

# Becoming Syntactic

Franklin Chang

Max Planck Institute for Evolutionary Anthropology

Gary S. Dell and Kathryn Bock

University of Illinois at Urbana–Champaign

Psycholinguistic research has shown that the influence of abstract syntactic knowledge on performance is shaped by particular sentences that have been experienced. To explore this idea, the authors applied a connectionist model of sentence production to the development and use of abstract syntax. The model makes use of (a) error-based learning to acquire and adapt sequencing mechanisms and (b) meaning–form mappings to derive syntactic representations. The model is able to account for most of what is known about structural priming in adult speakers, as well as key findings in preferential looking and elicited production studies of language acquisition. The model suggests how abstract knowledge and concrete experience are balanced in the development and use of syntax.

**Keywords:** sentence production, syntax acquisition, connectionist models

How do we learn to talk? Specifically, how do we acquire the ability to produce sentences that we have never said or even heard before? This question was central to the famous challenge from Chomsky (1959) to Skinner. Chomsky's view was iconoclastic: Speakers possess abstract syntactic knowledge, and the basis for this knowledge is in the human genes.

A system with abstract syntax is capable of producing novel and even unusual utterances that are nonetheless grammatical. Such knowledge is typically described in terms of syntactic categories (e.g., noun or verb), functions (e.g., subject or object), and rules (e.g., determiners precede nouns). The knowledge is abstract in the sense that it is not tied to the mappings between particular meanings and words. We accept the existence of abstract syntax, but in this article, we emphasize what is learned over what may be innate. We claim that the syntactic abstractions that support production arise from learners' making tacit predictions about upcoming

words in what they hear. If those predictions are erroneous, the learner makes changes to the system that generated the predictions. These ideas are made concrete in a connectionist model of the acquisition of production skills, one that accounts for data that reveal how experience adaptively alters these skills, most important, data concerning structural or "syntactic" priming in production.

Error-based learning algorithms in connectionist networks use the difference between a predicted output and the correct or target output to adjust the connection weights that were responsible for the prediction. One type of error-based learning, *back-propagation*, can be used to adjust the weights to hidden units in a network, units that are neither input nor output, thus allowing the model to learn arbitrary pairings of inputs and outputs (Rumelhart, Hinton, & Williams, 1986). One kind of back-propagation-trained network, called a *simple recurrent network* (SRN), has been particularly important in theories of language and sequence processing because it accepts inputs, and predicts outputs, sequentially (Elman, 1990). An SRN is a feed-forward three-layered network (input-to-hidden-to-output). It also contains a layer of units called the *context* that carries the previous sequential step's hidden-unit activations. By carrying a memory of previous states (akin to James's, 1890, notion of the "just past"), the system can learn to use the past and the present to anticipate the future.

SRNs provide some of the best accounts of how people extract generalizations in implicit sequence learning tasks (e.g., Cleeremans & McClelland, 1991; Gupta & Cohen, 2002; Seger, 1994). In these tasks, people learn to produce training sequences, such as a sequence of keypresses, and then are tested on novel sequences, showing that they have abstracted generalizations from the training. These models can easily be applied to language. SRNs that predict the next word at output given the previous word at input are able to learn syntactic categories and relationships from the sequential structure of their linguistic input (Christiansen & Chater, 1999, 2001; Elman, 1990, 1993; MacDonald & Christiansen, 2002; Rohde & Plaut, 1999). More complex models based on SRNs have been developed for comprehension, where meaning is predicted from word sequences (Miikkulainen, 1996; Miikkulainen & Dyer, 1991; Rohde, 2002; St. John & McClelland, 1990), and production, where word sequences are predicted from meaning

---

Franklin Chang, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany; Gary S. Dell and Kathryn Bock, Beckman Institute, University of Illinois at Urbana–Champaign.

This research formed a part of Franklin Chang's doctoral dissertation at the University of Illinois at Urbana–Champaign. An early version of the model was presented in a poster session at the March 2003 City University of New York Sentence Processing Conference in Boston, and the language acquisition results were presented at the Boston Conference on Language Acquisition, November 2004.

Preparation of this article was supported by National Science Foundation Grants SBR 94-11627 and 98-73450, National Institutes of Health Grants R01 HD21011, T32MH 1819990, HD-44455, and DC-00191, and a postdoctoral fellowship from the Department of Developmental and Comparative Psychology at the Max Planck Institute for Evolutionary Anthropology (Director: Michael Tomasello). We thank Kirsten Abbot-Smith, Morten Christiansen, Kara Federmeier, Cynthia Fisher, Yael Gertner, Evan Kidd, Elena Lieven, Danielle Matthews, and Michael Tomasello for their helpful comments on the article. Materials related to the model are available on the Web (at <http://osgood.cogsci.uiuc.edu/~fchang/becomingsyn.html>).

Correspondence concerning this article should be addressed to Franklin Chang, who is now at NTT Communication Science Laboratories, 2-4 Hikari-dai, Seika-cho, Soraku-gun, Kyoto 6190237, Japan. E-mail: [chang.franklin@gmail.com](mailto:chang.franklin@gmail.com)

(Chang, 2002; Chang, Dell, Bock, & Griffin, 2000; Dell, Chang, & Griffin, 1999; Miikkulainen & Dyer, 1991; Rohde, 2002).

Our model is a variant of the dual-path model of Chang (2002), a connectionist treatment of the acquisition of production skill. The original dual-path model augmented SRN approaches to language with architectural assumptions that enabled the network to acquire syntactic abstractions. Armed with these abstractions, the model was able to generalize in a symbolic fashion. It could accurately produce novel sentences, something that SRN-based models of production cannot reliably do (Chang, 2002). At the same time, though, because it used error-based learning in a recurrent network, it was, at least in principle, compatible with connectionist accounts of distributional learning in linguistic (e.g., Elman, 1990) and nonlinguistic (e.g., Gupta & Cohen, 2002) domains. Here, we ask whether the model withstands psycholinguistic scrutiny.

The model learns to produce by listening. When listening, it predicts (outputs) words one at a time and learns by exploiting the deviations between the model's expectations and the actually occurring (target) words. Outputting a word sequence is, of course, a central aspect of production, and so our version of the dual-path model can seamlessly transfer its knowledge gained by predicting during listening to actual production. We do not explicitly model comprehension, that is, the extraction of meaning from word sequences. However, we do assume that prediction is occurring during listening and that this affects processing and learning. Evidence for prediction during input processing comes from empirical demonstrations of the activation of the grammatical and semantic properties of upcoming words during comprehension (Altmann & Kamide, 1999; Federmeier & Kutas, 1999; Kamide, Altmann, & Haywood, 2003; Wicha, Moreno, & Kutas, 2003, 2004).

The model accounts for data from three experimental paradigms that purportedly test for adults' and children's use of syntactic representations: *structural priming*, *elicited production*, and *preferential looking*. The most central of these is structural priming. Structural priming creates structural repetition, which is a tendency for speakers to reuse previously experienced sentence structures (Bock, 1986). Of importance, the influence of a previously processed sentence, or *prime*, on the production of a *target* sentence persists over time and in the face of intervening sentences. Because of this persistence, it has been argued that structural priming is a form of implicit learning (Bock & Griffin, 2000) and that error-based learning is a way to model it (Chang et al., 2000).

In addition, we show that the model accounts for two other kinds of data that are important in the study of language development, data from elicited production and preferential-looking tasks. Results from these tasks have been at the center of a debate about the abstractness of syntax in children and the innate endowment for syntax. The issue concerns the abstractness of early transitive and intransitive structures. If these structures become abstract late in development, it supports *late-syntax* theories that posit that syntax becomes abstract through accumulated experience (Bowerman, 1976; Braine, 1992; Tomasello, 2003). If they are shown to be abstract early in development, then it is possible that experience alone is not enough for abstraction, and that would support *early-syntax* theories that assume that children have some innate linguistic propensity for abstract syntax (Gleitman, 1990; Naigles, 2002; Pinker, 1984). The debates arise in part because of experimental methodologies. Elicited production studies have tended to support

the late-syntax view (Tomasello, 2000), whereas preferential-looking studies show evidence for early structural abstraction (Naigles, 1990). We aim to resolve this debate by showing that the model, which acquires syntax gradually from prelinguistic architectural and learning assumptions, can account for the data from both methods.

The model presented here is ambitious, because there are no explicit, unified theories of the domains that it addresses. In syntax acquisition, there are no explicit theories that can explain structural priming; in sentence production, there are no explicit theories that can account for preferential-looking data; in language acquisition, there are no explicit theories that can deal with the problems of combinatorial behavior in neural systems (Fodor & Pylyshyn, 1988; Marcus, 1998, 2001; Pinker, 1989). The model attempts to provide a computational account of all of these phenomena. We present this account in four sections. The first (The Dual-Path Model) outlines the model architecture, the language that the model was trained on, and the accuracy of the model after training. The second section (Structural Priming) describes the testing of the trained model on structural priming results. The third section (Language Acquisition) deals with the model's account for language acquisition results in different tasks. In the fourth section (Successes and Limitations of the Model), we review and critique the model's behavior.

### The Dual-Path Model

Sentence production requires learning how to map between meaning (the message) and word sequences in a way that conforms to the syntax of a particular language (Bock, 1995; Levelt, 1989). An important property of this system is that people are able to use words in novel ways. For example, an editor of a celebrity gossip Web site created a verb to refer to the ravenous way that Catherine Zeta-Jones eats vegetarian food, as in "I had zeta-jonesed one too many carb-loaded dinners at Babbo to fit into my size zero skirt" (Safire, 2003; Spiers, 2003). This ability requires that the author ignore her experience with the proper name "Zeta-Jones" (e.g., occurs after "Catherine," occurs in proper noun positions) and treat it as a verb (after auxiliaries like *had*, can be in the past tense).

Whereas humans naturally come by the ability to use words flexibly, getting a learning system to do it is difficult. SRNs and other distributional learning schemes (e.g., Mintz, 2003; Mintz, Newport, & Bever, 2002; Redington, Chater, & Finch, 1998) are able to learn lexical categories based on the distributions of the words that they are exposed to. If an SRN is augmented with a message, the resulting model can also be taught to produce sentences, one word at a time, consistent with experienced messages (e.g., Chang et al., 2000). Moreover, its internal states can represent syntactic-semantic categories that are useful for mapping meaning onto particular words in particular positions.

Despite these merits, an architecture consisting of solely an SRN with a message input has a limited ability to use words in novel structural configurations, such as the verb "*zeta-jonesed*" (Chang, 2002; Marcus, 1998, 2001). For example, a model of this sort tested by Chang (2002) was unable to generalize a noun to a novel thematic role: If the model had not experienced the concept DOG as a GOAL, it could not produce sentences expressing this meaning. Moreover, the model was unable to produce adjective-noun combinations that it had not been trained on, which people can

readily do, as in the lyric by Echo and the Bunnymen “And an *ugly beauty* [italics added] was my own invention.” In short, this architecture was too bound by its training to exhibit the flexibility of the human production system.

To address the need for a flexible production system within a connectionist framework, Chang (2002) developed the dual-path model. The essential features of the current model’s architecture are the same as in the earlier work, but the input grammar had to be augmented to handle new empirical domains, and the message was simplified. Here we describe the basic properties of the dual-path model and how it has been adapted to account for the phenomena of interest in the present work.

### Incremental Word Prediction

The model’s task is to predict words, one at a time, using the immediately previous word and meaning as input (see Figure 1). The output layer is designated the *word* layer, in which each unit corresponds to a word in the model’s lexicon. The input layer is the *cword* layer, where the *c-* prefix reflects the status of the previous word as comprehended input, rather than predicted output. This layer also has units for every lexical item. When the model’s word output is erroneous (compared with heard next word; see the double arrow at the top of Figure 1), its connection weights are altered so as to reduce this error in the future.

The dual-path model adapted the word-prediction task to the task of sentence production by including a representation of the intended message. A message is a constellation of concepts organized by an event-semantic representation. When a message is present, the sequence of predicted words becomes the produced sentence, with each word output feeding back to serve as input for the production of the next word (see the dashed line in the dual-path model box in Figure 1). The message constrains the sequence to express what is intended.

The model only learns when it can compare its predictions against an externally generated utterance (and hence does not learn when it produces). Sometimes this prediction of an external utterance takes place when the message can be inferred from context (called a *situated* input event) and sometimes when message information is not inferable (called a *messageless* event).

