Emergent constraints on word-learning: a computational perspective

Terry Regier

Department of Psychology, University of Chicago, 5848 South University Avenue, Chicago, IL 60637, USA

In learning the meanings of words, children are guided by a set of constraints that give privilege to some potential meanings over others. These word-learning constraints are sometimes viewed as part of a specifically linguistic endowment. However, several recent computational models suggest concretely how word-learning – constraints included – might emerge from more general aspects of cognition, such as associative learning, attention and rational inference. This article reviews these models, highlighting the link between general cognitive forces and the word-learning they subserve. Ultimately, these cognitive forces might leave their mark not just on language learning, but also on language itself: in constraining the space of possible meanings, they place limits on cross-linguistic semantic variation.

Learning the meanings of words is a challenging inductive problem: when a new word is used to refer to some object or event, it is not clear which *aspect* of the object or event the word captures [1]. Young children seem to cope with this uncertainty by relying on a set of word-learning constraints. These constraints bias children toward adopting particular sorts of meanings for new words – and away from adopting others. There is considerable evidence for such word-learning constraints in children [2–5].

These constraints could be viewed as part of a human predisposition for language – a set of specifically linguistic semantic expectations, by analogy with the syntactic expectations that are thought to assist the child in acquiring grammar [6]. Another possibility, however, is that word-learning might emerge from cognitive processes that are not geared specifically for language [7,8], and that word-learning constraints are linguistic reflections of these more general processes [9,10].

This emergentist view has been made computationally concrete by several recent models of word-learning. No single one of these computational models provides a comprehensive account of word-learning – but when viewed as a group, they provide useful insights into possible mechanisms by which children learn words. This paper reviews these models, and distills four general principles from them: (1) general-purpose learning can account for important aspects of word-learning; (2) word-learning appears to reflect the use of 'accelerating representations', such that early learning builds expectations that in turn assist later learning; (3) the induction of word meaning can be informed both by direct perceptual experience, and by indirectly inferred semantic patterns; and (4) emergent constraints on word-learning can give rise to cross-linguistic semantic universals. These principles are described in turn, and exemplified through specific models.

Explaining word-learning through general learning processes

Many models of word-learning are grounded in general learning processes, rather than language-specific ones. Either implicitly or explicitly, these models suggest that although word-learning constraints are linguistic in nature, the learning mechanisms they spring from might not be. Instead, these linguistic constraints might emerge from general learning processes as they operate on linguistic experience.

Associative learning

Some word-learning constraints may emerge from associative learning. An example is the 'mutual exclusivity' principle, which holds that if an object has one name, it should not have another. Such a constraint could potentially be useful in word-learning: mutual exclusivity could prevent the child from overextending the name 'cat' to include dogs, if the child knew that dogs are named 'dog'. There is substantial evidence that children do in fact respect the mutual exclusivity principle, although it operates as a soft, probabilistic constraint, rather than a hard-and-fast rule [2,4]. There are several computational approaches to mutual exclusivity, showing that this principle can effectively guide generalization, even in the case of categories that are not actually mutually exclusive [11].

How might such a constraint emerge from associative learning? Consider a simple example: a world in which there are only two words, 'dog' and 'cat', each of which might refer to a DOG, or to a CAT, or to nothing. Figure 1 shows an associative network, a simple variant of MacWhinney's [12] and Merriman's [10] competition models, configured to learn which word names which object.

In this model, each time an object is presented, it activates the corresponding input node, which projects activation to the output nodes along weighted associative connections. When an object is presented together with a word, the connection between the object and the word is

Corresponding author: Terry Regier (regier@uchicago.edu).

http://tics.trends.com 1364-6613/03/$ - see front matter doi:10.1016/S1364-6613(03)00108-6

