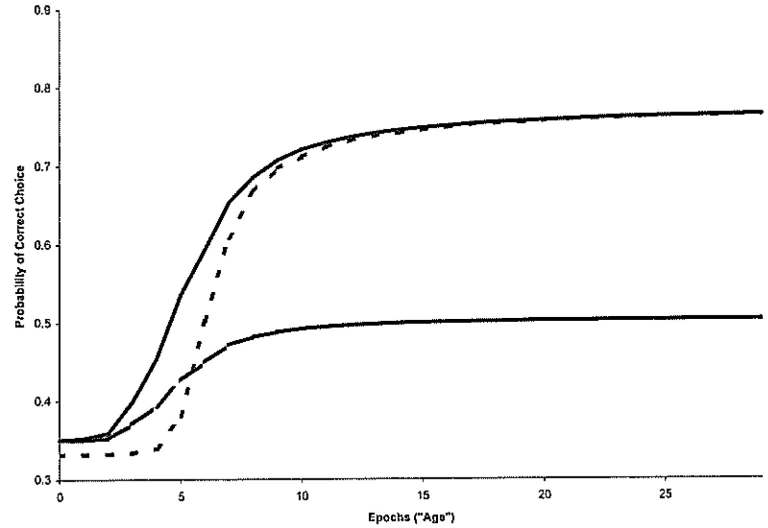

Figure 4: Probability of correct form-meaning pairings, according to different stages in training. The solid line shows the learning of a novel word, for reference; the dotted line indicates learning of the shape bias among a distractor; and the dashed line represents the learning of synonyms.

4.3 Mutual Exclusivity Assumption

Finally, we wished to determine whether the model would exhibit difficulty in learning synonyms, giving way to an eventual ability to learn them. To this end, we conducted another test, but this time the novel form was paired with a referent that was already named in the training set – a synonym. The results are shown in the dashed line in Figure 4: once more, the probability of a correct response improves significantly between 5 – 10 epochs. Importantly, synonyms are resisted throughout learning, as is true of children (Markman, 1989), and early in learning, they are not learned at levels appreciably greater than chance, even at ages when non-synonymous mappings are successfully learned (the dark line shows the original, non-synonymous, test for comparison). This early inability to learn synonyms, followed by an eventual ability, matches the empirical findings of Liitschwager and Markman (1994). The model exhibits this effect because of the same general principles of similarity and confusability that explain the improved learning of non-synonyms. Synonyms are the extreme case of words that are similar in meaning – they are identical in meaning, and thus interfere strongly with each other in memory. This interference is especially strong early in learning, when – as we have seen – even unrelated words interfere with each other. Later in learning, as the spaces stretch, there is some improvement.

5.0 Discussion

Children gain an increased ability for word learning during the second year of life, reflected in four distinct shifts that occur at around this age. Such parallel shifts may seem to suggest an underlying qualitative change in mechanism, such as the proposed referential symbolic insight. However, we hope to have shown that there is a simpler explanation for these phenomena – one that does not posit a sudden qualitative change in mechanism, but only a gradual, quantitative change in attentional deployment. This simpler explanation is grounded in independent principles of exemplar-based associative categorization – suggesting that some central aspects of early word-learning may emerge from very general learning mechanisms. While the ability to refer is clearly central to word-learning, it may not be the primary force behind the accelerated word-learning of young children.

Acknowledgments

The ideas behind this work grew out of interactions with Amanda Woodward, Linda Smith, Michael Gasser, and William Merriman. Programming and testing of the model were done in part by David Burkett, Rachael Cabasaan, Bryce Corrigan, and John O'Leary. This work was supported by grant DC03384 from the National Institutes of Health.

References

Behrend, D., Scofield, J., and Kleinknecht, E. 2001. Beyond fast mapping: Young children's extensions of novel words and novel facts. *Developmental Psychology 37*: 698-705.

Booth, A. E. and Waxman, S. R. 2002. Word learning is 'smart': Evidence that conceptual information affects preschoolers' extension of novel words. *Cognition 84: B11-B22.*

Carey, S. 1978. The child as word learner. In Halle, M., Bresnan, J., and Miller, G., editors, *Linguistic Theory and Psychological Reality*, pp. 264-293. MIT Press, Cambridge, MA..

Dromi, E. 1987. *Early Lexical Development*. Cambridge University Press, New York.

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

Landau, B., Smith, L., and Jones, S. 1988. The importance of shape in early lexical learning *Cognitive Development 3:299-321.*

Liittschwager, J. and Markman, E. 1994. Sixteen and 24-month-old's use of mutual exclusivity as a default assumption in second label learning. *Developmental Psychology 30: 995-968.*

Markman, E. 1987. *How children constrain the possible meanings of words.* In Neisser, U. (ed), 1987. *Concepts and conceptual development: Ecological and intellectual factors in categorization. Emory symposia in cognition, 1. pp. 255-287.* Cambridge University Press, New York.

Markman, E. 1989. *Categorization and naming in children: Problems of induction.* MIT Press, Boston, MA.

McShane, W. 1979. The development of naming. *Linguistics 17:879-905.*

Namy, L. 2001. What's in a name when it isn't a word? 17-month-olds' mapping of nonverbal symbols to object categories. *Infancy 2*: 73-86.

Namy, L., and Waxman, S. 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.*

Regier, T. Under review. The emergence of words: From associations to symbols. Ms.

Schafer, G. and Plunkett, K. 1998. Rapid word learning by fifteen-month-olds under tightly controlled conditions. *Child Development 69*: 309-320.

Smith, L. 2000. Learning how to learn words: an associative crane. *Becoming a Word Learner: A Debate on Lexical Acquisition* Golinkoff (ed) Oxford University Press, New York.

Smith, L., Jones, S., Landau, B., Gershkoff-Stowe, L., and Samuelson L. 2002. Object name learning provides on-the-job training for attention. *Psychological Science 13: 13-19.*

Stager, C.L. and Werker, J. F. 1997. Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature 338: 381- 382.*

Woodward, A. and Hoyne, K. 1999. Infants' learning about words and sounds in relation to objects relation to objects *Child Development 70:65-72.*

Woodward, A., Markman, E., and Fitzsimmons, C. 1994. Rapid word learning in 13- and 18-month-olds. *Developmental Psychology 30: 553-566.*