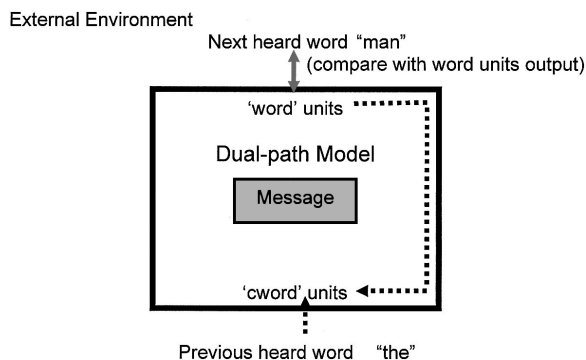


Figure 1. Incremental prediction in the dual-path model. The model is a device that takes the previous word as input and predicts the next word as output, constrained by the message. The *c-* prefix in *cword* reflects the status of the previous word as comprehended input, rather than predicted output.

### The Sequencing and Meaning Systems

The model’s architecture has two pathways for influencing the prediction–production of each word, one that maps from the concepts in the message, called the *meaning system*, and one that maps from the model’s *sequencing system* (see Figure 2). Both systems ultimately converge on the model’s *word* output layer, ensuring that message-consistent words are produced at the right time.

The sequencing system (see Figure 3) was designed to learn information that would ensure sentences were sequenced in a syntactically appropriate manner. The system had a SRN architecture, which has been used in other models for acquiring aspects of syntax (Elman, 1990, 1993; Rohde & Plaut, 1999). It mapped from the previous word in a sequence (*cword* units) to the next word in a sequence (*word* units) through a set of compression units (*ccompress* and *compress*) and a set of *hidden* units. The *hidden* units copied their activations into a set of *context* units, and these activations were passed as inputs to the *hidden* units. The compression units kept the *hidden* layer from directly sequencing particular words and instead forced it to create word classes (Elman, 1993). This made the network more syntactic, because the word classes that were the most useful for sequencing lexical items were syntactic categories.

For the sequencing system, the *cword* layer represents the previously predicted and/or heard word in the sequence. When a sentence is heard rather than produced, the *cword* units are set to the sum of the predicted *word* output and the actual heard input (normalized so that the sum does not exceed 1). In essence, the model’s *cword* “perception” is a blend of what it expects (previous predicted word) and what it hears (previous heard word). Because the *cword* units include predicted activation, the knowledge gained from heard input transfers readily to production, that is, when there is no heard input. In production, the *cword* activations are just the previous *word* output. This feeding back of produced output is particularly important for production because it helps the model keep track of the kind of sentence that it is producing and where it is in the sentence (Jordan, 1986).

The meaning system contains the message. The most important part of the message consists of concepts and event roles, and the bindings between the concepts and roles. Typically, these bindings are represented in connectionist models by using separate units that represent the binding between role and concept, such as DOG–AGENT and DOG–PATIENT (Chang et al., 2000; McClelland & Kawamoto, 1986; St. John & McClelland, 1990). Fodor and Pylyshyn (1988) criticized this approach, because in these networks, the ability to know or say that “John loves Mary” (JOHN–AGENT, MARY–PATIENT) is completely independent of the ability to know or say the related idea of “Mary loves John” (MARY–AGENT, JOHN–PATIENT). This increases the difficulty of language learning, because the system must individually learn that DOG–AGENT and DOG–PATIENT are both dogs and DOG–AGENT and CAT–AGENT are both agents (Pinker, 1989).

A similar problem exists in vision, because the retinal input is a two-dimensional image where object information is embedded in location-specific retinal fields (e.g., the flower and dog are not separately represented in the top part of Figure 4). During spatial processing, objects are individuated and linked to their location in space (e.g., middle section of Figure 4). The brain can do this



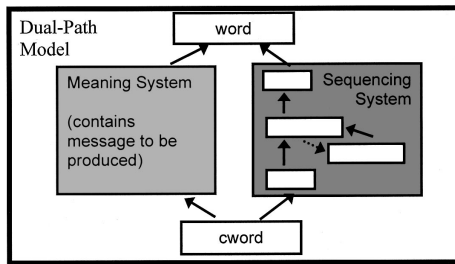


Figure 2. Two pathways in the model: a meaning system and a sequencing system. The sequencing system is an simple recurrent network.

because it has specialized pathways that separately process object properties, called the *what* system, and location properties, called *where* system (Milner & Goodale, 1995; Mishkin & Ungerleider, 1982) and the ability to temporarily bind these concepts to their locations (Karnath, 2001). Thus, it is likely that the spatial system gives a language learner a location-independent notion of the concept DOG before language learning begins (see the *what* side of the middle part of Figure 4).

Because concept–location bindings are required for spatial processing, similar mechanisms could be used for concept–role bindings in message encoding (see the bottom part of Figure 4). Chang (2002) implemented concept–role binding in messages by temporarily increasing weights between concepts (*what* units) and roles (*where* units). For example, setting the weight from DOG to AGENT to a high value identifies a dog as the agent. The dynamic binding used in the dual-path model’s weight-based message could be seen as a general feature of relational processing (Hummel & Biederman, 1992; Shastri & Ajjanagadde, 1993). Alternately, the message system could be a specialized part of the spatial system, which might help explain data showing tight links between spatial and language processing (e.g., Altmann & Kamide, 1999; Bloom, Peterson, Nadel, & Garrett, 1996; Clark & Carpenter, 1989; Griffin & Bock, 2000; Jackendoff, 1983; Lakoff, 1987; Lakusta & Landau, 2005; Landau & Jackendoff, 1993; Langacker, 1987). Our account is consistent with either approach.

The *what*–*where* character of the model’s message works together with its dual-path nature to promote generalization in sentence production. The sequencing system has only limited contact with the meaning system. Specifically, it does not connect directly to the concepts bound to the roles, but only to the roles. Therefore, when it learns to sequence, say, “dog” in “The dog carries the flower,” the sequencing system really only learns how to order the role that is linked to the dog concept. It does not sequence “dog” directly. Later, when the model is asked to produce “The cat carries the rose,” the cat concept is linked via fast-changing weights to the same role. Consequently, what the model learns about how to sequence this role transfers fully to cat. More generally, Chang (2002) showed that the dual-path model successfully generalized nouns to novel thematic roles and sentence structures and produced novel adjective–noun pairs in contrast to SRN-based models that lacked both the dual-path architecture and the *what*–*where* message structure. Moreover, the dual-path model’s generalization with verbs turned out to be constrained in a manner similar to what has been found with children (C. L. Baker, 1979; Gropen, Pinker, Hollander, Goldberg, & Wilson, 1989).

In addition to its ability to generalize in a humanlike manner, the model also accounted for neuropsychological data related to its assumptions of two paths. As the model learned, the meaning and sequencing pathways became differentially sensitive to particular word categories (content and function words, respectively). This allowed lesioned versions of the model to mimic double dissociations in aphasia related to the distinction between agrammatism and anomia, which are classically associated with syntactic and semantic lesions, respectively (Gordon & Dell, 2002, 2003). Thus, the model’s separation of meaning and sequencing receives some motivation from the aphasia literature and, more generally, from theoretical treatments of brain function that separate procedural (e.g., sequential) from declarative (e.g., semantic) properties of language (e.g., Cohen & Eichenbaum, 1993; Ullman, 2001).

### Detailed Assumptions About Messages

Figure 5 shows the full structure of the implemented model, including additional assumptions about the message. Here, we describe the meaning system in terms of its three parts: the meaning-to-word component, the word-to-meaning component, and the event semantics.

The meaning-to-word component of the meaning system involves three layers of units: *where*, *what*, and *word* units (see Figure 6). As already mentioned, the *where* units represented event roles in the message. These are denoted with single capital letters (e.g., A, X, Y, Z, D) and are explained later in the next section. The *what* units represented the lexical semantics of words (e.g., SLEEP, MARY, DOG). The links between the *where* and the *what* units are dynamic, set before the production of each sentence to represent a particular message. The connections between the *what* and the *word* units, in contrast, were learned as the model experienced particular words and concepts. These connections correspond to concept-to-lemma links in theories of lexical access in production (e.g., Caramazza, 1997; Dell, 1986; Dell, Schwartz, Martin, Saffran, & Gagnon, 1997; Levelt, Roelofs, & Meyer, 1999; Rapp & Goldrick, 2000).

For example, to represent the message for “Mary sleeps,” the *where* unit for the action role (role A) would be weight-linked to the *what* unit that represented the lexical semantics for sleep. The *where* unit for the role Y would be weight-linked to the *what* unit for Mary. Because of the fast-changing weights (see the thick links in Figure 6), whenever the system activated a

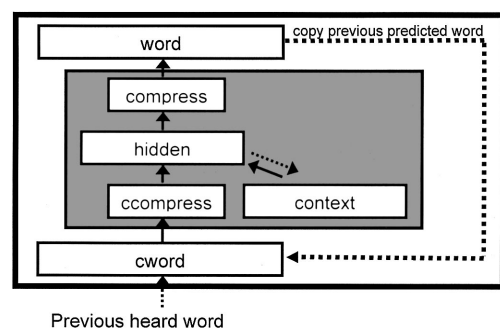


Figure 3. Sequencing system (simple recurrent network, right side of Figure 2); *ccompress* and *compress* are a set of compression units.

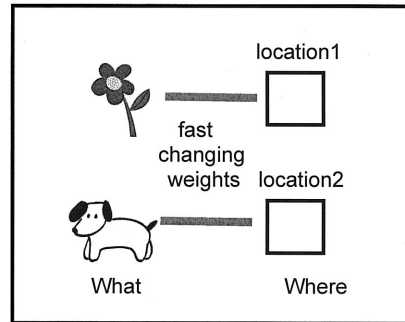
### Visual Input

Two-dimensional retinal image  
of a single flower-dog object.



### Spatial System

Dog and flower have been  
individuated and bound to their  
locations.



### Message

Dog and flower have been  
dynamically bound to roles to  
encode that dog is acting on the  
flower (e.g., carrying).

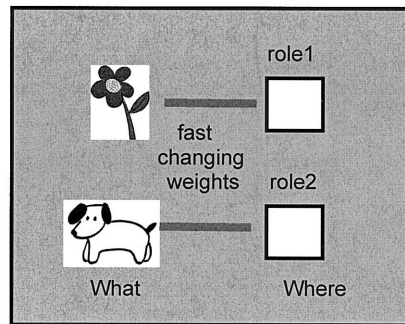


Figure 4. Visual input is translated into object (*what*) and locational (*where*) representations in the spatial system. The message makes use of this segmentation and dynamic binding mechanisms to assign objects to roles.

*where* unit, such as A or Y, the appropriate lexical-semantic content in the *what* units (SLEEP or MARY) would also become activated. (Control of activation of the *where* units resides in the sequencing system; see the connection in Figure 5 from *hidden* to *where*.) Because of the learned connections between the *what* and *word* layers, the activated *what* unit then biases for the output of the word that was learned in association with that unit (“sleep” or “Mary”).

The second part of the meaning system represents the word-to-meaning component (see Figure 7). It was used to relate incoming or already produced words (*cword* layer) to the message so that the model knows which components of the message have already been heard or produced. This mapping was just a reverse of the lexical meaning for output component of the message.

This reversal of the meaning-to-word component mapped a previous word (*cword*) to its lexical semantics (*cwhat*) and then to its role (*cwhere*). At the same time that the production message was set, the message links between *cwhat* and *cwhere* were set in the reverse direction (see the thick links in Figure 7, e.g., SLEEP → A, MARY → Y). If the *cword* unit for “Mary” was activated,

this would lead to the activation of the *cwhat* unit for MARY. Then, because of the message links (e.g., MARY → Y), the model would know that the Y role had been produced, and that information could be used in turn by the sequencing system to determine what to produce next. To facilitate memory of the roles that have been previously produced, the previous *cwhere* activations were copied into a *cwherecopy* layer that collected a running average of these previous states (see the dashed lines in Figure 7). The *cwhere* and *cwherecopy* units influence sequences through their connections to the *hidden* layer (see Figure 5).

