To appear in:

*Encyclopedia of Language and Linguistics – 2nd Edition.* (Keith Brown, Editor). Oxford: Elsevier. 2005.

# Computational approaches to language acquisition

Jeffrey L. Elman
Department of Cognitive Science
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093-0515

**ABSTRACT**

Computational models have become increasingly important in the field of child language acquisition over the past several decades. This reflects not only the increased accessibility of hardware and software, but more importantly, significant advances in the field of machine learning that have led to more sophisticated understanding about biologically plausible learning mechanisms. Computational models have been useful as tools to make verbal theories more precise, to explore complex interactions and emergent phenomena, and to generate predictions that lead to new empirical research. Issues that have been the target of modeling include so-called "burst" phenomena; U-shaped curves in learning; discovery of words, category, and grammatical structure; alternative to rule-based accounts; and the Poverty of the Stimulus hypothesis.

## WHY COMPUTATIONAL MODELS?

One of the striking developments in the field of child language acquisition within the past two decades has been the dramatic increase in the use of computational models as a way of understanding the acquisition process. In part, this has been driven by the wide-spread availability of inexpensive but powerful computers and the development of software that has made modeling more widely accessible than in previous periods. But there is a more interesting and scientifically significant reason for the phenomena. Computational theories of learning themselves have matured significantly since the middle of the $20^{th}$ century. The renaissance in neural network (or connectionist) approaches, and more recently, the linkages with mathematical approaches such as Bayesian inference, information theory, and statistical learning have provided a much more sophisticated perspective on a number of issues relevant to language acquisition.

These models are necessarily used in conjunction with empirical approaches, but provide an important complement to such approaches. At the very least, computational models can be thought of as enforcing a level of detail and specificity in a theory or account that a verbal description might not possess. Furthermore, even in simple models, there may be interactions among the model's components that are sufficiently complex that only through empirically running a simulation is it possible to know how the model will behave. Computer models also afford the opportunity to explore aspects of a phenomenon that may not be easily tested in the real world (either because the corresponding situation has not yet been studied, or perhaps may be infeasible to test). By systematically exploring the full parameter space of a theory one can sometimes gain insight into the deeper principles that underlie a behavior. And of course, a model may be amenable to analytic techniques that are not practical with real children. With children we can never do more than make inferences about the internal mechanisms that drive a behavior. Computer models, on the other hand, can in principle be complete understood. Finally, such models can serve as hypothesis generators. They often suggest novel ways of understanding a phenomenon. Of course, the validity of the hypothesis ultimately depends on empirical testing with real children.

In general, there have emerged two complementary approaches to modeling. In the first, the goal is determine that a problem can be solved without making specific claims that the solution implemented in the computer model is the same as in the child. These approaches tend to be more mathematical in nature. Work involving Bayesian inference, information theory, and statistical learning are of this sort. The second approach attempts to model the acquisition process a bit more directly. Learning plays a central role in these approaches, and the models' behavior at intermediate stages is as much a focus as the ability to ultimately master the task. Connectionist models are examples of this second sort of model.

Because of the field of computational approaches to language acquisition has grown so explosively in recent years—and is beyond the scope of the present review—what follows will be organized around the major issues that have been addressed (leaving aside a large number of interesting but less central phenomena). For excellent discussion of related computational approaches, see Brent (1996), Munakata and McClelland (2003) and MacWhinney (1998, 1999).

## ISSUES AND RESULTS

It is useful to focus on the modeling work in terms of two major questions that have been addressed (bearing in mind the caveat above that, just as the field of language acquisition is itself large and diverse, there are many models that fall outside the scope of this taxonomy). These questions have to do with: (1) Oddities in the "shape of change" (e.g., discontinuities or nonlinearities in acquisition, as in U-shaped curves); (2) What information is available in the input a child receives, and what can she infer from it (e.g., the problem of segmenting words or discovering grammatical categories or syntactic regularities). How can learning proceed in the face of putatively insufficient information (e.g., "Baker's paradox" or the so-called "Poverty of the stimulus" problem)?

We shall consider each of these in detail, considering first what the issues are and then the computational models that have endeavored to understand the phenomena.

### Explaining the shape of change

The simplest and possibly most natural pattern of development would probably be to assume a linear increase in performance over time. Such a pattern would be consistent with the assumption that the mechanisms that subserve learning remain relatively constant, and thus the increase in what is learned over every time increment should also remain constant. In fact, very few developmental patterns illustrate such linear tendencies. Development seems to proceed in fits and spurts, sometimes interrupted by long periods where little appears to change and sometimes even by phases where performance temporarily deteriorates.