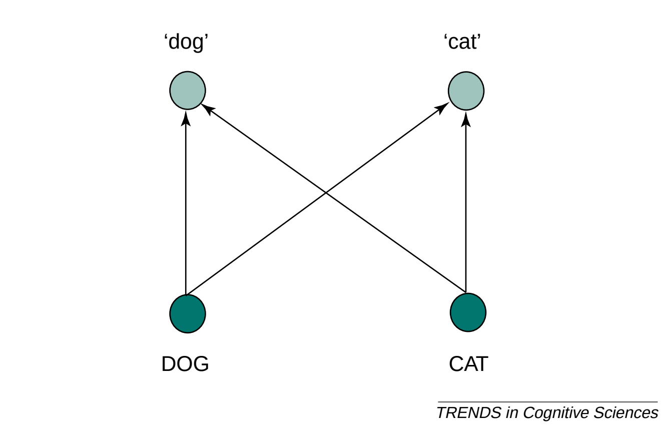

Fig. 1. An associative network, configured to learn the words 'dog' and 'cat'.

incremented. The activation for a word is the summed weighted activation from all object nodes, and the probability of a word being uttered, given an object as input, is given by normalizing:

$$p_i = \frac{a_i}{\left(\sum_j a_j\right) + n}$$

where a_i is the activation of word i, and n denotes noise in the system, and is here set to 0.3 (changing this parameter would yield different p_i values, but would not affect the qualitative behavior of the model). This equation captures the idea of competition between words: when one word is active, others become less likely to be produced. We initialize all weights to 0.1, and then expose the network to multiple instances of the object DOG paired with the word 'dog'. Under this training, the DOG–'dog' link grows stronger with time, and the probability of producing 'cat' as a name for the DOG decreases, as shown in Figure 2. This is a mutual exclusivity effect: the DOG already has a name ('dog'), which makes another name less likely. Merriman [10] has shown that this simple associative structure, with a competitive output rule, accounts for a variety of empirically observed mutual exclusivity effects.

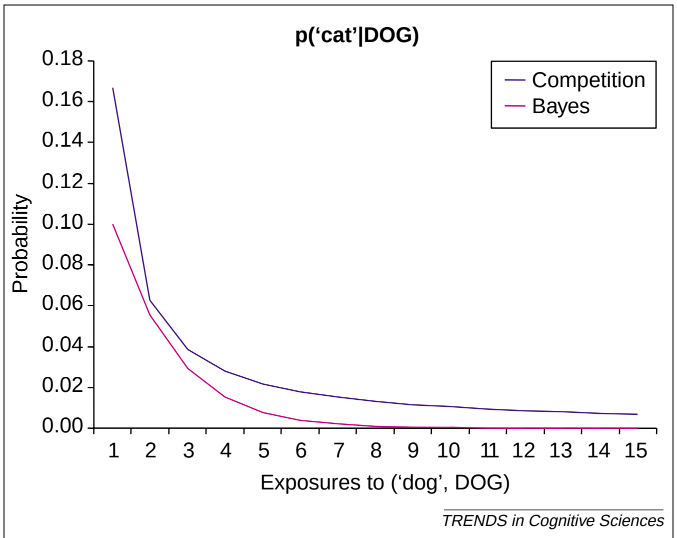

Fig. 2. Mutual exclusivity emerging from lexical competition in an associative competition model, and in a Bayesian model. In both models, the probability of using the word 'cat' to describe a DOG decreases with the number of times the DOG has instead been named a 'dog'.

Associative models have also accounted for many other aspects of word-learning, including the linking of sound and meaning [13–16], the perceptual grounding of meaning [11,17,18], language dysfunction [16,19], generalization patterns [17,20,21], and the effect of word co-occurrence patterns on word-learning [22,23]. Several of these models are discussed later in this review [11,13,17,22,23].