Learning to link heard words and meaning in comprehension is a basic problem in language acquisition, because of the multiple possible semantic interpretations in a scene for a word (Gleitman, 1990). The model trains the *cword*–*cwhat* links by using the previous *what* unit activations as a training signal on the *cwhat* layer (see the double arrow in Figure 7). The model simply associates the current *cword* input with all the active *what* features. By experiencing enough situations in which, for example, the *cword* “cat” was heard when the *what* feature CAT was active, the correct connections can be established. Because activation of the *what* units from inferred meanings controls both the learning of

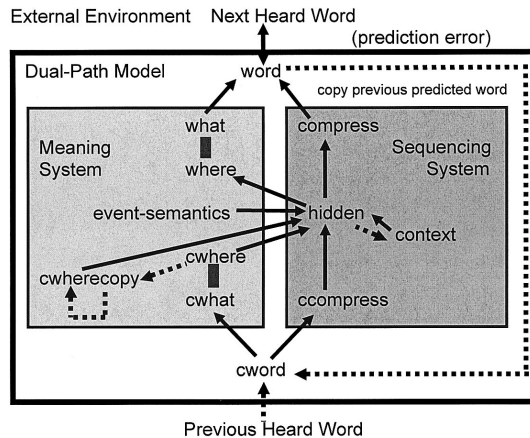


Figure 5. Dual-path model and environment; *ccompress* and *compress* are sets of compression units. Single arrows are learned weights. Dashed arrows are fixed copy connections. Thick gray lines are fast-changing message weights. Double arrows are for target comparison (*what* layer output as target for *cwhat* layer not shown).

words and *cwords*, this approach helps to ensure that the links in the word production and comprehension systems are similar.<sup>1</sup>

The word-to-meaning component plays an important role in bringing the model into line with incremental psycholinguistic theories of production (Bock, 1982; Dell, Burger, & Svec, 1997; Kempen & Hoenkamp, 1987; Levelt, 1989). For the dual-path model to learn to produce sentences incrementally and deal with syntactic alternations, events at the “choice point” in the sentence are critical. This is the point at which a structural alternative is uniquely determined (e.g., a double-object dative vs. a prepositional dative is determined right after the main verb, as in “The girl gave a flower to the boy/the boy a flower”). At the choice point, information from previous words (*cword* input) interacts with the message (the event semantics component of the message as described later) to bias events in the sequencing system. The model learns how to do this from predicting the sentences of others. Recall that the *cword* inputs are always the normalized sum of the previous predicted output activations and the external inputs. During prediction of another’s words during comprehension, the external input at the choice point provides the information for predicting the structure of the oncoming sentence. For example, in the passive “The flower was carried by the dog,” the early production of the word “flower” signals a passive, because in the target message, the flower is the patient (*cword* “flower” activates *cwhat*

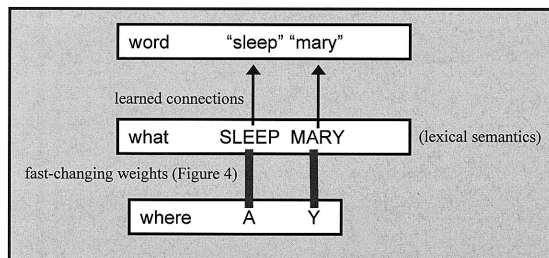


Figure 6. Meaning-to-word component (top of meaning system in Figure 5).

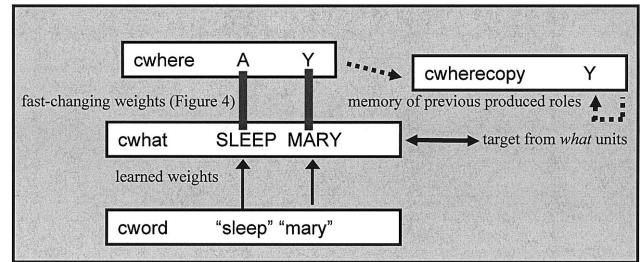


Figure 7. Word-to-meaning component (bottom of meaning system in Figure 5).

FLOWER, which is linked to the patient role in the *cwhere* units). Because the *cwhere* units help to signal the structure of the rest of the sentence, the sequencing system depends on this information to activate roles in the *where* units when appropriate. In production a similar process occurs, except there is no external input word. Hence, at a choice point, the predicted output is the *cword* input, and the model must use its previously learned representations to sequence the correct structure.

The third part of the message is the event-semantics units, which represents some of the relational information that the sentence conveys (middle of meaning system in Figure 5). The distinction between event semantics (functions–relations) and lexical semantics (concepts) is assumed in most semantic frameworks (Grimshaw, 1990; Jackendoff, 1983, 1990; Pinker, 1989; Talmy, 2000). Event semantics influences structure selection, whereas lexical semantics influences word selection. Among other things, such an arrangement helps to explain how speakers generalize nouns as verbs (Clark & Clark, 1979). For example, the semantics of the noun “Google” (the Web site) does not tell one which syntactic frame is appropriate when one wants to use it as a verb. However, because people normally use the Web site to search for something, the typical action involves two arguments. Knowing that it takes two arguments allows speakers in different languages to pick out the appropriate frame if they want to use it as verb, as in English “I googled myself,” German “Ich habe mich gegoogelt” (have-auxiliary, verb-final), or Japanese “jibun-o guuguru shita” (omitted subject, light verb).

In the implemented model, event-semantics represented the number of arguments and features of tense and aspect and was directly available to the sequencing system (and hence syntactic decisions). Altogether, then, the message for a particular sentence can be thought of as the combination of the *where–what* links, *cwhat–cwhere* links, and *event-semantics* information. Whereas these message links are set before each sentence, all the rest of the links in the model are learned through back-propagation (i.e., *cword* → *cwhat*, *what* → *word*, *event-semantics* → *hidden*, *hidden* → *where*, *cword* → *ccompress*, *compress* → *word*, and all internal

<sup>1</sup> Although our implementation uses different units and connections for production and comprehension within the meaning system, we take no stand on the issue of common versus separate representations for these functions (see a similar approach to single-word production in Plaut & Kello, 1999, where shared meaning–word links are implemented with separate weights and units). It is simply easier to train and debug the model when the use of the message for production purposes and its use for determining the role of the last produced or input word are kept distinct.



links within the sequencing system except for the process of copying *hidden* unit activations onto the *context* units).

Before describing the results of the empirical tests of the model, it is useful to summarize the model's critical assumptions in tabular form (see Table 1). These assumptions identify properties of the human brain that constrain language acquisition, and they include claims about learning mechanisms, architecture, and semantic representations. Most were described in the previous section, except for the XYZ roles assumption, which is described in the next section. We assume that the brain's learning and processing mechanisms function throughout life (learning-as-processing assumption) and are sensitive to the difference between expected input and the actual external input (prediction error assumption). This approach to learning and processing is present in many psychological models (e.g., Botvinick & Plaut, 2004; Gupta & Cohen, 2002; Plaut, McClelland, Seidenberg, & Patterson, 1996). The dual-pathways assumption finds motivation in neuropsychological and neurophysiological research. Studies of aphasia and other pathologies (e.g., Ullman, 2001) and functional imaging studies (e.g., Indefrey & Levelt, 2004; Levelt, Praamstra, Meyer, Helenius, & Salmelin, 1998) associate frontal and temporal-parietal areas with distinct functions, frontal areas being associated more with syntax and sequential output (Botvinick & Plaut, 2004; Keele, Ivry, Mayr, Hazeltine, & Heuer, 2003; Petersson, Forkstam, & Ingvar, 2004), and temporal-parietal areas being the locus of stored lexical representations (Rogers et al., 2004). The sequencing-by-SRN assumption capitalizes on the known ability of SRNs to learn and produce sequences in a manner consistent with human learning data (Cleeremans & McClelland, 1991; Gupta & Cohen, 2002; Seger, 1994). The meaning assumptions in the model stem from mechanisms that are also required for spatial processing. Multimodal spatial processing requires the ability to individuate objects and bind them to locations, and these abilities could also be used in message encoding (what-where assumption). Moreover, scene processing requires an algorithm for storing scene-based information, and this helps to motivate the model's role representation (XYZ assumption; see the next section and Chang, 2002, for evidence supporting this approach).

Overall, we assume that the brain is organized during development such that language develops in healthy individuals in similar ways (architectural innateness as argued by Elman et al., 1996). This organization is assumed to be a product of how the brain evolved to support sequence learning and scene processing in contexts without language (Conway & Christiansen, 2001; Milner & Goodale, 1995). When language is learned in an individual, these neural systems become at least somewhat specialized for language. The mature system may then exhibit some degree of dissociation between language and nonlanguage abilities (sequencing, spatial processing) if it is damaged, as in aphasia that spares nonlinguistic processing of scenes or sequences. Our claim, then, is that at the level of mechanism (but not necessarily in terms of learned representations), there are similarities between language and nonlanguage processing.

What is not assumed in our framework, but rather must be learned, are the mappings and representations that are needed for language-specific processing. The system must learn to produce words (*what* → *word*) and comprehend words (*word* → *cwhat*). It must forge syntactic categories (e.g., nouns, verbs) and constructions (e.g., passive, double-object dative). It must also learn how to select between alternative language-specific structures depending on either event-semantics or previously selected words. Because we view the critical assumptions as providing part of the universal prelinguistic basis for language acquisition, we occasionally make reference to typological patterns in the syntax of languages of the world, because these patterns help to specify the space of possible languages that are learnable and also suggest patterns that are more or less frequent (M. C. Baker, 2005).

### *Input Environment: Message-Sentence Pairs*

The model was trained by exposing it to sentences and their meanings. These message-sentence pairs were generated by an input environment grammar, designed to teach the model about a variety of sentence types including those used in structural priming and language acquisition studies. Table 2 presents examples of all

Table 1  
*Critical Assumptions of Dual-Path Model*

Assumption	Description
Learning (L) assumptions	
(L1) Learning as processing	The mechanisms through which language processing skill is initially acquired during childhood continue to function throughout life.
(L2) Prediction error	During both comprehension and production, learning occurs when a predicted word deviates from a target word.
Architectural (A) assumptions	
(A1) Dual pathways	There are separate meaning and sequencing systems, with restricted communication between them. The prediction-production of each word reflects the convergence of the outputs of the systems.
(A2) Sequencing by simple recurrent network	Sequential prediction-production is enabled by a context or short-term memory that changes as each word is predicted-produced.
Representational (R) assumptions	
(R1) What-where	Message representations consist of dynamic bindings between concepts ( <i>what</i> ) and roles ( <i>where</i> ).
(R2) XYZ roles	The participants in an event are associated with particular abstract roles, and these roles must distinguish transitive and intransitive agents.

Table 2  
Sentence Types in Input Environment Grammar

Sentence type	Example sentence
Animate intransitive	a dog sleep -ss.
Animate <i>with</i> intransitive	a cat was bounce -ing with the big brother -s.
Inanimate intransitive	a small flower is fall -ing.
Locative transitive	the cat is run -ing around me. i kick a cup. (active voice) it was bake -par by him. (passive voice)
Theme-experiencer	she surprise -ss a cat. (active voice) she is hurt -par by you. (passive voice)
Cause-motion	john put -ss the bread on the new mother.
Transfer dative	mary send -ss a mother the apple. (double object dative) a sister is give -ing the water to a man. (prepositional dative)
Benefactive dative	it bake -ss her a cake. (double object dative) a aunt bake -ss the apple for the grandma. (prepositional dative)
Benefactive transitive	mary carry -ed the orange for the grandma.
State-change	a uncle is fill -ing the sink with beer.
Locative alternation	a aunt spray -ed beer on the bath. (location-theme) the sister -s shower a bad bottle with water. (theme-location)

of the constructions present in the grammar. Because the grammar is complex and, technically speaking, independent of the model, we only summarize it here and present the details in the Appendix.

The input environment consisted of messages and sentences from a simplified grammar of English single-clause sentences. Here is an example:

Message: A=JUMP,

Y=BIRD, DEF

Event-Semantics:

AA=0.5 XX=0.5 PROG=0.5

Sentence: The bird is jump -ing.

Each message was composed of slot–filler combinations, where each slot was individuated by a capital letter identifying the type of slot (e.g., A, X, Y, Z, D), and each filler consisted of one or more concepts (e.g., BIRD, JUMP) and event-semantics features (e.g., AA, XX, PROG). The sentence associated with the message was an ordered set of words (e.g., “*bird*”) and inflectional morphemes (e.g., “-ing”). Notice that the message–sentence pair did not include a syntactic frame. The model had to develop its syntactic knowledge through learning.

In order to train the model on message–sentence pairs, the pairs had to be turned into inputs to the model. The sentence part of the pair is straightforward. When the model is exposed to the pair, each word–morpheme of the sentence is designated, in turn, as the “target” or desired output of the model. Exposing the model to a message was more complicated. Explaining it requires more details about, in turn, elements of messages in the grammar, the kinds of events associated with message information, and message representations in the model.

*Elements of messages in the grammar.* The letters X, Y, and Z designate abstract thematic roles. The particular set of roles assumed here is called the *XYZ role representation*. The XYZ role

representation was developed by testing a variety of representational schemes in the dual-path model. (A similar role scheme, called the “*spatial message*”, was described in Chang, 2002.) Although it does not correspond to any single linguistic theory, the XYZ system combines linguistic approaches to meaning (Dowty, 1991; Goldberg, 1995; Jackendoff, 1983, 1990; Levin & Rappaport Hovav, 1995; Van Valin & LaPolla, 1997) with conceptions of the nature of attention in event perception. Given that attention to spatial locations influences and is influenced by comprehension (Altmann & Kamide, 1999; Kamide, Altmann, & Haywood, 2003; Kamide, Scheepers, & Altmann, 2003; Knoeferle, Crocker, Scheepers, & Pickering, 2005) and production (Bock, Irwin, Davidson, & Levelt, 2003; Griffin & Bock, 2000), it seemed appropriate to develop an approach to assigning thematic roles that might be easily mapped onto visual scene analysis. Such an approach also meshes well with our assumption that the role–concept binding mechanism is akin to the mechanism binding locations and objects in spatial representations.