Noteworthy examples of such nonlinearities abound in the realm of language acquisition, and have played a major role in theorizing about the mechanisms that make language acquisition possible. The special ability of children to learning languages (the Critical Period) is a notable example of such a non-linearity. One influential explanation of this effect is that it reflects the existence of a specialized neural mechanism, the Language Acquisition Device, that is operative only during childhood (Chomsky, 1957; Johnson & Newport, 1989).

Another well-documented set of nonlinearities is exemplified by the rapid increases in word comprehension, production, and knowledge of grammar that occur in young children during their second year of life (as in Figure 1, from Bates & Goodman, 1997). Clearly, something dramatic seems to be happening at the point where, for example, the child manifests a burst in the rate at which she learns new words. Many theorists have interpreted such bursts as evidence that something new has appeared in the child (e.g., Golinkoff, Mervis, & Hirsh-Pasek, 1994; Markman, 1989; McShane, 1979).

*<Figure 1 near here>*

A final example has played a particularly important role in the theoretical literature: the apparent U-shaped curve that characterizes children's mastery of the past tense of the English verbal system. At the earliest stage, children know a small number of verbs, mostly of very high frequency and tending to be irregular; they typically produce the past tense correctly. At the second stage, the number of verbs in the child's productive vocabulary increases and includes a larger number of regulars, some of which may be lower in frequency. At this stage, both observational evidence (over-generalization of the "+ed" pattern for regular verbs) and experimental data (ability to generate the regular version of nonce verbs) suggest that the child has learned a rule for forming the past tense. During the second stage, the rule is incompletely learned and misapplied to many (previously correctly produced) irregulars, resulting in a decline in overall performance. Finally, at the third stage, the correct forms for both regulars and irregulars are produced and the child appears to have learned not only the rule—that applies to regulars—but the exceptions. These data have provided a powerful argument in favor of the psychological reality of rules.

## The critical period

A number of computational models have addressed these issues and in many cases provided alternative hypotheses for the phenomena. In attempting to understand how neural networks might deal with complex recursive structure in language, Elman (1993) discovered that the network was able to process complex sentences only when it began either by initially exposed to simple sentences (a kind of neural network "motherese"), or if it began the learning process with a restricted working memory (similar to the limited WM found in young children, e.g., Santelmann & Jusczyk, 1998)). Elman called this the "starting small" effect. It is similar in spirit to Newport's "less is more" hypothesis (Newport, 1990). In both accounts, the limitation on processing resources acts like a filter that temporarily hides the more complex aspects of language from the network (or child). Learning the simpler phenomena first creates a foundation of knowledge that makes it possible to subsequently learn more complex regularities. These accounts suggest that rather than being enabled by a special mechanism (the LAD) that is lost in adulthood, the explanation for the Critical Period is that—paradoxically—it is maturational limitations that facilitate the learning of language. However, it is also possible that there are multiple factors that result in Critical Period effects. Using a model based on Hebbian learning (a computational paradigm closely related to the changes in Long Term Potentiation of synaptic junctions that results from synchronous firing of neurons) Munakata and Pfaffly (2004) demonstrated that even though the mechanism for plasticity did not change, what was learned early in a network's life constrained what it could learn later. Marchman (1993) has demonstrated similar effects in networks that learn the past tense; networks that suffer simulated brain damage early in life recover much better compared to networks that are lesioned after much learning has occurred.

## The vocabulary burst

A number of models have attempted to understand what factors might lead to the rapid acceleration in learning of new words that typically occurs in the middle of the second year of life. Plunkett, Sinha, Moller, and Strandsby (1992) trained networks to associate linguistic labels with visual images and observed that a burst-like increase in ability to learn labels occurred after early training. They also found that comprehension performance in the networks always exceeded (and preceded) production, that the networks exhibited prototype effects, and that they also show under- and over-extension phenomena found in children. Plunkett et al. attribute these behaviors to the network's need to develop robust conceptual categories. Prior to this time, learning is slow and errorful. Once categories are learned, they facilitate the learning of new words. A similar effect was found in Elman (1998), who also found that there was a direct causal connection between vocabulary growth and the later emergence of grammar (cf. Bates & Goodman, 1997). The effect arose because essentially grammar was understood as a generalization over the commonalities in syntactic behavior of many words; with a small vocabulary these patterns are not evident, and so vocabulary growth is a necessary prerequisite to discovery of grammatical patterns (cf. Tomasello, 2000 for a similar account in the acquisition literature).