Bayesian inference

Bayesian models [24–26] are another genre of word-learning models built on general-purpose processes of induction. In addition, these models address the question of whether word-learning is rational, in the sense that it conforms to the normative standard of probabilistic inference given by Bayes' rule:

$$p(H/o) \propto p(o/H)p(H).$$

Here H is a hypothesis, and o is observed evidence. Tenenbaum and Xu's [24] Bayesian model of word-learning assumes that a single word – let's say 'cat' – is being learned. Each hypothesis H then corresponds to a possible meaning for 'cat', specified as a set of objects, for example {CAT} (the correct meaning), or {CAT or DOG} (an overly broad meaning). The learner must determine how probable each meaning is, given observations of objects being named 'cat'. At the core of this model is the assumption that each time a word is used, the particular object that it refers to was drawn randomly from the set of objects that constitutes the word's meaning. Thus, the likelihood of seeing 'cat' used to label some object o, on the hypothesis that the meaning of 'cat' is H, is:

$$p(o/H) = \begin{cases} \frac{1}{|H|} & \textit{if } o \in H \\ 0 & \textit{otherwise} \end{cases}$$

where $|H|$ denotes the number of objects in the set H. This simple idea can constrain word-learning: when the word 'cat' is applied to a CAT, this likelihood gives greater weight to narrow meanings like {CAT} than it does to broader meanings like {CAT or DOG}, because narrow meanings encompass fewer objects (|H| is smaller). This preference for narrow meanings is coupled with a prior p(H) that favors meanings that are dissimilar from others – on the intuition that more distinctive meanings are more deserving of a name. The model accounts well for adults' generalization behavior, when given 1–3 exposures to exemplars of a new word ('fast mapping' [27,28]). Niyogi extends this approach to account for the influence of syntax on the generalization of newly-learned words [26].

Interestingly, we can combine the insights of MacWhinney's competition model [12] and Tenenbaum and Xu's Bayesian model [24] to arrive at a Bayesian account of the mutual exclusivity principle. Thus, this word-learning constraint might not only be emergent, as we have already seen; it may also be rational, as shown in Box 1.

Box 1. Mutual exclusivity as rational inference

Consider a world of two words, 'dog' and 'cat', in which each word can mean either nothing, {DOG}, or {CAT}. Table I shows the resulting hypothesis space.

Table I. The hypothesis space for a model containing the two words 'dog' and 'cat' [a]

H	'dog'	'cat'	p(('dog',DOG)|H)
0	none	None	0
1	none	CAT	0
2	none	DOG	0
3	CAT	None	0
4	CAT	CAT	0
5	CAT	DOG	0
6	DOG	None	$1/(2\|N\|) = 1/(2 \times 1) = 1/2$
7	DOG	CAT	$1/(2\|N\|) = 1/(2 \times 1) = 1/2$
8	DOG	DOG	$1/(2\|N\|) = 1/(2 \times 2) = 1/4$

[a]For each hypothesis (H), the meaning of the words 'dog' and 'cat' are given, together with the likelihood of seeing a DOG labeled 'dog', given H.

We assume a uniform prior p(H) for simplicity, and a likelihood that captures random sampling of naming events from a hypothesis. (Tenenbaum and Xu [24] address the case of a single word applying to several referents; we apply their random selection principle to the inverse case of several words applying to one referent.) Specifically, to determine the likelihood of observing word *w* being used to name object *o* given hypothesis *H*, we first randomly select an object *o* from the world (either DOG or CAT, each with probability 1/2), and then randomly select a word from among those that can name *o* in *H*. Thus:

$$p((w,o)|H) = \begin{cases} \dfrac{1}{2|N|} & \text{if } w \text{ is a name for } o \text{ in } H \\ 0 & \text{otherwise} \end{cases}$$

where $|N|$ is the number of names available in *H* for object *o*. Critically, the likelihood of a word-object pairing is lower the more names a hypothesis specifies for the object; thus, hypotheses that violate mutual exclusivity have lower likelihoods. Given this prior and likelihood, we may use Bayes' rule to determine the probability of each hypothesis given a set of observations. From this, we can determine the conditional probability of using word *w* given object *o* and accumulated evidence *e*, through a weighted average across hypotheses:

$$p(w|o,e) = \sum_i p(w|o,H_i)p(H_i|e)$$

where $p(w|o,H)$ is also determined by randomly selecting a name for *o*: its value is $1/|N|$ if *w* is a name for *o* in *H*, and zero otherwise.