In order to learn language, children must have an algorithm for assigning thematic roles. How does the child know, when viewing a scene where a doll falls off a table, that the doll is an agent (as in “The doll jumped”) or a theme (as in “The doll fell”)? Given the subtlety of even basic distinctions such as these (Levin & Rappaport Hovav, 1995), it is not clear what algorithm children use to assign roles during online visual scene analysis. The XYZ roles alleviate this problem by forcing role assignment into a fixed order that approximates the way that scenes are viewed. The first role that is assigned is the “central” Y role. This role should be linked to the element of the scene that is most saliently changed or moved, or affected by the action (“*doll*” in both of the preceding examples). This typically includes the subject of unergative and unaccusative intransitives (“*The bread floated*,” “*Marty jumps*”) and the object of transitives (“*drink the milk*,” “*hit Marty*”). If the action on Y is caused by another element of the scene, whether this element is volitional (“*The girl eats the bread*”) or not (“*The noise surprised the girl*”), the causal element is assigned to the X role. If an action on Y involves movement to a location indexed by another element, as in transfer scenes (“*The boy gave the girl the dress*”) or in caused-motion events (“*The boy hit the ball to the girl*”), that element is assigned to the Z role. The Z role also indexes adjunct relationships to Y, for example adjuncts in *with* intransitives (“*John jumps with Mary*”) or locations (“*John is walking near Mary*”). In terms of traditional roles, X subsumes agents, causers, and stimuli; Y subsumes patients, themes, experiencers, and figures; and Z subsumes goals, locations, ground, recipients, and benefactors.

The most unusual aspect of the XYZ format is that it uses the same role for intransitive agents and transitive patients. There is, nonetheless, evidence for such a treatment. Goldin-Meadow and Mylander (1998) found that children without a language model (deaf children of nonsigning parents) treat these two elements in similar ways in the sign language that they invented with their parents. They tend to gesture about these elements before producing the gesture for the action, suggesting that there is a prelinguistic basis for treating them as the same role. Typologically, this ergative pattern in mapping appears in one fourth of the world’s language (Dixon, 1994). By assuming that messages have an ergative structure, one helps to motivate the existence of this mapping pattern. The sequencing system, on the other hand, is biased toward accusative mappings (e.g., where transitive agents



and intransitive subjects have similar surface features, as in English), because it learns sequences of syntactic categories that will tend to map the most prominent arguments (transitive agents and intransitive subjects) into similar positions. The model has these biases but must also learn how these biases work for individual verbs within different constructions in particular languages (Levin & Rappaport Hovav, 1995; Palmer, 1994).

*Kinds of events associated with message information.* Before describing how the message appears in the model, we must consider the states of the world that yield message information in the first place. Many theories of language acquisition (Gleitman, Cassidy, Nappa, Papafragou, & Trueswell, 2005; Pinker, 1984; Tomasello, 2003), assume that children can infer some extralinguistic meaning of heard utterances, thereby acquiring information that is useful for learning the mapping between meaning and linguistic forms. Of importance, there are many situations in which the child knows aspects of intended meaning before hearing an utterance, such as viewing a familiar picture book, playing known games, or taking part in common daily rituals (e.g., meals) in which the sequence of events and utterances is well-known. These are represented to the model as situated events—events in which the child can infer the message and hears a word sequence that expresses it. When the model experiences a situated event, learning results from the differences between the child's predictions of words (based on the inferred meaning and any prior words) and each heard word.

In reality, children might only be able to infer a part of the whole adult meaning. To simulate the noisiness of the inferable meaning, we also have messageless events, which have no message, but which are processed in the same way as situated ones. Learning occurs because prediction occurs, albeit prediction unconstrained by inferred meaning. What is important is not whether the message is all there or not (as we have implemented), but rather the consistency of the relationship between inferable parts of meaning and utterances.

The model experiences only situated and messageless events during training. To simulate the composition of materials in particular experiments, we chose event types that correspond most closely to the experimentally presented events. For example, to simulate the production of a sentence from a pictured event by adult participants in structural priming experiments, we used a variant of a situated event called a *production* event. The speaker's intended meaning is represented by a corresponding message, just like the situated event. Rather than using external input as *cword* input, however, the model uses its own produced output as a *cword* input. In this way, the output of a production event is a word sequence constrained by a message but unconstrained by any external input.

*Message representation in the model.* Now we are in a position to describe how the message is instantiated in the model. Each of the XYZ role units is represented in the message by a *where* and *cwhere* unit. Each role-filler concept in the message has its own *what* and *cwhat* unit. When the message specifies a role-concept link (e.g., X = BIRD), the weight between the role unit in the *where* layer (e.g., X) and the concept in the *what* layer (e.g., BIRD) is set to an arbitrary value of 6, high enough to ensure that the connected *what* unit was active when the host *where* unit was. The same was done for the *cwhat*–*cwhere* connection. In this instance, the link runs from concept to role, reflecting the fact that during input processing, input words map to concepts and then to their roles. These connections function the same as any other connection

in the model. They spread activation and backpropagate prediction error. The only difference is that their weights are not learned but are set anew for each message. Presumably, the message is set in humans by an independent planning system that is devoted to determining the appropriate communicative means for particular situational goals (see Levelt, 1989, chapter 4).

The second part of the message was the event-semantics units. Before a sentence is experienced in a situated event or formulated in a production event, the activation of these units is set and kept on during the entire sentence. The event semantics provides information about the overall form of the event—most important, the kinds of arguments present. This is accomplished by event-semantic units for each of the argument roles (unit XX for the X role, YY for the Y role, and ZZ for the Z role), one unit for the action (unit AA), and one unit for any preposition (unit DD).

The activation of the event-semantics units encodes the relative prominence of roles in the message (Goldberg, 1995; Grimshaw, 1990). When the YY unit was more activated than the XX unit, then the Y role was more prominent than the X role. The relative prominence of the roles influenced the model's output by causing words associated with more prominent roles to be placed earlier in the sentence. The model learned to do this because the environment input grammar reflected such a correlation, and the model's architecture allowed event-semantics to affect the sequencing system. In the message–sentence pairs from the grammar, the sentence structure associated with a message placed prominent arguments earlier. For example, a more prominent Y than X role would be associated in the training input with a structure in which Y is the subject (e.g., a passive). Hence, the event semantics helps the sequencing system learn language-specific frames for conveying particular sets of roles by giving the sequencing system information about the number of arguments and their relative prominence.

### *Producing a Sentence in the Model: An Example*

To illustrate how the model creates a sentence, here we give an extended example of a production event. In a production event, a message is present and the model's ongoing output, rather than heard input, drives the process. We assume that the model has been fully trained, and so its production will be accurate. (The actual training of the model is taken up in next section.)

The example sentence is “The boy is carried by the grandma.” With our lexicon, this sentence consists of the sequence “the boy is carry -par by the grandma” (-par being the past participle morpheme). The message associated with this sentence is given in Table 3, and the trace of the production process is shown in Table 4.

Before production begins, the message must be placed into the model and the event-semantics set. Roles and fillers are linked in the *what*–*where* and *cwhat*–*cwhere* units. For instance, the X *where* role unit is linked to the *what* concept unit for GRANDMA. GRANDMA receives the X role because it is a transitive agent. The X and Y roles are also linked to other modifying concepts. Here, for example, both *grandma* and *boy* are definite, and this feature is instantiated using a complex coding (see the Appendix for details) that Table 3 annotates as the DEF *what* feature. The *event-semantics* unit XX is set to 0.25, and YY is set to 0.5. Their relative activations bias production of a passive structure, because YY (corresponding to the Y role linked to BOY) is more activated than XX. Tense and aspect information is also associated with the event semantics. In this case, the event was not in the past and not

Table 3  
*Example Message for the Utterance “The boy is carried by the grandma”*

Role	Concept	Event semantics
A (action)	CARRY	AA=0.5 XX=0.25 YY=0.5
X (transitive agent)	GRANDMA, DEF	
Y (transitive patient)	BOY, DEF	
Z		

progressive, so the past tense and progressive aspect units were not activated. The model also initializes the activations of all of its other units as described in the Appendix.

The production of the first word (Time Step 1 in Table 4) entails the spread of activation from the *event semantics* and *context* layer (the latter having been initialized uniformly to 0.5) to the *hidden* layer. Because this is the first word, there is no input from the “c” layers (marked with a dash in Table 4). From the *hidden* layer, the activation contacts the role units in the *where* layer (e.g., X, Y), and hence the concepts GRANDMA, BOY, and DEF all become somewhat activated. Because the sequencing system has learned that articles come before nouns in the input, the activation converging on the output *word* layer favors “the” over “grandma” and “boy.” The *word* layer uses an approximate winner-take-all activation function (see the Appendix), and hence the *word* unit *the* suppresses other activated *word* units.

At Time Step 2, “the” is copied back to the *cword* units—the self-feedback property of production. Because both the agent and the patient are definite, the model cannot use the “the” to signal which structure to pursue, and the *cwhere* units are not strongly activated. Because the sequencing system has not settled on an appropriate structure, it again activates both the X and Y roles, which in turn activate BOY, GRANDMA, and DEF. The knowledge that the article “the” was produced can be used by the sequencing system (*cword* input) to suppress all the articles and to activate nouns through the syntactic categories in the *compress* units. This leaves the competition for output between “grandma” and “boy.” Theoretically, either one could win, but everything else being equal, there will be a bias for “boy” to win because the model has experienced passive structures when the YY *event-semantics* unit is more active than the XX one. Therefore, we assume that “boy” is chosen at *word* layer. Its activation is copied to the *cword* layer (Time Step 3), and the model moves on to produce the next word.

At Time Step 3, there is significant input to the *cwhat*–*cwhere* layers from the *cword*, “boy,” leading to the activation of the

*cwhere* unit Y, thus allowing the *hidden* units of the sequencing system to know that the model has produced the patient first. This information then leads, over the next several time steps, to learned states consistent with production of a passive, including most immediately a form of “to be.” The correct form “is” occurs because of the lack of plural features in the *what* layer for the subject, and the reflection of this fact in the sequencing system, and the lack of past-tense features in the event semantics. The model’s treatment of singular as default is in keeping with accounts of agreement in production (for a review, see Eberhard, Cutting, & Bock, 2005). At Time Steps 4 through 6, the learned sequence of states within the sequencing system biases for, first, activation of the action role A in the *where* layer leading to production of the verb stem, “carry,” followed by “-par” and “by.” The *cwherecopy* information becomes important at Time Steps 7 and 8. These units retain the fact that the patient has already been produced, and so it helps the sequencing system recognize that the *where* unit for X should now be active. Because X is associated with both DEF and GRANDMA, the sequencing system again has to use its acquired knowledge that the article precedes the noun. Finally, at Time Step 9, because the *cwherecopy* information and *cwhere* activations show that all of the roles have been produced that are active in the event semantics, the model generates an end-of-sentence marker (a period).

In summary, production in the model involves incremental competition between words that are activated by the message. The sequencing system attempts to make a grammatical sequence out of the winners of this competition, thereby constraining the patterns of activation. Next, we address how the model was trained to achieve this.

### Training the Dual-Path Model

Connectionist models, like people, vary in their experiences and in the knowledge they gain from their experience. To ensure that

Table 4  
*Schematic Summary of Model Events During Incremental Production of Sentence*

Time step	Cword	Cwhat	Cwhere-copy	Cwhere	Where	What	Word
1	—	—	—	—	X, Y	BOY, GRANDMA, DEF	the
2	the	—	—	—	X, Y	BOY, GRANDMA, DEF	boy
3	boy	BOY	—	Y	—	—	is
4	is	—	Y	—	A	CARRY	carry
5	carry	CARRY	Y	A	—	—	-par
6	-par	—	Y, A	—	—	—	by
7	by	—	Y, A	—	X	GRANDMA, DEF	the
8	the	—	Y, A	—	X	GRANDMA, DEF	grandma
9	grandma	GRANDMA	Y, A	X	—	—	.

Note. Dashes indicate that units are strongly biased and can not be easily labeled.

our modeling results are general, we created multiple model subjects, allowing for statistical tests of model properties. The only difference among the model subjects was their training experiences. Twenty training sets of 8,000 message–sentence pairs were generated and used to train 20 model subjects. As a means of yielding efficient learning of the grammar, about three quarters of the pairs on average were situated events and one quarter were messageless events. Because children hear approximately 7,000 utterances in a day (estimated in Cameron-Faulkner, Lieven, & Tomasello, 2003, from child-directed speech in corpora) and this input refers to a limited set of people, objects, and events that are part of the child’s world, the model’s input underestimates both the amount of input and the semantic predictability of this input.