## The English past tense: Rules or connections?

The final example of nonlinearities in language acquisition is the U-shaped performance shown in English by many children as they learn the correct past tense forms of verbs. This phenomenon has long been seen as demonstrating the psychological reality of rules, insofar as we appear to be observing the moment in time when the rule for the past tense is being acquired (Berko, 1958; Pinker, 1991). Rumelhart and McClelland (1986) challenged this assumption by showing that when a neural network was trained, on an verb by verb basis, to produce the past tense of English verbs, it not only manifest a similar U-shaped performance curve, but also replicated in detail many of the more specific empirical phenomena found in children. Rumelhart and McClelland suggested that the network account provided an alternative to the traditional interpretation involving rules. Not surprisingly, this claim provoked a controversy that continues to this day (Prince & Pinker, 1988). The debate has been lively, if at times acrimonious. And although the theoretical interpretation remains controversial, One of the most important outcomes of this debate has been the broadening—in terms both of languages studied and level of detail—of empirical research in English but also other languages including German, Hebrew, Icelandic, Italian, Norwegian, Polish, Spanish (Behrens, 1998; Clahsen, Aveledo, & Roca, 2002; Dabrowska, 2001; Dromi, Leonard, Adam, & Zadunaisky-Ehrlich, 1999; Eddington, 2002; Marchman, Plunkett, & Goodman, 1997; Marcus et al., 1992; Orsolini, Fanari, & Bowles, 1998; Ragnarsdottir, Simonsen, & Plunkett, 1999; Wittek & Tomasello, 2002; Xu & Pinker, 1995). This is an excellent example of ways in which computational models can refine the questions that are addressed and encourage new avenues of empirical investigation. The debate has also led to a more sophisticated understanding of the implications of the competing accounts not only for acquisition but other aspects of language processing and historical change (Hare & Elman, 1995; Marcus et al., 1992; McClelland & Patterson, 2002; Patterson, Lambon Ralph, Hodges, & McClelland, 2001; Plunkett & Juola, 1999; Plunkett & Marchman, 1993; Plunkett & Nakisa, 1997; Ragnarsdottir et al., 1999; Seidenberg & Bruck, 1990; Seidenberg & Hoeffner, 1998).

## What information is available to a child, and what can be learned?

Although obviously a child’s experience places a critical role in the learning process, the relationship between what the child hears and what she ultimately knows is in many cases not transparent. Indeed, it has been claimed that in some cases, there is no evidence at all for this knowledge (Crain, 1991; Pinker, 1984). The putative insufficiency of the evidence available to a child—often called the Poverty of the Stimulus problem—has led to the conclusion that significant amounts of linguistic knowledge must be “pre-known” by a child. This knowledge constitutes a Universal Grammar that is part of the biological endowment every child begins with as she begins the task of learning the specific features of her own language.

There are two three issues that must be considered when evaluating such a hypothesis. The first involves what the actual input is that is available to children. Although that input is in fact massive in terms of word tokens, there is now reason to believe that it reflects a restricted range of the adult language (Cameron-Faulkner, Lieven, & Tomasello, 2003). Second, it is also clear that for a long period of time, children are actually much more conservative in their productions and stick closely to what they hear (Cameron-Faulkner et al., 2003; Lieven, Behrens, Speares, & Tomasello, 2003; Theakston, Lieven, & Tomasello, 2003; Tomasello, 2003). Nonetheless, it is also obviously true that at some point children venture into uncharted territory, so the problem of what motivates such creative use of language is real. This leads to the third question, which is what theory of learning is assumed. At least some of the Nativist accounts have assumed a very weak kind of learning, essentially little more than a mental tabulation of utterances (e.g., Pinker, 1984, p. 49ff). Computational models have been most successful in addressing this third question, by exploring the properties of more powerful—but hopefully psychologically plausible—learning mechanisms.

## Discovering where the words are: The segmentation problem

Unlike written language, in which words are delimited by white space or punctuation characters, spoken language yields few explicit clues as to where the boundaries between words are. For the infant, this poses a serious challenge, complicated by the fact that even the definition of what counts as a word differs dramatically across languages. How does the child thus learn (a) what can serve as a word? And (b) where the words are in continuous speech?