This model, like the competition model, exhibits a mutual exclusivity effect: as the model sees the DOG repeatedly labeled 'dog', the conditional probability of using 'cat' given the object DOG decreases (see Fig. 2 in main text). This constrains generalization, in a manner that is seen in children. In both models, the effect of mutual exclusivity emerges from competition [12] between words – in the output rule of the competition model, and in the random selection of a name in the Bayesian model.

Accelerating representations

In some word-learning models, early learning produces expectations that enable faster subsequent learning – which further strengthens the expectations, leading to yet faster learning. We may think of these expectations as 'accelerating representations': they permit a slow entry into word-learning to give way to accelerated learning as the expectations gradually become more accurate (cf. 'autonomous bootstrapping' [29]). This concept could help to explain the vocabulary spurt – a sometimes sudden increase in the rate and ease with which one-year-old children learn words. The vocabulary spurt is sometimes taken to suggest a conceptual insight into the referential nature of words [30,31] – but it might instead reflect the child's increasingly accurate expectations about the form and content of words.

Regier *et al.*'s associative word-learning model illustrates this idea [13]. In this model, word forms and word meanings reside in separate similarity spaces, and are linked through bi-directional associative connections.

At the heart of the model is the concept of memory interference: similar sounds (e.g. 'cat' and 'cad') interfere with each other in memory and are easily confused; the same is true of similar meanings (e.g. {CAT} and {DOG}). Early in learning, all words and meanings are somewhat similar, and the resulting interference hampers word-learning. However, this is soon alleviated. Word-learning drives selective attention preferentially toward dimensions of form that are predictive of meaning, and toward dimensions of meaning that are predictive of form. For instance, when learning object names, attention would be allocated to object shape, but not color – because shape, unlike color, is a good predictor of object name. (The model assumes that syntactic and other cues indicate which general class of names, e.g. object names, is being learned.) This attentional deployment effectively 'stretches' predictive dimensions, but 'compresses' non-predictive ones [10,32–34], as shown in Fig. 3. The result is that stimuli cluster together in categories that are psychologically distant from each other – and therefore no longer interfere with each other, enhancing future learning. An entirely analogous process causes word *forms* to cluster into categories corresponding to individual words, highlighting relevant differences (e.g. voicing) but downplaying irrelevant ones (e.g. pitch, in English).

Attention here is an accelerating representation. It is both driven by word-learning, and the enabler of increasingly fast subsequent learning [34], through a reduction in memory interference. This subsequent learning then further strengthens attention, which further accelerates learning. This simple dynamic provides a unified account of four roughly simultaneous changes in word-learning ability that children undergo during the second year of life [13]. First, they shift from slow learning, requiring 9–12 training trials to learn a word for an object [35,36], to one-trial learning, or 'fast mapping' [27,28] – accounted for by the reduction in interference. Second, young children can map two words to two different objects only if the words sound quite different, and not if they sound similar – whereas slightly older children can learn similar or dissimilar pairs of words [37,38]. Here, the explanation is that the initial interference from the early compressedness of space will be exacerbated by any similarity between the word forms, making similar-sounding words especially vulnerable; this extreme interference is eventually reduced through attentional stretching. Analogously,