Each model subject experienced 60,000 message–sentence pairs randomly selected from its training set, with weight updates after each pair. To see how well the model generalized to novel sentences, a testing set of 2,000 sentences was randomly generated from the grammar. The tests were all production events, because our empirical focus is on production and particularly generalization in production. The message grammar can generate approximately  $8.067158 \times 10^{10}$  different messages,<sup>2</sup> and the overlap in sentences between training and test was small (less than 1% overlap).

Production accuracy can be indexed in many ways. We defined two measures, *grammaticality* and *message accuracy*, which are determined by analyzing the model’s output word sequences. For example, consider the sequence “sally is hurt -par by a cat.” Each such word sequence was augmented with syntactic information, to yield a lexical–syntactic sequence: NOUN:sally AUX:is VTRAN:hurt MOD:-par PREP:by DET:a NOUN:cat. Lexical–syntactic sequences were derived both for the model’s outputs and for the sentences from the training or testing sets that the model was trying to produce. The model’s output sequence was considered grammatical if the whole syntactic sequence (“NOUN AUX VTRAN MOD PREP DET NOUN”) matched the syntactic sequence for any sentence in the training set. The output sequence’s message was accurate if its lexical–syntactic sequence and the intended lexical–syntactic sequence mapped onto the same message by a set of transformational rules set up for this purpose (e.g., NOUN:sally AUX:is VTRAN:hurt MOD:-par PREP:by DET:a NOUN:cat → ACTION:HURT X=NOUN:CAT Y=NOUN:SALLY). The message accuracy ignored differences in minor semantic features such as definiteness, number, aspect, and tense.

Figure 8 shows the grammaticality and message accuracy of the model for the training and testing sets every 2,000 epochs. An epoch is a point during training when weights are actually altered. Here, such alteration occurs at the end of every training sentence. The figure shows two important convergences. First, at the end of training, performance on the training and testing sets converges, showing that the model treats novel sentences from the grammar in the same way as trained exemplars (see the Appendix for why grammaticality is initially higher for test). In other words, the model generalizes extremely well. Second, at the end of training, grammaticality and message accuracy converge. Because message accuracy presupposes a grammatical lexical–syntactic sequence, this convergence shows that the grammatical sentences produced also tend to be the appropriate ones for the intended message. After 60,000 epochs, grammaticality is 89.1% and message accuracy is 82% on the test set. In summary, after training on just 8,000 examples (each example trained an average of 7.5 times), the

model can correctly produce the target utterance for most of the 80 billion meanings that can be expressed by the grammar. This is clearly an example of effective learning from an impoverished training set of positive examples, with no direct negative evidence.

Before we turn to the application of the model to empirical data, it is useful to summarize the structure of the modeling work and to preview its application to data. This summary is presented in Figure 9. We started with a formal grammar of an English-like single-clause language with a variety of constructions and grammatical phenomena. This grammar generated message–sentence pairs, which were used to train the internal syntactic representations in the dual-path model. The model consisted of a connectionist architecture with one part set up for sequencing items and another for representing meaning, instantiating a set of assumptions about learning. The architecture and learning assumptions were chosen to simulate a neural system that must sequence behavior, represent states of the world, and relate the behavior and the states to each other. For our purposes, these assumptions are given; they can be viewed either as representing innate factors or as arising from interactions between more basic innate factors and early experience.

After training, the model was applied to psycholinguistic data. Versions of the model with less training were applied to developmental data, and well-trained versions were applied to structural priming data from studies using adult participants. The next section begins the presentation of these applications.

### Structural Priming

One of the most convincing sources of evidence that people make use of abstract syntactic representations while speaking comes from the phenomenon of structural or syntactic priming. *Structural priming* is a tendency for speakers to reuse the abstract syntactic structures of sentences that they have produced before (Bock, 1986; Bock & Loebell, 1990). Structural priming experiments require people to produce meanings that can be conveyed in at least two structures, that is, they require materials with structural alternations. For example, the prepositional dative (e.g., *The man showed a dress to the woman*) and the double-object dative (e.g., *The man showed the woman a dress*) convey a similar meaning. When messages originate in a form that allows either of these structures to be produced (e.g., a picture of the event MAN SHOWS DRESS, WOMAN SEES DRESS), speakers can make a tacit choice between these two sentence structures. When the event description is preceded by a prepositional dative prime (e.g., *The rock star sold some cocaine to the undercover agent*), speakers are more likely to choose the prepositional dative structure than they would be otherwise. Similarly, when the event description is

<sup>2</sup> To give an example of how the total number of possible sentences is calculated, let us examine one of the largest frames in the language: the cause-motion construction with adjectives in each noun phrase as in the sentence “A bad girl push -ed the old car to the happy boy -s.”  $6$  (adjectives)  $\times 22$  (animate nouns)  $\times 3$  (number–definiteness)  $\times 5$  (transitive verbs for this frame)  $\times 2$  (tenses)  $\times 3$  (aspect–command)  $\times 6$  (adjectives)  $\times 16$  (inanimate nouns)  $\times 3$  (number–definiteness)  $\times 10$  (prepositions)  $\times 6$  (adjectives)  $\times 34$  (animate–inanimate nouns)  $\times 3$  (number–definiteness) =  $2.09392128 \times 10^{10}$ . Doing this for each frame and summing the results together yields the total number of possible sentences.



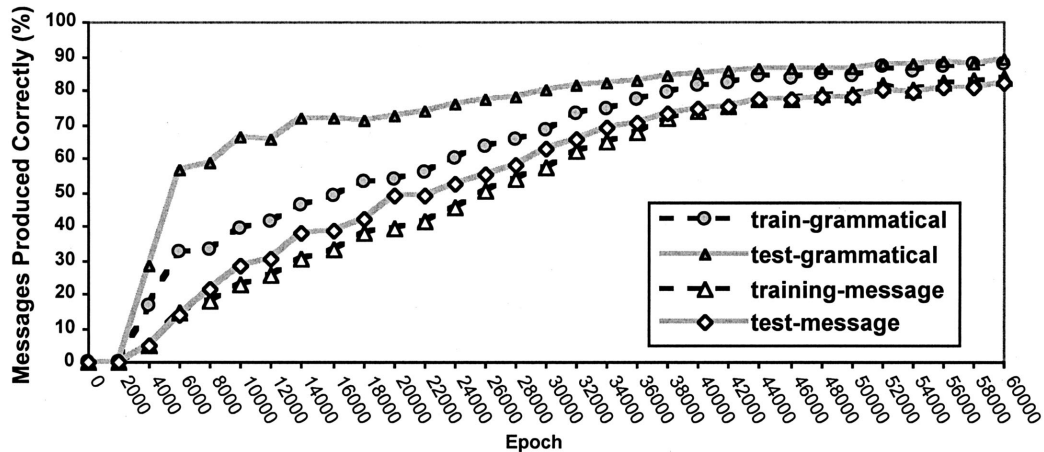


Figure 8. Average message and grammaticality accuracy during training for training and test sets.

preceded by a double-object dative structure (e.g., *The rock star sold the undercover agent some cocaine*), the likelihood of a double-object dative description increases. This greater tendency to use the primed structure when expressing the target message has been found for a variety of syntactic structures. It occurs in the absence of lexical and conceptual repetition and in the face of thematic role differences. It does not depend on similarities in prosodic patterns. These things suggest that priming involves abstract structural frames.

Two mechanisms have been offered to account for structural priming: activation and learning. The activation account postulates that structural priming is the result of the activation of a structural frame, which makes the frame easier to access (Bock, 1986; Branigan, Pickering, & Cleland, 1999). Activation-based phenomena tend to have a short life span (lexical priming with semantic or phonological primes typically lasts less than a few seconds; Levett et al., 1999), and so an activation account predicts that structural priming should disappear over a delay or after other sentences are processed. Contrary to this prediction, though, a variety of researchers have found that structural priming persists undiminished over time or the processing of other sentences (Bock & Griffin,

2000; Boyland & Anderson, 1998; Branigan, Pickering, Stewart, & McLean, 2000; Hartsuiker & Kolk, 1998; Huttenlocher, Vasilyeva, & Shimpi, 2004; E. M. Saffran & Martin, 1997).

Bock and Griffin (2000) argued that the long-lasting nature of priming, its lack of dependence on explicit memory (Bock, Loebell, & Morey, 1992), and the fact that people are not conscious of the priming manipulation (Bock, 1986) support the idea that structural priming is a form of implicit learning. Implicit learning has been characterized as a change in the strength of the connections in a neural network, unlike activation, which corresponds to the firing of network units (Cleeremans & McClelland, 1991). Finding that priming lasts over the processing of other stimuli means it is unlikely that priming is due to continued activation in the networks that support sentence production, and it suggests instead that the mechanism is enhanced strength in the connections between representational units that support the use of syntactic structure.

Because language learning also requires the ability to implicitly learn syntax, it is possible that structural priming stems from the same mechanisms. Chang et al. (2000) implemented this hypothesis. They built a connectionist model based on an SRN and trained it to produce a small English-like grammar. Then, they gave it prime-target pairs that were similar to the experimental conditions in some structural priming experiments (Bock, 1986; Bock & Loebell, 1990). When the model processed the prime, its learning algorithm continued to function, so that processing the prime induced additional learning. This learning then affected the way that the target was produced, creating structural priming.

The Chang et al. (2000) model, however, was more a model of priming than a model of production. It was developed with the priming data in mind and ignored crucial production functions. Most important, like other pure SRN-based models (e.g., see models in Chang, 2002), it could not routinely produce sentences that it was not trained on. Because creativity in sentence production and structural priming are both thought to depend on abstract syntactic representations, solving the generalization problem might provide a better account of structural priming.

We next describe the application of the dual-path model to structural priming. In doing so, we outline and defend three important claims: First, structural priming is a form of error-based implicit learning, with the same learning mechanism responsible

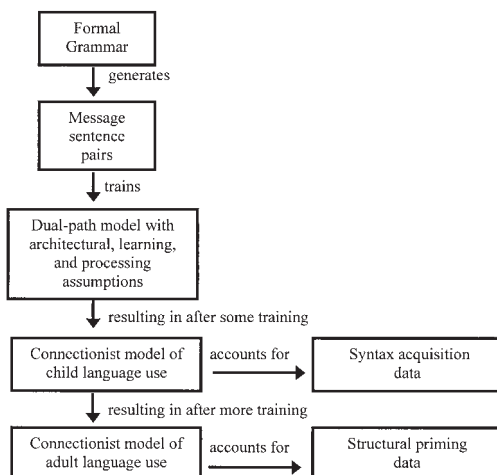


Figure 9. Structure of the modeling work

both for syntax acquisition and for priming. Second, the model acquires syntactic representations that are insensitive to thematic role similarities unless role information is necessary for learning a structural alternation. This property of the model enables it to explain how structural selection occurs on the basis of thematic role distinctions as well as cases in which structural priming appears to ignore thematic role distinctions. Third, the model's syntactic representations lead to priming that is insensitive to closed-class elements such as prepositions and inflectional morphology. This is also in agreement with the data. To help support these claims, the model's internal representations are explored to show how the model implements these results.

### Testing Structural Priming in the Model

To make prime–target pairs to test priming, message–sentence pairs were first generated by the input environment grammar. These message–sentence pairs were put into prime–target pairs such that the prime and target did not share message elements. Because most lexical elements had corresponding message elements, reducing message overlap also helped to reduce overlap in lexical items. For all of the priming tests, only three target structures were used: datives, transitives, and locative alternators. Each structural priming test set took 100 prime–target message pairs and crossed them with the two prime conditions and with two *target message preferences*, yielding 400 total prime–target pairs. Examples of the crossing of prime sentences and target messages for datives and transitives are shown in Tables 5 and 6, respectively, including both the same-structure and different-structure primes (in later examples, only the different primes are shown).

*Target message preference* refers to the way that event semantics biased toward one or the other alternation. In Tables 5 and 6, the *event-semantics* units XX, YY, and ZZ have activation values (e.g., XX = 0.5) that bias toward particular structures. A prepositional-dative biased dative target is one in which the YY *event-semantics* unit is more activated than the ZZ unit, and a double-object biased dative target is associated with the reverse. There are corresponding differences in the activation of XX and YY units for biasing actives and passives.

Structural priming in the models involved presenting the prime–target pairs to the model with learning turned on. First, before each prime–target pair the weights were set to the weights of the model at the end of training (Epoch 60,000). Each prime sentence was

then presented to the model word by word, and back-propagation of error was used to calculate the weight changes for all the weights in the network. After processing of the prime sentence, the weights were updated, and the production of the target sentence began. Then the target message was set, and the target sentence was produced word by word. Finally, the resulting structure was recorded.