A number of computation approaches have converged in a similar insight, which is that at least to a first approximation, sequences of sounds that are highly associated are good candidates to be words. The manner in which this hypothesis is implemented varies (Brent & Cartwright, 1996; Christiansen, Allen, & Seidenberg, 1998; Elman, 1990) but the essential idea is that word boundaries are thus locations where the conditional probability of the next sound, given what has preceded it, is low. This can be seen in Figure 2, which shows the errors made by a network that has learned to predict the next letter in a sequence of words (white space removed) that make up a child's story (Elman, 1990). Error tends to be greatest at the onsets of words, and decreases as more of a word is heard. Error maxima thus constitute likely word boundaries.

*<Figure 2 near here>*

Another issue that concerns word learning is the problem of determining the syntactic and semantic categories of words. Here again, strong claims have been advanced that at least the categories much be innate, as well as innate principles that guide the child in make such determinations. The arguments have included the claim that the kind of distributional information available to a child (e.g., words in the same category tend to have similar distributional properties`; Harris, 1954; Maratsos & Chalkley, 1980) will fail given the complexity of real language input. However, a number of computational models have suggested otherwise. Considerably more information of this sort appears to be available to a child than might be imagined (Cartwright & Brent, 1997; Elman, 1995; Mintz, 2002; Redington, Chater, & Finch, 1998). Again, these models differ in their details, but share the same insight that a word's privilege of occurrence is a powerful indicator of its category. Importantly, there is an increasingly empirical literature involving learning of artificial languages by infants and young children that is highly consistent with the type of learning embodied in the computational models (see Gomez & Gerken, 2000; Saffran, 2001, for a discussion of this work).

## Discovering grammar? The poverty of the stimulus problem

Perhaps the strongest claims regarding what the necessity for children's innate linguistic knowledge arise in the context of grammar. As with the past tense debate, the controversy has been heated. It has also been complex, since it interacts not only with the long-standing Nature vs. Nurture debate but also questions regarding the extent to which linguistic knowledge is modular and independent from other cognitive processes (i.e., domain-specific), and whether the uniqueness of language to our species also reflects specialized neural—and presumably, also genetic—substrates that are entirely absent in other species. For two very different answers to these questions, see (Elman, Bates, Johnson, Karmiloff-Smith, & et al., 1996) and (Pinker, 1994).

One basic question that arose early in the discussion is whether connectionist models were capable at all of capturing some of the apparently recursive nature of natural language (Fodor & Pylyshyn, 1988). Even if recursion in human language is only partial, there is good evidence that some kind of abstract internal representations must underlie the surface forms. Symbolic accounts that make use of syntactic trees provide one mechanism that might explain why, for example, the verb *is* in (1) is in the singular, agreeing with *woman,* rather than any of the other nouns in the sentence.

(1) The woman who Mary and Bob introduced me to last summer while I was visiting them in Paris on my way to Prague is really quite interesting.

Similarly, tree-structured representations provide a formalism that makes possible hypotheses about why is (2) an acceptable sentence, whereas (3)—which is very similar in meaning—is ungrammatical. (It should be noted, however, that accounts of such differences are elusive and there is still not complete agreement within any framework about the explanation for these sorts of differences.)

(2) Who did you believe Annie saw? (Possible answer: I believed Annie saw Elvis.)

(3) *Who did you believe the claim Annie saw? (Possible answer: I believed the claim Annie saw Elvis.)

Claims that neural networks were in principle unable to deal with such linguistic complexities may be premature (Boden & Blair, 2003; Elman, 1991; Rodriguez, Wiles, & Elman, 1999). Their solution to the problem of recursion differs from classical discrete automata, but recurrent neural networks definitely have sufficient power to deal with complex grammatical constructions

(Siegelmann, 1995). More relevant to issues in language acquisition is whether these complex grammatical regularities can actually be learned, given the input to which a child might be exposed.

A number of computational models suggest a positive answer (e.g., Christiansen & Chater, 1999; Elman, 1991; Elman, 1993). One particularly challenging problem, and the one we will conclude with, was posed by Crain (Crain, 1991) and concerns what has been called Aux Inversion as a hypothesis to explain how certain kinds of questions are formed.

Crain argued that based on the evidence available to a child, such as question-answer pairs shown in (4) and (5), any account of grammar acquisition that relies solely on learning would be expected to produce the incorrect generalization that, for any declarative, the corresponding interrogative involves inversion of the first verb and first noun, as captured schematically by the rule shown in (6).