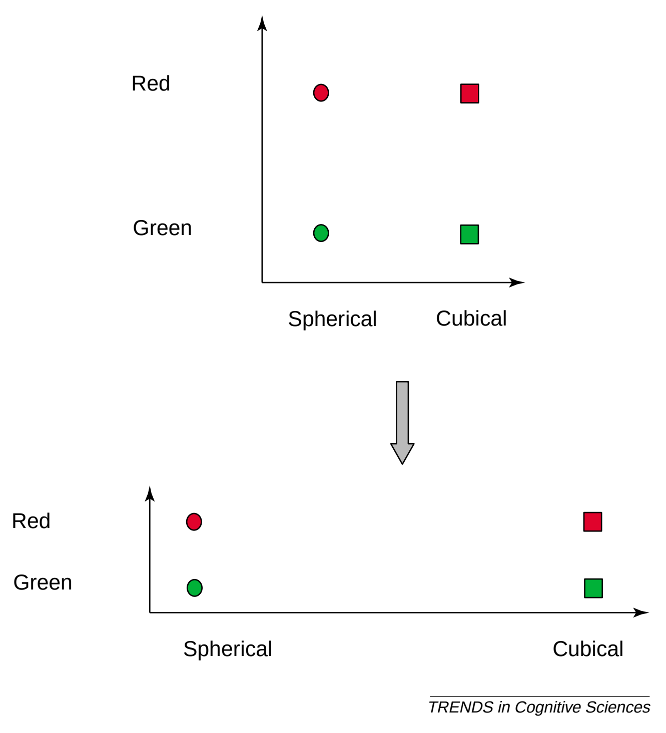

Fig. 3. Selective attention in word-learning. This psychological space has two dimensions – shape and color – and has four stimuli located in it: a red ball, a green ball, a red box, and a green box. The stimuli are shown before and after selective attention has been allocated toward the shape dimension and away from the color dimension. This allocation makes shape differences psychologically salient, and color differences less so, yielding the categories BALL and BOX, and facilitating the learning of new object names also based on shape.

this dynamic can account for children's early inability to learn synonyms, followed by a later ability to learn them [39]. Again, there is initially exacerbated interference, this time due to the semantic similarity (identity) between synonyms,and again, overall interference lessens as space stretches, allowing learning. In this case, however, the synonym-induced component of interference is unaffected by stretching, because the two meanings are the same, not just similar. This residual interference means that although synonyms can eventually be learned, there is a persisting resistance to them, as there is in children [2,4]. Finally, the increased attention to predictive dimensions also accounts for children's growing tendency to generalize new object nouns on the basis of object shape [36,40].

Another example of accelerating representations can be found in Siskind's [41] word-learning model. This model, which learns to link words with symbolically-represented conceptual structures (see also [42]), exhibits slow initial learning, requiring several exposures before a word is learned, followed by later one-trial learning. As words are learned for some concepts, that leaves fewer concepts unnamed, yielding a smaller hypothesis space – and a simpler inductive problem – for later word-learning. Plunkett *et al.*'s connectionist model of word-learning [17] also exhibits an acceleration in learning, owing to the development of internal representations that capture widely shared patterns of variation in the input. Yet another example is found in Bayesian models: the posterior probability after one observation is often taken as the prior probability for the next observation – this can also be seen as an accelerating representation.

Grounding meaning in the world and in words

Several computational models ground words in perceptual representations of objects and events in the world [11,17,18,43–45]. Most others ground word meaning in more abstract featural representations, but still on the assumption that there is some concrete element of experience to which the word is being linked.

However, much word-learning does not occur in this fashion – people eventually learn words for things that are not grounded in their personal experience at all (e.g. 'prehistoric'). Instead, we acquire much of our knowledge of word meanings through reading, inferring the meanings of new words from their textual context. Landauer and Dumais [22] (see also Burgess and Lund [46]) present a model of this process. Their approach, latent semantic analysis (LSA), induces knowledge of the meanings of English words given only a large corpus of English text. The central concept is that the more similar two words are in meaning, the more likely they are to appear in the company of the same set of other words, across different chunks of text. The input to LSA is a matrix in which each row represents an English word, each column represents a chunk of text (e.g. a paragraph), and the cell holds the frequency with which that word occurs in that chunk. LSA then re-represents these co-occurrence data in a space of reduced dimensionality. Each word is represented as a vector in this space, indirectly capturing the textual contexts in which the word can be used. Semantic similarity between two words is measured by the angle between their vectors. LSA was trained using 4.6 million words of text taken from an encyclopedia. Given a standardized test of synonyms, it then performed comparably to foreign-born applicants to US colleges.

More importantly, however, it suggested a solution to a core puzzle in language induction, concerning the rate at which words are learned. Schoolchildren, in reading, add one new word to their comprehension vocabularies every five paragraphs or so – an impressive rate. LSA roughly matched this rate of acquisition, and did so in an illuminating manner. For the model, a paragraph of text had a greater effect on the learning of words that were *not* in that paragraph than on those that were. Thus, if people learn in an analogous fashion, the high speed of vocabulary acquisition through reading can be explained in part through the effects of indirect learning. The general picture that emerges from LSA is one of multiple weak constraints among the meanings of different words, such that information concerning one word can effect the entire web of interrelated meanings.