Prime processing for the model consisted of the kind of training event that is associated with hearing a contextually unsupported or isolated sentence, a messageless event. This assumes that structural priming takes place during language comprehension. Bock and Loebell (1990) first raised the possibility of structural priming during comprehension, writing that

it is unknown whether priming is possible from comprehension to production, or vice versa. Assuming that production mechanisms are distinct from parsing mechanisms, a strict procedural view would predict no intermodality priming. However, if the assumption [that production and parsing are distinct] is wrong, even a procedural account would predict intermodal effects. (p. 33)

Later, Branigan and colleagues found evidence for priming from comprehension to production (Branigan, Pickering, & Cleland, 2000; Branigan et al., 1995), and further experiments have shown that the magnitude of priming from comprehended primes is similar to those that are comprehended and produced (Bock, Dell, Chang, & Onishi, 2005). Accordingly, in the application of the model to priming data, priming arises from the input processing of a messageless event. In such an event, the *cword* input includes external input and there is no message present during the prediction process. The target, which is typically a picture in priming studies, was associated with a production event—a message was present, and the *cword* input consisted of the model's produced output.

The same scoring procedures that were used for the training and testing sets were used to score the target sentence. First, the meaning of the produced target sentence had to match the target message. (Message accuracy on targets was 82% overall. Human studies sometimes yield lower results; e.g., 53% target responses occurred in Bock et al., 1992.) If the produced target sentence and message mismatched, the trial was eliminated; if they matched, the target's structure was assessed. The structure was what remained after stripping lexical information from the lexical–syntactic parse (e.g., NOUN AUX VTRAN MOD PREP DET NOUN). One

Table 5  
*Prime and Message Pairs for Dative Priming*

Prime-target types	Prime sentence	Target message
Prepositional dative Prepositional-dative-biased target	the mother -s give the orange to a grandma.	A=THROW X=UNCLE Y=BOTTLE Z=AUNT EVSEM: AA=0.5 XX=0.5 YY=0.475 ZZ=0.451 (e.g., the uncle throw -ss the bottle to the man)
Double object dative Prepositional-dative-biased target	the mother -s give a grandma the orange.	A=THROW X=UNCLE Y=BOTTLE Z=AUNT EVSEM: AA=0.5 XX=0.5 YY=0.451 ZZ=0.475 (e.g., the uncle throw -ss the man the bottle)
Prepositional dative Double-object-biased target	the mother -s give the orange to a grandma.	A=THROW X=UNCLE Y=BOTTLE Z=AUNT EVSEM: AA=0.5 XX=0.5 YY=0.451 ZZ=0.475 (e.g., the uncle throw -ss the man the bottle)
Double object dative Double-object-biased target	the mother -s give a grandma the orange.	A=THROW X=UNCLE Y=BOTTLE Z=AUNT EVSEM: AA=0.5 XX=0.5 YY=0.451 ZZ=0.475 (e.g., the uncle throw -ss the man the bottle)

Table 6  
*Prime and Targets for Transitive Priming*

Prime-target types	Prime sentence	Target message
Passive transitive Active-biased target	a apple is sculpt -par by the woman.	A=PUSH X=MARY Y=MAN EVSEM: AA=0.5 XX=0.5 YY=0.475 (e.g., mary push -ss a man.)
Active transitive Active-biased target	the woman sculpt -ss a apple.	
Passive transitive Passive-biased target	a apple is sculpt -par by the woman.	A=PUSH X=MARY Y=MAN EVSEM: AA=0.5 XX=0.475 YY=0.5 (e.g., a man is push -par by mary.)
Active transitive Passive-biased target	the woman sculpt -ss a apple.	

structure for each alternation was arbitrarily chosen to be the target structure, and the percentage of the two alternate structures that corresponded to the target structure was calculated. The prepositional dative was chosen to be the target structure for datives, the active for transitives, and the theme–locative for locative alternations. This treatment of the model's priming behavior enabled it to be directly compared with priming experiments, which also used these dependent variables.

The 20 trained model subjects were used for testing. Repeated-measures analysis of variance was performed on the percentages of the target structure, using model subject as the random factor. Effects were considered significant when the probability associated with them was less than .05. To emphasize the priming results in the graphs, the *priming difference* was calculated by subtracting from the percentage of target structures produced after the target-structure prime the percentage of targets produced after the alternative-structure prime.

The goal of the priming studies with the model was to assess the qualitative fit to the human data. That is, the aim was to see whether the model exhibits reliable priming under the same conditions that people do. In humans, structural priming experiments yield different magnitudes of priming, presumably due to variability in learning due to differences in task and materials, and speaker variables such as attention span, motivation, and the strength of existing representations. In the model, the magnitude of priming depends on the *learning rate* (a parameter in back-propagation that scales weight changes). In the present model, the learning rate during the testing of priming was the average (0.15) of the initial and final learning rates during training (0.25, 0.05).

Another factor that influences priming in the model is the alternation parameter. Recall from Table 3 that the passive was signaled by having the activation of the XX unit be 0.25 and the YY unit be 0.5. The alternation parameter determines the difference in these activations; here it is 0.5, meaning that the lesser activated unit, XX, has 50% of the activation of the more activated one, YY. During training, the alternation parameter for situated events was 0.5 half of the time and 0.75 the rest of the time. Using two values of this parameter during training taught the model to use the relative activation level of the two units, rather than the absolute level. This also simulated the natural variation in prominence in different learning situations. When the model was producing target sentences to simulate priming, we reasoned that differences in activation were even smaller; the alternation parameter was set at 0.95. Our reasoning was based on the fact that, in

structural priming experiments, stimuli are chosen that are not strongly biased for one structure (e.g., active) over another (e.g., passive). Diminishing the differences in the activation of the XX, YY, and ZZ units similarly makes the model's messages less biased. It has the effect of making the model more willing to alternate structures—what we call “flippability.” In humans, flippability in structural priming depends on the task (e.g., pictures create more flippability than sentence repetition), the properties of the sentences, the mixture of structures elicited, and the experience of speakers. Because the goal of this work was to understand priming mechanisms rather than accounting for the quantitative details of the data, the values of the learning rate and flippability parameters were not determined anew for each experiment being modeled. Instead, conservatively, they were held constant at levels that led to priming magnitudes that approximated the average magnitude found in all of the experiments. For this reason, the magnitude of priming in the model may not exactly match that found in individual experimental studies.

Consider how a prepositional dative prime, “The mother give -s the orange to the grandma,” affects the description of a dative target during a priming trial in the model. Before the prime is processed, the weights are set to the final adult weights achieved during training. Furthermore, because prime processing is assumed to involve messageless prediction, no message is set. When processing begins, the model predicts what the initial word will be. Assume that it correctly predicts “the” because many sentences begin with “the.” When the initial “the” of the prime is experienced, the error will thus be quite small and there will be little weight change. Now the model tries to predict the next word given “the” as *cword* input. The lack of a message means that the model will almost certainly fail to predict “mother” (although it will likely predict some noun), and hence there will be considerable weight change when “mother” is experienced. Important changes will probably occur at the word “give” and “to,” because these words are strongly linked to the prepositional dative structure. After processing the prime, the weights are updated. For the target part of the trial, the message is given to the model (e.g., A=THROW X=UNCLE Y=BOTTLE Z=AUNT EVSEM: XX=0.5 YY=0.451 ZZ=0.475). Because the ZZ value is higher than the YY value, the model would normally have a slight bias to produce a double-object structure. However, because these values are closer together than they are normally in training, the model is less sure about this choice. After the model produces “The uncle throw -s the,” the weight changes that strengthened the units associated with the



prepositional dative during prime processing increase the likelihood that the Y role is activated next (because the model already has learned that prepositional dative units activate the Y role after the verb), thereby increasing the chance of a prepositional dative.

### Structural Priming as Implicit Learning

The model instantiates the idea that adaptation to recent sentence structures in adults could be due to the same sorts of mechanisms that are used to learn the language initially. The experimental evidence that supports this idea comes from the many studies that have shown persistence of priming in speech (but not writing; cf. Branigan et al., 1999, and Branigan, Pickering, Stewart, & McLean, 2000) over time. Hartsuiker and Kolk (1998) found that priming lasted over a 1-s interval (because lexical activation disappears in milliseconds, this was deemed sufficient to test whether priming was activation). Boyland and Anderson (1998) found that priming lasted over 20 min, and E. M. Saffran and Martin (1997) found that priming was evident a week later for patients with aphasia. In young children, priming from a block of primes can persist over a block of test trials (Brooks & Tomasello, 1999; Huttenlocher et al., 2004).

In two experimental studies, Bock and Griffin (2000; see also Bock et al., 2005) separated primes and targets with a list of intransitive filler sentences (0, 1, 2, 4, or 10 fillers) and found that structural priming was statistically undiminished over these different lags between prime and target. To test whether the model's priming persists over as many as 10 filler sentences, dative and transitive prime–target pairs like those given in Tables 5 and 6 were generated from the input environment grammar. Lags of 0, 4, and 10 were used to separate the prime and the target. The 20 model subjects were tested as described earlier, with learning during the processing of the priming sentence being the sole way to influence the production of the target message. These prime–target pairs were separated by a list of fillers made up of animate and inanimate subject intransitives (e.g., a girl laugh -ed) generated by the input environment grammar (approximating the fillers in the human study). The filler sentences, being isolated unrelated sentences, were processed in the same way as the primes were, as messageless events. The word-sequence output for the target message was coded, and the priming difference between the two prime structures was calculated as described before. These results are presented in Figure 10 along with the corresponding human results from Experiment 2 in Bock and Griffin's summary Table 2 (lag 0 results are averaged from Experiments 1 and 2).

For the analysis, the dependent measure was the percentage of target utterances in the model's output (i.e., active transitive and prepositional dative) out of those that conveyed the target message. The design crossed the factors of sentence type (dative or transitive), prime structure (same or different), and lag (0, 4, 10). The analysis found a single main effect of prime structure, which was due to the fact that the model produced more of the target structures (actives, prepositional datives) when preceded by a prime of the same structure than with primes of the other structure (same prime = 52.3%, different prime = 48.0%),  $F(1, 19) = 45.81, p < .001$ . There was a nonsignificant trend toward an interaction between prime structure and lag,  $F(2, 38) = 3.14, p = .055$ , due to reduced priming between Lag 0 and Lag 10. And if we look at just the Lag 10 position, we find a significant effect of prime type,  $F(1, 19) = 34.14, p < .001$ , and no interaction with target type,  $F(1,$

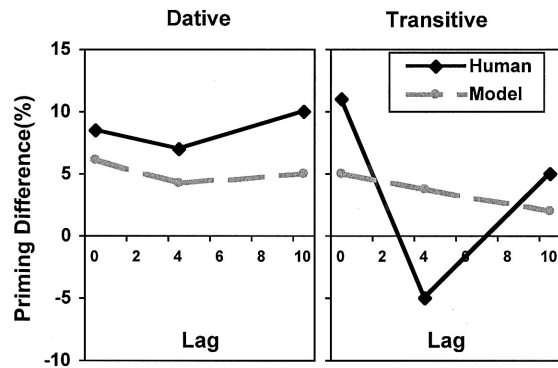


Figure 10. Dative and transitive priming over lag (human results from Bock & Griffin, 2000).

19) = 2.56,  $p = .126$ , demonstrating that priming in the model persists over lag. Note that in the data from Bock and Griffin (2000), there was a large dip in the priming at Lag 4 for transitives, a dip not present in the model's priming. This dip may be a chance event: No significant interaction with lag accompanied the main effect of priming, and a replication of Bock and Griffin by Bock et al. (2005) yielded no dip. Hence, the model and the data agree on the persistence of priming. The lag results in the model demonstrate that the assumptions of learning as processing and of prediction error lead to weight changes that are structure specific, so that learning about intransitives during processing of the fillers does not impact transitive or dative priming. This ability is necessary to explain how priming lasts over lag and also to explain how learning from individual sentences can lead to changes in syntax.

The model instantiates the claim that prediction error in the sequencing system during comprehension of the prime is the basis for structural priming. Therefore, simply comprehending the prime should be as effective as the procedure of hearing and then producing the prime. This was demonstrated in the replication of Bock and Griffin's (2000) lag study by Bock et al. (2005), which used only comprehended primes. Averaged over Lags 0, 4, and 10, there was a similar amount of priming regardless of whether the prime was only comprehended or produced aloud. Figure 11 shows this averaged data compared with the averaged production-to-production priming effect from Bock and Griffin. Because the model's priming is assumed to come from its comprehension of the prime, it accounts well for the equivalence of the priming in these two circumstances. These results support the view that the sequencing system acquires structural representations that serve both comprehension and production (e.g., Bock, 1995; Hartsuiker & Kolk, 2001; Kempen & Harbusch, 2002; MacDonald, 1999; MacKay, 1987; Vigliocco & Hartsuiker, 2002).