(4) a. Mary is happy.
b. Is Mary happy?
(5) a. Timmy can swim awfully fast.
b. Can Timmy swim awfully fast?
(6) *Declarative*: **Noun AUX . . .**
*Interrogative*: **AUX Noun . . .**

But this rule would be wrong, because it predicts incorrectly that the interrogative that corresponds to (7a) would be (7b). In reality, the correct interrogative is (7c) (For convenience, underlining shows the location of the auxiliary prior to inversion.)

(7) a, The boy who is smoking is crazy.
b. *Is the boy who __ smoking is crazy?
c. Is the boy who is smoking __ crazy?

Crain argues that children do not hear the sort of data (e.g., questions of the form in (7c)) until well past the period where they can be shown—by experimentally eliciting them—to correctly produce these forms. He concludes that this is strong evidence for the existence of an innate constraint that makes requires that abstract constituent structure be the basis for learning grammatical regularities. He calls this the “parade case” for an innate constraint.

To test this claim, Lewis and Elman (2001) constructed a simulation in which a recurrent neural network was trained on examples of well-formed sentences; the training data were generated to mimic the types and frequencies of sentences found in the Manchester corpora from the CHILDES databank (MacWhinney, 2000; Theakston, Lieven, Pine, & Rowland, 2001). Crucially, although there were many sentences of the form shown in (4) and (5), no sentences of the forms shown in (7) were included. The network was then tested on both ungrammatical (7b) and grammatical (7c) inputs. Its clear preference was for the grammatical questions, despite never having seen similar sentences during training.

How did the network learn the true grammatical generalization? It turns out that there are many other sentences present in the input (to children as well as these networks) that provide ample evidence for the fact that noun phrases act as constituents. These include sentences such as those shown in (9).

(9) (a) The bike with wheels belongs to me.
(*Not:* The bike with wheels *belong to me.)

(b) The cats my dog chases belong to our neighbor.
(*Not*. The cats my dog chases *belongs to our neighbor.)

The input to the network is thus sufficient to motivate a number of generalizations. These involve learning about different grammatical categories (nouns, verbs, prepositions, complementizers, etc.); selectional restrictions imposed by verbs on their arguments; the form of simple declaratives; the form of simple interrogatives; and the fact that agreement relations (among others) must respect constituenthood. Although these are logically independent generalizations, they have the opportunity to interact. The critical interaction occurs when a complex sentence is also an interrogative. The network has never seen such interactions, but its ability to partial out independent generalizations also makes it possible to combine generalizations as they may interact.

There is an important lesson here, and it is a clear demonstration of the ways in which computational models—particularly those that involve learning—can yield new insights into old problems. To a large degree, the question of what can be learned from the available input hinges crucially on what counts as input. Many of the claims regarding Poverty of the Stimulus have taken a straightforward and literal view of the input. If the target generalization to be learned involves strings of the form *X*, then the relevant input consists of strings of the form *X*. But this is a very limited view of the relationship between our experience and what we make of it. The Lewis and Elman simulation suggests that some of the more complex aspects of language learning may involve a good deal of what is really indirect evidence, and that inductive mechanisms of the sort instantiated in neural networks are capable of combining that evidence in novel ways to yield outcomes that are not transparently related to the input. Whether this is in fact also true of children of course can only be determined through empirical research. The importance of the computational simulations is that they open up a logical possibility that previously had been ruled out.

## REFERENCES

Bates, E., & Goodman, J. C. (1997). On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia, and real-time processing. *Language and Cognitive Processes, 12*, 507-584.

Behrens, H. (1998). How difficult are complex verbs? Evidence from German, Dutch and English. *Linguistics, 36*(4), 679-712.

Berko, J. (1958). The child's learning of English morphology. *Word, 14*, 150-177.

Boden, M., & Blair, A. (2003). Learning the dynamics of embedded clauses. *Applied Intelligence, 19*(1-2), 51-63.

Brent, M., & Cartwright, T. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition, 61*(1-2), 93-125.

Brent, M. R. (1996). Advances in the computational study of language acquisition. *Cognition. Special Issue: Compositional language acquisition, 61*(1-2), 1-38.

Cameron-Faulkner, T., Lieven, E., & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science, 27*(6), 843-874.

Cartwright, T., & Brent, M. (1997). Syntactic categorization in early language acquisition: Formalizing the role of distributional analysis. *Cognition, 63*(2), 121-170.

Chomsky, N. (1957). Syntactic structures. *The Hague: Mouton.*

Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language & Cognitive Processes. Special Issue: Language acquisition and connectionism, 13*(2&3), 221-268.