Syntactic context is another potentially powerful clue to meaning, and is also available in the surface form of text. Some models [23,26] exploit this fact, relying on syntactic cues to set semantic expectations concerning the meanings of words that are being learned.

Constraints and semantic universals

Word-learning constraints are a possible source of cross-linguistic semantic universals. For if words are learned in the same constrained manner across languages, the meanings of words in different languages should bear some mark of the constraints that produced them.

Regier's connectionist model of spatial term learning [11] illustrates this idea. The model learns to categorize spatial events according to the spatial system of a given target language. Because languages differ in their spatial categorization Scheme, the model must accommodate a variety of linguistic structurings of space.

The search for a simple model led to one with a constraint on its operation: fine-grained spatial distinctions are learned more reliably at event *endpoints* than at event beginnings. This predicts a semantic universal tendency: across languages, spatial terms describing motion into some spatial configuration should tend to be more semantically specific than words describing motion out of that configuration – because the configuration is present at endpoint only for the 'motion-in' term. An English example is the term 'on top of'. This spatial term is semantically finer than its opposite 'off' – because 'off' can refer to removal from the top, side, or bottom of an object (e.g. 'take it off the wall').

There is empirical support for this account. Zheng and Regier (in preparation) have found that adults perceive spatial distinctions more clearly at event endpoints than at event beginnings. They also found that adult native speakers of each of English, Chinese and Japanese, who described a set of simple spatial actions, tended to use semantically narrower spatial terms for motion-in events than for analogous motion-out events, as predicted. Finally, Bowerman has found that children learning different languages tend to overgeneralize words for motion-out more than they do words for motion-in [47]. These results highlight the potential for computational models of constrained word-learning to explain regularities in semantic structure across languages, not just within them.

Objections and limitations

Many of the models discussed in this review assume an associative basis of some sort for word-learning. This basic assumption is one that has encountered two broad sorts of objection.

The first objection is that word-learning is too fast to be a reflection of an associative or statistical process [8,30,31]. As we have seen, children can eventually learn a new word given only a very few exposures. But is this really a problem? It is true that one often thinks of associative learning as requiring multiple exposures and gradual learning, but we have seen that one-trial learning can be exhibited by associative [13] as well as probabilistic [24] word-learning models. Thus speed in and of itself is not a strong argument against associative – or more generally, statistical – learning.

A more pointed objection is that in learning words, children do not passively link words with whatever they see when the words are spoken [8,48]. Instead, they appear to actively impose social interpretations on the behavior of other people, and to rely on their knowledge of others' mental states to guide their word-learning [49,50]. A related objection is that non-human animals, who might lack some of these social abilities but are certainly capable of associative learning, do not learn words [50,51].

This objection is compelling, and constitutes a strong argument for the centrality of social understanding in word-learning (see also Questions for future research). However, it does not argue against the reasonable idea that social cues and associative learning can operate together, and that word-learning might emerge from the collaboration of species-specific social sensitivities and widely-shared general learning mechanisms. Given the simplicity, independent motivation, and explanatory power of associative models, this idea seems well worth pursuing. One intriguing possibility is that social cues might act in part to filter out irrelevant aspects of experience, and to focus the child's attention on a word on the one hand, and a referent on the other, so that these two are linked through association.

Box 2. Questions for future research

- Word use is a fundamentally social activity, involving reference and an awareness of the mental state of one's conversation partner. How is this fact most appropriately brought to bear on computational models of word-learning?
- Is word-learning usefully thought of as a form of categorization? Of inference? How much of word-learning can be explained in these terms?
- How does the phonological form of words interact with semantic forces in learning the link between form and meaning?
- What neuropsychological constraints can inform models of word-learning?