### Syntax and Meaning

Data from structural priming experiments constitute key psycholinguistic evidence about the relation between syntax and meaning. If surface syntax and meaning were inextricably linked, we would expect greater overlap in meaning to lead to greater structural priming. However, as we show, for the most part this is not the case. In priming experiments, similarity between the prime and the target in thematic roles, argument status, and transitivity



Figure 11. Comprehension-based versus production-based priming (human results from Bock & Griffin, 2000; Bock et al., 2005).

does not consistently yield priming apart from the priming attributable to the structure alone. The insensitivity of syntax to these factors suggests that some aspects of syntax are isolable from some aspects of meaning. If our model is to capture these findings, its internal states should respond to syntactic similarities without requiring support from meaning similarities. In this section, we review the priming studies that explore the relation between syntax and meaning and test whether the model can reproduce their findings.

Our review of the influence of meaning on structural priming focuses on two contrasting sets of findings. The first one, exhibited in experiments from Bock and Loebell (1990; see also Bock et al., 1992, and Potter & Lombardi, 1998), points to the robustness of structural priming in the face of variation in meaning. These and other studies support the syntactic nature of priming that we referred to earlier. The second type of finding (Chang, Bock, & Goldberg, 2003) is a case of meaning priming the locations of arguments in sentences. In Chang et al. (2003), the priming alternation—the locative alternation—was one in which the two forms had the same surface syntactic structure, and hence surface syntax cannot be the basis of the priming. Thus, we have one set of studies emphasizing that priming does not depend on meaning (e.g., Bock & Loebell, 1990) and another (Chang et al., 2003) demonstrating that it can depend on meaning. The model, we claim, resolves this conflict. We focus first on Bock and Loebell's data and particularly on the effect of the model's XYZ roles in accounting for their data, and then turn to Chang et al. (2003). Finally, we use the model's behavior to state a specific hypothesis about the relation between structural frames and meaning.

In two experiments, Bock and Loebell (1990, Experiments 1 and 2) showed that thematic role overlap did not increase the magnitude of structural priming. In their first experiment, they compared priming from prepositional locatives (e.g., *The wealthy widow drove an old Mercedes to the church*), prepositional datives (e.g., *The wealthy widow gave an old Mercedes to the church*), and double-object datives (e.g., *The wealthy widow sold the church an old Mercedes*). The prepositional dative sentence had a dative verb, which encoded a transfer relationship that required a recipient argument. The prepositional locative had a transitive motion verb, which specified movement to a goal adjunct phrase. Thus, there was a difference in thematic roles (recipient–goal), argument status (argument–adjunct), and verb class (dative–transitive), and these differences can be related to aspects of meaning. If meaning is tightly linked to syntax, then any one of these similarities or differences should contribute to priming. Bock and Loebell used target pictures that elicit dative structures, and so it seems that the similarity of the prepositional dative primes to the dative pictures

in roles, argument status, and verb transitivity should lead to more priming than the prepositional locative primes. When Bock and Loebell tested these structures, however, they found that prepositional locatives primed prepositional datives as much as prepositional datives did (see Figure 12). Because so many aspects of meaning are varied here, it suggests that surface syntax can be isolated from meaning.

To see if the model can produce this result from Bock and Loebell (1990), prime sentences for prepositional dative, double-object dative, and prepositional locatives were generated from the input environment grammar (see Table 7). The dative sentences had transfer dative verbs that required the preposition *to*. The prepositional locatives had transitive verbs, and the locative phrase always used the preposition *to*.

Figure 12 shows the model and human priming effects. Like the human data, the model exhibited priming of prepositional dative responses by prepositional dative primes relative to double-object control primes,  $F(1, 19) = 25.86, p < .001$ . More important, the model (and people) showed reliable priming of prepositional dative responses from the structurally similar, but semantically distinct, prepositional locatives,  $F(1, 19) = 21.55, p < .001$ .

Why are prepositional locatives as effective at priming prepositional datives as prepositional datives themselves? Part of the answer may have to do with the model's XYZ roles. Both the locative and goal prepositional phrases use the Z role. However, this cannot be the whole story because the two sentence types also differ in other ways—for example, their verbs have different subcategorization patterns. The transitive verbs (e.g., *kick*, in Table 7) can occur without the locative phrase (adjuncts are optional). Clearly, the model is able to generalize over these differences and thus gives the appearance of abstracting a single syntactic construction for prepositional locatives and prepositional datives.

Bock and Loebell's (1990) second experiment provided another and arguably stronger test of the hypothesis that priming is insensitive to meaning. The key comparison involved passives (e.g., *The 747 was alerted by the airport's control tower*), intransitive locatives (e.g., *The 747 was landing by the airport's control tower*), and actives (e.g., *The 747 radioed the airport's control tower*). The passives and locatives have similar surface structures, but they differ in meaning-related features such as thematic roles and verb transitivity. In the passive, the verb is transitive, the subject is the patient, and the *by* phrase is an agent, while in the locative, the verb is intransitive, the subject is an agent, and the *by* phrase is a location. Bock and Loebell presented pictures that elicited transitive sentences and found that locatives primed passives as much as passives themselves did (see Figure 13). This suggested that overlap in verb type and roles did not modulate the priming effect, which provides further evidence for isolability of syntax.

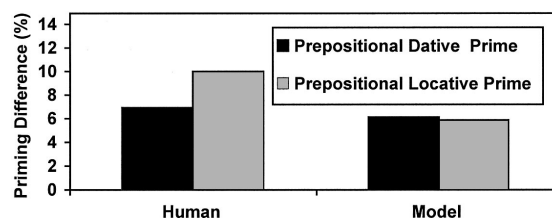


Figure 12. Prepositional locative–dative priming (human results from Bock & Loebell, 1990, Experiment 1).

Table 7  
*Prime and Targets for Bock and Loebell (1990, Experiment 1) Test*

Prime types	Prime sentence	Target message
Prepositional dative	the mother -s give the orange to a grandma.	A=THROW X=UNCLE Y=BOTTLE Z=MAN
Prepositional locative	the mother -s kick the orange to a grandma.	(e.g., the uncle throw -ss the man the bottle)
Double object dative	the mother -s give a grandma the orange.	

The XYZ message does use different roles for passive and locatives (unlike the prepositional locatives and datives, both of which used the Z role for the prepositional phrase). Thus, applying the model in this case provides a good test of whether roles and verb transitivity are inseparable from the model's structural representations. To test this in the model, we used passive transitives, active transitives, and intransitive locatives as primes for transitive messages (Table 8). The results are shown in Figure 13. The model's priming pattern was similar to that of Bock and Loebell's (1990) Experiment 2. Passive responses to the targets were promoted by passive primes more than active primes,  $F(1, 19) = 19.35$ ,  $p < .001$ , and, critically, locative primes also promoted more passive responses,  $F(1, 19) = 14.12$ ,  $p = .001$ . Locatives in the model primed the way that passives do, even though they differ in roles and verb class.

These results suggest that the model has abstracted syntactic structures during the process of mapping from meaning into word sequences. The model's XYZ message format, however, differed from traditional thematic roles in ways that are important for these structural priming phenomena. In traditional-role theories, intransitives as well as transitives distinguish the roles of agent and patient. In the XYZ message, the subjects of intransitives are always coded as the Y role, and this made the subject of intransitive locatives similar to the subject of transitive passives. Also in some traditional approaches, beneficiaries, recipients, goals, and locations would be distinguished as different roles. In the XYZ message, the same role unit (Z) is used for all of them, making the location in prepositional locatives similar to the recipient in prepositional datives.

Because the XYZ role representation appears to support human-like priming, it is important to see whether the model's match to the data is specifically limited to this representational scheme. If the model's results are due to its ability to acquire abstract syntactic mappings—that is, processes that link messages to sequences of words—and not to the nature of the roles it uses, the type of role representation should be irrelevant. Traditional roles should work as well as XYZ roles.

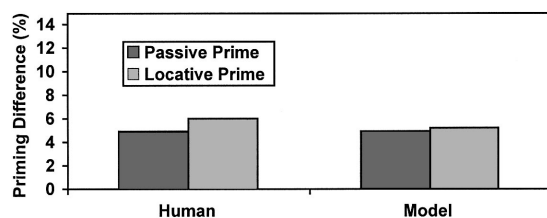


Figure 13. Locative-passive priming (human results from Bock & Loebell, 1990, Experiment 2).

To test this, we generated a version of the model using traditional roles. The traditional-roles message differed from the XYZ message in that (a) the argument of causative intransitives played the same role as the causal argument in transitives (X) and (b) the third arguments of transfer datives, benefactive datives, and other locations were distinguished. This means that the subject of locatives used the same role unit as the subject of active transitives, unlike XYZ messages in which the subjects of locatives use the same role as the subjects of passive transitives. Distinguishing the third (Z) argument of datives in the traditional-roles model means that prepositional datives and prepositional locatives were associated with different sets of roles, and that benefactive and transfer datives, whether prepositional or double object, no longer shared all three roles. In all other ways (training and testing sets, architecture, parameters), the models were identical.

The traditional-roles model was tested on its ability to exhibit the priming patterns in Bock and Loebell's Experiments 1 and 2 (1990), as well as another study (Bock, 1989) that demonstrated that benefactive datives prime transfer datives as much as transfer datives (see the following section for more detail about this study). These studies were chosen for this analysis because their materials exhibited more role overlap in their prime and targets when the XYZ message was used in contrast to the weaker role overlap in the traditional-roles message. Thus, we might expect less or no priming with the traditional-roles message. Twenty models were trained using the same messages as in the original model. In Figure 14, the priming differences are reported for the prepositional locative (prepositional locative with double-object dative control), locative (locative with active control), and benefactive dative (benefactive prepositional dative with double-object dative control) primes for the XYZ message and the traditional-roles message.

The main question is whether the model requires the XYZ message to get priming. It does not. Although the model with XYZ

Table 8  
*Prime and Targets for Bock and Loebell (1990, Experiment 2) Test*

Prime types	Prime sentence	Target message
Passive transitive	a apple is sculpt -par by the woman.	A=PUSH X=MARY Y=MAN (e.g., mary push -ss a man.)
Locative	the woman is walk -ing by a apple.	
Active transitive	the woman sculpt -ss a apple.	



roles led to more overall priming,<sup>3</sup>  $F(1, 19) = 4.98, p = .038$ , the model with traditional roles exhibited priming in each experiment: prepositional locative,  $F(1, 19) = 17.06, p < .001$ ; locative,  $F(1, 19) = 6.45, p = .020$ ; benefactive,  $F(1, 19) = 14.77, p = .001$  (see Figure 14). Because in each of these experiments the roles in the prime and the roles in the target differ, the existence of priming in each case demonstrates that the basis for the model's priming is not role ordering.

Overall, these results suggest that the dual-path model will learn abstract syntactic relationships regardless of which message type is used. This is likely because the message is not available during the comprehension of the prime. Only changes in the sequencing system can influence priming, and so only when roles are represented within the sequencing system will the system show message effects interacting with structural priming.

The comparison of priming with the two role systems suggests that the model learns abstract structural frames that are not simple thematic role ordering schemes. Rather, they have some similarity to surface syntactic structures. This raises the question of whether and how role information influences syntax during production. Chang et al. (2003) addressed this question using the locative alternation, which has the same surface syntactic structure but different role orders in its two versions. One form, the *theme–locative* structure, illustrated in *The man sprayed water on the wall*, puts the theme (*water*) before the location (*wall*). The *locative–theme* structure puts the location before the theme, as in *The man sprayed the wall with water*. Both forms have the same order of surface syntactic categories NP (noun phrase) V (verb) NP PP (prepositional phrase), so differences in priming between the two cannot be explained by priming of these categories or their order. Chang et al. (2003) found differences between theme–locative and locative–theme primes, with theme–locative structures increasing in likelihood after theme–locative relative to locative–theme primes. This supports the idea that role information is entering into the formulation process.

To test this in the model, we generated theme–locative and locative–theme sentences (see Table 9) from the input environment grammar, and the model subjects were tested and coded as before. Figure 15 shows the increase in theme–locative responses associated with theme–locative as opposed to locative–theme primes. The model exhibited the priming effects,  $F(1, 19) = 15.89, p < .001$ , that were present in the human data. This result is important because it shows that the model's (and people's) mapping choices *can* be influenced by role information in the prime, specifically when the alternatives share the same surface syntactic structure but differ in the order of roles. More generally, it suggests that the production system makes use of role information within

the sequencing system when the information is needed for distinguishing alternative mappings from messages to structures.