Christiansen, M. H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science, 23*(2), 157-205.

Clahsen, H., Aveledo, F., & Roca, I. (2002). The development of regular and irregular verb inflection in Spanish child language. *Journal of Child Language, 29*(3), 591-622.

Crain, S. (1991). Language acquisition in the absence of experience. *Brain and Behavioral Sciences, 14*, 597-611.

Dabrowska, E. (2001). Learning a morphological system without a default: The polish genitive. *Journal of Child Language, 28*(3), 545-574.

Dromi, E., Leonard, L. B., Adam, G., & Zadunaisky-Ehrlich, S. (1999). Verb agreement morphology in Hebrew-speaking children with specific language impairment. *Journal of Speech, Language, & Hearing Research, 42*(6), 1414-1431.

Eddington, D. (2002). Dissociation in Italian conjugations: A single-route account. *Brain & Language Special Issue: Mental lexicon II, 81*(1-3), 291-302.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*(2), 179-211.

Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning, 7*, 195-224.

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition, 48*, 71-99.
Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition, 48*(1), 71-99.
Elman, J. L. (1995). Language as a dynamical system. In R. P. a. T. v. Gelder (Ed.), *Mind as motion: Dynamical perspectives on behavior and cognition* (pp. 195-225). Cambridge, MA: MIT Press.
Elman, J. L. (1998). *Generalization, simple recurrent networks, and the emergence of structure*. Mahwah, NJ: Lawrence Erlbaum.
Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., & et al. (1996). Rethinking innateness: A connectionist perspective on development.
Fodor, J. A., & Pylyshyn, Z. W. (1988). *Connectionism and cognitive architecture: A critical analysis*. Cambridge, MA: MIT Press/Bradford Books.
Golinkoff, R. M., Mervis, C. B., & Hirsh-Pasek, K. (1994). Early object labels: The case for a developmental lexical principles framework. *Journal of Child Language, 21*(1), 125-155.
Gomez, R. L., & Gerken, L. A. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Science, 4*(5), 178-186.
Hare, M., & Elman, J. L. (1995). Learning and morphological change. *Cognition, 56*, 61-98.
Harris, Z. S. (1954). Distributional structure. *Word, 10*.
Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21*(1), 60-99.
Lewis, J. D., & Elman, J. L. (2001). A connectionist investigation of linguistic arguments from poverty of the stimulus: Learning the unlearnable. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the twenty-third annual conference of the cognitive science society* (pp. 552-557). Mahwah, NJ: Lawrence Erlbaum.
Lieven, E., Behrens, H., Speares, J., & Tomasello, M. (2003). Early syntactic creativity: A usage-based approach. *Journal of Child Language, 30*(2), 333-367.
MacWhinney, B. (1998). Models of the emergence of language. *Annual Review of Psychology, 49*, 199-227.
MacWhinney, B. (2000). The childes project: Tools for analyzing talk, vol. 1: Transcription format and programs (3rd ed.).
MacWhinney, B. (Ed.). (1999). *The emergence of language*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
Maratsos, M., & Chalkley, M. A. (1980). The internal language of children's syntax: The ontogenesis and representation of syntactic categories. In K. Nelson (Ed.), *Children's language* (Vol. 2). New York: Gardner Press.
Marchman, V. (1993). Constraints on plasticity in a connectionist model of the English past tense. *Journal of Cognitive Neuroscience, 5*, 215-234.
Marchman, V. A., Plunkett, K., & Goodman, J. (1997). Overregularization in English plural and past tense inflectional morphology: A response to Marcus (1995). *Journal of Child Language, 24*(3), 767-779.

Marcus, G., Ullman, M., Pinker, S., Hollander, M., Rosen, T. J., & Xu, F. (1992). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development, 57.*

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

McClelland, J. L., & Patterson, K. (2002). 'words or rules' cannot exploit the regularity in exceptions: Reply to pinker and Ullman. *Trends in Cognitive Sciences, 6*(11), 464-465.

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

Mintz, T. H. (2002). Category induction from distributional cues in an artificial language. *Memory & Cognition, 30*(5), 678-686.

Munakata, Y., & McClelland, J. L. (2003). Connectionist models of development. *Developmental Science, 6*(4), 413-429.

Munakata, Y., & Pfaffly, J. (2004). Hebbian learning and development. *Developmental Science, 7*(2), 141-148.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science, 14*, 11-28.

Orsolini, M., Fanari, R., & Bowles, H. (1998). Acquiring regular and irregular inflection in a language with verb classes. *Language & Cognitive Processes, 13*(4), 425-464.