Conclusions

Word-learning is generally thought to be an underdetermined inductive problem, such that children require a set of constraints to tackle it. This view has been bolstered by the considerable empirical evidence for such constraints. However, several recent computational models of word-learning suggest that these constraints need not spring from language-specific forces. General-purpose learning mechanisms, accelerating representations, and perceptual and textual forces might all combine to constrain word-learning, and cross-linguistic variation in meaning. As computational models of word-learning continue to draw on constraints that spring from a variety of such sources, including social cues, they will help to explain how the very center of language – the link between form and meaning – builds on the rest of cognition.

Acknowledgements

This work was supported by NIH grant DC03384. I thank Susanne Gahl and the anonymous reviewers for helpful comments on an earlier draft of this paper.

References

1 Quine, W. (1960) *Word and Object*, MIT Press
2 Markman, E. (1989) *Categorization and Naming in Children: Problems of Induction*, MIT Press
3 Golinkoff, R. *et al.* (1992) Young children and adults use lexical principles to learn new nouns. *Dev. Psychol.* 28, 99–108

4 Merriman, W. and Bowman, L. (1989) The mutual exclusivity bias in children's word learning. *Monogr. Soc. Res. Child Dev. No. 54*
5 Mervis, C.B. and Bertrand, J. (1994) Acquisition of the novel name-nameless category (N3C) principle. *Child Dev.* 65, 1646–1662
6 Chomsky, N. (1986) *Knowledge of Language: Its Nature, Origin and Use*, Praeger
7 Markson, L. and Bloom, P. (1997) Evidence against a dedicated system for word-learning in children. *Nature* 385, 813–815
8 Bloom, P. (2000) *How Children Learn the Meanings of Words*, MIT Press
9 MacWhinney, B. ed. (1999) *The Emergence of Language* Erlbaum
10 Merriman, W. (1999) Competition, attention, and young children's lexical processing. In *The Emergence of Language* (MacWhinney, B., ed.), pp. 331–358, Erlbaum
11 Regier, T. (1996) *The Human Semantic Potential: Spatial Language and Constrained Connectionism*, MIT Press
12 MacWhinney, B. (1989) Competition and lexical categorization. In *Linguistic Categorization* (Corrigan, R. *et al.*, eds), pp. 195–242, John Benjamins
13 Regier, T. *et al.* (2001) The emergence of words. In *Proc. 23rd Annu. Conf. Cogn. Sci. Soc.* (Moore, J. and Stenning, K., eds) pp. 815–820, Erlbaum
14 Cottrell, G. and Plunkett, K. (1994) Acquiring the mapping from meaning to sounds. *Connect. Sci.* 6, 379–412
15 Gupta, P. and MacWhinney, B. (1997) Vocabulary acquisition and verbal short-term memory: computational and neural bases. *Brain Lang.* 59, 267–333
16 Miikkulainen, R. (1997) Dyslexic and category-specific aphasic impairments in a self-organizing feature map model of the lexicon. *Brain Lang.* 59, 334–366
17 Plunkett, K. *et al.* (1992) Symbol grounding or the emergence of symbols? Vocabulary growth in children and a connectionist net. *Connect. Sci.* 4, 293–312
18 Nenov, V. and Dyer, M. (1994) Perceptually grounded language learning: II. Dete: a neural/procedural model. *Connect. Sci.* 6, 3–41
19 Plaut, D. *et al.* (1996) Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychol. Rev.* 103, 56–115
20 Schyns, P. (1991) A modular neural network model of concept acquisition. *Cogn. Sci.* 15, 461–508
21 Li, P. and MacWhinney, B. (1996) Cryptotype, overgeneralization and competition: a connectionist model of the learning of english reversive prefixes. *Connect. Sci.* 8, 3–30
22 Landauer, T. and Dumais, S. (1997) A solution to Plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychol. Rev.* 104, 211–240
23 Gasser, M. and Smith, L.B. (1998) Learning nouns and adjectives: a connectionist account. *Lang. Cogn. Proc.* 13, 269–306