Patterson, K., Lambon Ralph, M. A., Hodges, J. R., & McClelland, J. L. (2001). Deficits in irregular past-tense verb morphology associated with degraded semantic knowledge. *Neuropsychologia, 39*(7), 709-724.

Pinker, S. (1984). Language learnability and language development. *Cognitive science series. Cambridge, MA, US: Harvard University Press, 435*.

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA, USA: Harvard University Press.

Pinker, S. (1991). Rules of language. *Science, 253*(5019), 530-535.

Pinker, S. (1994). *The language instinct*. New York, NY, USA: William Morrow & Co, Inc.

Plunkett, K., & Juola, P. (1999). A connectionist model of English past tense and plural morphology. *Cognitive Science, 23*(4), 463-490.

Plunkett, K., & Marchman, V. (1993). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition, 48*(1), 21-69.

Plunkett, K., & Nakisa, R. C. (1997). A connectionist model of the Arabic plural system. *Language & Cognitive Processes. Special Issue: Cognitive models of speech processing: Psycholinguistic and computational perspectives on the lexicon, 12*(5-6), 807-836.

Plunkett, K., Sinha, C., Moller, M. F., & Strandsby, O. (1992). Symbol grounding or the emergence of symbols? Vocabulary growth in children and a connectionist net. *Connection Science: Journal of Neural Computing, Artificial Intelligence & Cognitive Research Special Issue: Philosophical issues in connectionist modeling, 4*(3-4), 293-312.

Prince, A., & Pinker, S. (1988). Rules and connections in human language. *Trends in Neurosciences, 11*(5), 195-202.

Ragnarsdottir, H., Simonsen, H. G., & Plunkett, K. (1999). The acquisition of past tense morphology in Icelandic and Norwegian children: An experimental study. *Journal of Child Language, 26*(3), 577-618.

Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science, 22*(4), 425-469.

Rodriguez, P., Wiles, J., & Elman, J. L. (1999). A recurrent neural network that learns to count. *Connection Science, 11*(1), 5-40.

Rumelhart, D. E., & McClelland, J. L. (1986). *On learning the past tenses of English verbs*. Cambridge, MA: MIT Press.

Saffran, J. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language, 44*(4), 493-515.

Santelmann, L. M., & Jusczyk, P. W. (1998). Sensitivity to discontinuous dependencies in language learners: Evidence for limitations in processing space. *Cognition, 69*(2), 105-134.

Seidenberg, M., & Bruck, M. (1990). *Consistency effects in the generation of past tense morphology.* Paper presented at the Thirty-first Meeting of the Psychonomic Society, New Orleans.

Seidenberg, M. S., & Hoeffner, J. H. (1998). Evaluating behavioral and neuroimaging data on past tense processing. *Language, 74*(1), 104-122.

Siegelmann, H. T. (1995). Computation beyond the Turing limit. *Science, 268*(5210), 545-548.

Theakston, A. L., Lieven, E. V. M., Pine, J. M., & Rowland, C. F. (2001). The role of performance limitations in the acquisition of verb-argument structure: An alternative account. *Journal of Child Language, 28*(1), 127-152.

Theakston, A. L., Lieven, E. V. M., & Tomasello, M. (2003). The role of the input in the acquisition of third person singular verbs in English. *Journal of Speech, Language, Hearing Research, 46*(4), 863-877.

Tomasello, M. (2000). The item based nature of children's early syntactic development. *Trends in Cognitive Science, 4*, 156-163.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Wittek, A., & Tomasello, M. (2002). German children's productivity with tense morphology: The perfekt (present perfect). *Journal of Child Language, 29*(3), 567-590.

Xu, F., & Pinker, S. (1995). Weird past tense forms. *Journal of Child Language, 22*(3), 531-556.

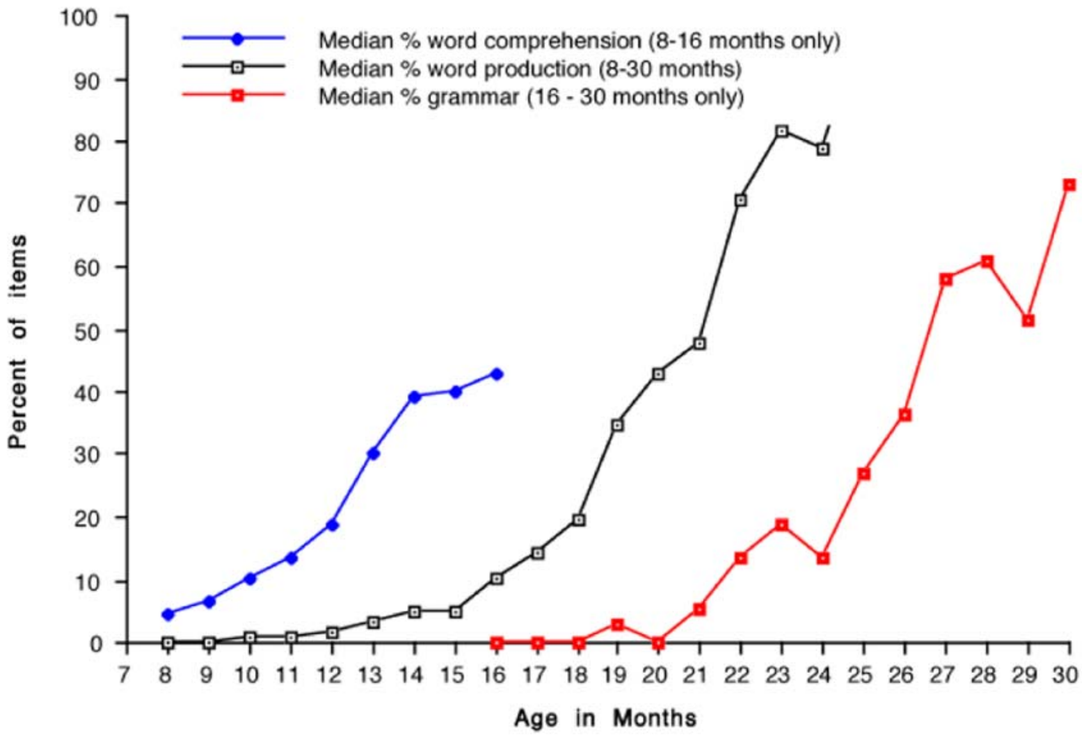


Figure 1. Median growth scores for word comprehension expressed as a percentage of available items (from Bates & Goodman, 1997).

 

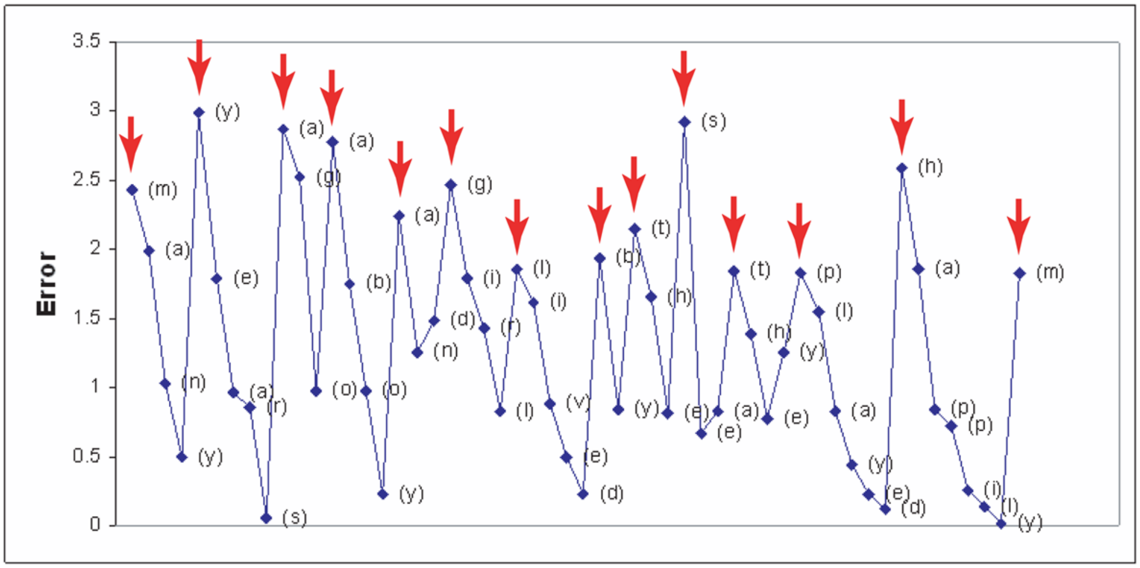


Figure 2. Performance of a simple recurrent network that has learned to predict the next letter in a short story. Error maxima are highly correlated with the onsets of a new word (from Elman, 1990).