24 Tenenbaum, J. and Xu, F. (2000) Word learning as Bayesian inference. In *Proc. 22nd Annu. Conf. Cogn. Sci. Soc.* (Gleitman, L. and Joshi, A., eds) pp. 517–522, Erlbaum
25 Tenenbaum, J. and Griffiths, T. (2001) Generalization, similarity and Bayesian inference. *Behav. Brain Sci.* 24, 629–640
26 Niyogi, S. (2002) Bayesian learning at the syntax-semantics interface. In *Proc. 24th Annu. Conf. Cogn. Sci. Soc.* (Gray, W. and Schunn, C., eds) pp. 697–702, Erlbaum
27 Behrend, D. *et al.* (2001) Beyond fast mapping: young children's extensions of novel words and novel facts. *Dev. Psychol.* 37, 698–705
28 Carey, S. (1978) The child as word learner. In *Linguistic Theory and Psychological Reality* (Halle, M. *et al.*, eds), pp. 264–293, MIT Press
29 Brent, M.R. (1996) Advances in the computational study of language acquisition. *Cognition* 61, 1–38
30 Lock, A. (1980) *The Guided Reinvention of Language*, Academic Press
31 McShane, J. (1979) The development of naming. *Linguistics* 17, 879–905
32 Nosofsky, R. (1986) Attention, similarity, and the identification-categorization relationship. *J. Exp. Psychol. Gen.* 115, 39–57
33 Kruschke, J. (1992) ALCOVE: an exemplar-based connectionist model of category learning. *Psychol. Rev.* 99, 22–44
34 Smith, L. (2000) Learning how to learn words: an associative crane. In *Becoming a Word Learner: A Debate on Lexical Acquisition* (Golinkoff, R. *et al.*, eds), pp. 51–80, Oxford University Press
35 Schafer, G. and Plunkett, K. (1998) Rapid word learning by fifteen-month-olds under tightly controlled conditions. *Child Dev.* 69, 309–320
36 Woodward, A. *et al.* (1994) Rapid word learning in 13- and 18-month-olds. *Dev. Psychol.* 30, 553–566
37 Stager, C. and Werker, J. (1997) Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature* 388, 381–382
38 Werker, J. *et al.* (2002) Infants' ability to learn phonetically similar words: effects of age and vocabulary size. *Infancy* 3, 1–30
39 Liittschwager, J. and Markman, E. (1994) Sixteen and 24-month-olds' use of mutual exclusivity as a default assumption in second label learning. *Dev. Psychol.* 30, 955–968
40 Smith, L. *et al.* (2002) Object name learning provides on-the-job training for attention. *Psychol. Sci.* 13, 13–19
41 Siskind, J. (1996) A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition* 61, 39–91
42 Suppes, P. *et al.* (1995) Comprehension grammars generated from machine learning of natural languages. *Mach. Learn.* 19, 133–152
43 Roy, D. and Pentland, A. (2002) Learning words from sights and sounds: a computational model. *Cogn. Sci.* 26, 113–146
44 Siskind, J. (1994) Grounding language in perception. *Artif. Intell. Rev.* 8, 371–391
45 Regier, T. and Carlson, L. (2001) Grounding spatial language in perception: an empirical and computational investigation. *J. Exp. Psychol. Gen.* 130, 273–298
46 Burgess, C. and Lund, K. (1997) Modelling parsing constraints with high- dimensional context space. *Lang. Cogn. Proc.* 12, 177–210
47 Bowerman, M. (1996) Learning how to structure space for language: a crosslinguistic perspective. In *Language and Space* (Bloom, P. *et al.*, eds), pp. 385–436, MIT Press
48 Sabbagh, M. and Gelman, S. (2000) Buzzsaws and blueprints: what children need (or don't need) to learn language. *J. Child Lang.* 27, 715–726
49 Baldwin, D. *et al.* (1996) Infants' reliance on a social criterion for establishing word-object relations. *Child Dev.* 67, 3135–3153
50 Tomasello, M. (1999) *The Cultural Origins of Human Cognition*, Harvard University Press
51 Tomasello, M. and Akhtar, N. (2000) Five questions for any theory of word learning. In *Becoming a Word Learner: A Debate on Lexical Acquisition* (Golinkoff, R. *et al.*, eds), pp. 179–186, Oxford University Press