In summary, the results show that the model can learn syntactic representations that are purely structural and also representations that incorporate meaning. This ability depends on two facets of the dual-path model. One is due to the SRN, which is used during prediction during the prime and also during production. This network prefers to learn representations that just encode surface syntactic categories, and that accounts for the tendency to learn purely structural representations. However, because the model uses error-based learning, if purely structural representations are inadequate for predicting how a sentence should be produced, as in the case with the locative alternation, then the network will learn representations that allow it to distinguish the thematic difference in the alternation.

### Syntax and the Lexicon

Although the preceding analysis of the priming data and the model's account of them suggest that syntactic frames are isolable from meaning, it is possible that syntactic structures are grounded in lexical representations. Many theoretical frameworks project syntax from lexical items (Haegeman, 1994; Pollard & Sag, 1994), adult processing theories make extensive use of the lexicon in syntax (Ferreira, 1996; Garnsey, Pearlmutter, Myers, & Lotocky, 1997; Levett et al., 1999; MacDonald, Pearlmutter, & Seidenberg, 1994; Pickering & Branigan, 1998; Vosse & Kempen, 2000), and language acquisition theories often emphasize the lexical nature of early syntax (Bates & Goodman, 2001; Bowerman, 1976; Braine, 1976; Lieven, Behrens, Speares, & Tomasello, 2003; Lieven, Pine, & Baldwin, 1997; Pine & Lieven, 1997; Theakston, Lieven, Pine, & Rowland, 2001; Tomasello, 1992). The heart of the sequencing system of the dual-path model is an SRN, and these networks have been shown to learn lexically specific representations (e.g., Chang, 2002; Marcus, 1998). Thus, there are good reasons to expect syntactic knowledge to be closely linked to lexical items, both in people and in the model. At the same time, the structural priming literature presents several studies in which lexical overlap is not necessary for structure to transfer from prime to target. The most striking example of such a finding is cross-language structural priming, in which a prime from one language affects a bilingual individual's choice of target structure in another language (Hartsuiker, Pickering, & Veltkamp, 2004; Loebell & Bock, 2003; Meijer & Fox Tree, 2003). The priming requires similar structures between the languages, but it occurs despite no lexical overlap. Results like these suggest that adult syntax must involve some nonlexical abstractions.

To examine this issue, we tested the model in structural priming experiments that probed the relationship between syntactic frames

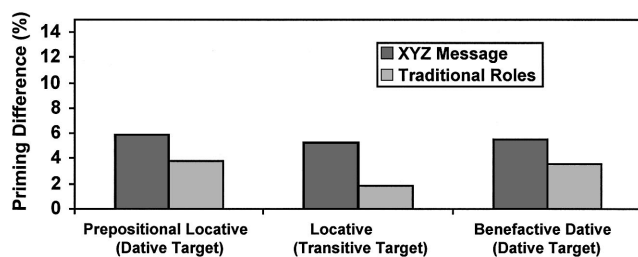


Figure 14. Priming differences for XYZ and traditional-roles message.

<sup>3</sup> We do not take these results as demonstrating that the XYZ roles necessarily prime more than the traditional roles in the model. Recall that the parameters that control overall priming magnitudes, the learning rate and the alternation parameter, were set to values that led to average priming magnitudes similar to those found in the data. This was for the principal model, the one with the XYZ roles. In the test of the traditional-roles version of the model, we used these same values, rather than search for values that optimized the priming level. Hence, the overall magnitude of priming in the models is not necessarily comparable. The key finding is that both versions exhibit priming in these conditions, showing that the model exhibits evidence of abstraction for both kinds of roles.

Table 9  
*Prime and Targets for Chang, Bock, and Goldberg (2003) Test*

Prime types	Prime sentence	Target message
Theme-locative	a mouse spray -ss coffee into the bottle -s.	A=RUB X=MARY Y=CAKE
Locative-theme	a mouse spray -ss the bottle -s with coffee.	D=OVER Z=CUP (e.g., "mary rub -ss the cake over the cup")

and lexical items that normally occur in them. Analogous to the case in which we examined the influence of meaning on structural priming, the key questions are whether priming is found in the absence of lexical overlap and whether this overlap increases priming. In this section, two studies are reviewed and modeled. One was by Pickering and Branigan (1998), who varied the extent to which prime and target verbs and their inflections matched. They found that lexical overlap was not required for priming, and verb inflections did not enhance priming, but the use of the same verb stem in the prime and the target increased priming. Another study by Bock (1989) found that priming in datives was not affected by whether or not the prime and target shared prepositions. These studies suggest that syntactic frames can prime regardless of lexical overlap and that sometimes lexical and morphological overlap does not increase priming.

Pickering and Branigan (1998) used a sentence-completion task that allowed them to control the verb and verb morphology that speakers used for both prime and target. Their first two studies looked at whether verb overlap increased priming, and they found that priming occurred when the prime and target verbs differed but also that verb overlap increased the magnitude of priming. In the next three experiments, they manipulated overlap in tense (present vs. past tense), aspect (completed or progressive), and number (singular or plural agreement) and found that priming was unaffected by having the same morphology. To see if the model's representations treat verbs and morphology in a way that is similar to findings in humans, it was tested in three conditions (see Table 10 for examples): different-verb, same-verb, and same-verb-tense conditions. The different-verb condition was a set of 100 transfer dative messages generated by the input environment grammar where the prime and target had different verbs. The prime sentences were always present tense, and the target sentences were always past tense. The same-verb condition was simply the same message-sentence pairs, except with the prime verb changed so that it was the same as the target verb. The same-verb-tense condition was just the same-verb condition with the prime sentences changed to past tense.

The model's semantic information about verb morphology is represented in the *event-semantics* units. Making this information

available to the sequencing system leads to the possibility that structural frames in the trained model will differ based on tense and aspect, and this might lead to more priming when there is greater morphological overlap, contrary to the human data. If the model does not show more priming when verbs share tense, it would suggest that it can learn to isolate syntactic frames from morphology and that learning plays a role in deciding what features inhabit the frames.

The human and model results are illustrated in Figure 16. The model's dependent measure is the average difference between priming conditions in the percentage of prepositional datives out of prepositional dative and double-object datives. Pickering and Branigan's results were originally presented as the percentage of prepositional dative structures out of all structures (prepositional datives, double-object datives, others) and the percentage of double-object datives out of all structures. To make the experiments and the model more comparable, the "other" items were excluded, and Pickering and Branigan's (1998) results were converted into percentages of prepositional datives out of prepositional datives and double-object datives.<sup>4</sup> The priming difference between prepositional dative and double-object dative priming conditions was computed and used as the dependent measure in Figure 16.

There was significant priming in the model,  $F(1, 19) = 24.01$ ,  $p < .001$ , but priming did not interact with type of overlap,  $F(2, 38) < 1$ ,  $p = .408$ . The lack of an interaction of priming with overlap suggests that the model is insensitive to both verb and morphological overlap. This insensitivity is different from the human data. Although Pickering and Branigan (1998) found a lack of sensitivity to overlapping verb inflections (like the model), they found considerable enhancement of priming when the verb itself was the same (unlike the model). We return to this discrepancy later.

Another example of priming's being insensitive to the identity of closed class elements involves the influence of prepositions in dative priming. Bock (1989) compared transfer dative primes, which mark the goal with the preposition "to" (e.g., *A cheerleader offered a seat to her friend*), and beneficiary dative primes, which use "for" (e.g., *A cheerleader saved a seat for her friend*). The targets were all transfer datives. The finding (see Figure 17) was that both kinds of datives lead to reliable priming of approximately the same amount. Hence, whether the prepositions themselves

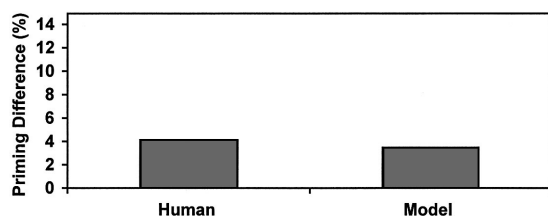


Figure 15. Locative alternation priming (human results from Chang, Bock, & Goldberg, 2003).

<sup>4</sup> Same- and different-verb results came from Table 1 in Pickering and Branigan (1998). Same-tense-verb results came from Table 3. To make the model and human results comparable, we converted the human results so that they excluded utterances that were classified as "other." Same verb =  $.47/((.47 + .22) - .29)/(.29 - .38) = .25$ , Different verb =  $.40/((.40 + .25) - .35)/(.35 - .29) = .07$ , Same verb tense =  $.50/((.50 + .19) - .34)/(.34 + .32) = .20$ .

Table 10  
*Prime and Targets for Pickering and Branigan (1998) Test*

Overlap types	Prime types	Prime sentence	Target message
Different verb	PD	the aunt -s lend the apple -s to the man.	A=PASS
	DO	the aunt -s lend the man the apple -s.	X=BROTHER Y=ORANGE
Same verb	PD	the aunt -s pass the apple -s to the man.	Z=GRANDMA
	DO	the aunt -s pass the man the apple -s.	EVSEM=PAST (e.g., a brother pass -ed the grandma the orange)
Same verb tense	PD	the aunt -s pass -ed the apple -s to the man.	
	DO	a brother pass -ed the grandma -s the orange.	

Note. PD = prepositional dative; DO = double object dative.

match between prime and target was not relevant. To test for this effect in the model, we used the different-verb sentences from the previous test (because the verbs in this condition's prime and target were both required to be transfer datives), and the *to* dative primes were changed into *for* datives by linking the action A role to a benefactive verb and the preposition-semantics D role to the semantics for the preposition *for*. The benefactive verbs were required to be verbs that allow the double-object structure. Prepositional dative and double-object versions of both verbs were tested. Table 11 shows example prime–target pairs.

The model showed priming with both transfer (*to*) datives and beneficiary (*for*) datives,  $F(1, 19) = 23.13, p < .001$ , and no interaction of prime and verb type,  $F(1, 19) = 1.77, p = .200$  (see Figure 17). The results suggest that the model abstracts representations that are insensitive to lexical items such as prepositions.

To summarize, the model learned to produce sentences with the appropriate words and morphology, but the syntactic representations that support these behaviors are, to a large extent, abstracted away from particular lexical items. Although there are many routes for lexical information to influence syntax in the model (e.g., *cword* → *ccompress* → *hidden*, *hidden* → *compress* → *word*), the model also seems to be learning some representations that operate independently of these items. This comports with views that syntax has some abstract elements that do not have lexical content (e.g., the double-object construction; Goldberg, 1995). The model's behavior arises from the dual-pathway assumption, which keeps syntax from incorporating too much lexical information, and the sequencing-by-SRN assumption, which reduces the influence of the lexicon on the *hidden* units by using compression units.

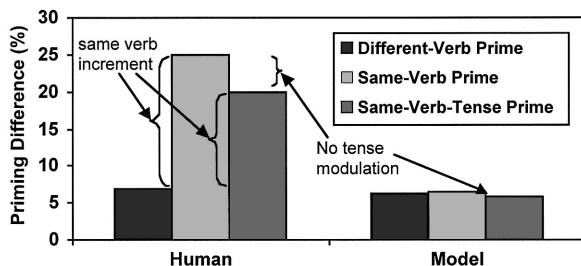


Figure 16. Verb and morphology overlap (human results from Pickering & Branigan, 1998, Experiments 1, 2, and 3).

### Understanding the Mechanism

In a complex architecture like the dual-path model, syntactic knowledge is distributed in weights between different layers. To understand the priming mechanism behind a particular experiment, one has to determine the appropriate layers that are involved. In this section, we are interested in determining the source of locative-passive priming (e.g., “The 747 was landing by the control tower”), because this priming effect is more abstract and difficult to understand (Bock & Loebell, 1990).

As an overview, we first tracked the difference in the activation of the agent and patient *where* units (which we refer to as the *agent-bias score*) during the production of active and passive transitive targets after active and locative primes. This assessed whether priming has to influence role activation in the subject position in order to select structures. To see how the *hidden* units affect the agent and patient *where* units, we collected an *influence score* for each *hidden* unit. We then collected the *hidden* unit activation difference between the active and locative primes (this is called the *prime difference score*). By multiplying the influence score by the prime difference score, we get the *weighted prime difference score*, which tells us which *hidden* units strongly influence priming. We now describe each of these steps in detail.

To start, we must first see how the model produces active versus passive structures during target production. Chang (2002) found that the activation of the *where* units (corresponding to thematic roles) at choice points was related to the structure actually produced. For this reason, the activation of the *where* units X (agent in transitives) and Y (patient in transitives) was recorded during the production of all the target sentences in the active-locative test set. These average activations were computed for each position in active and passive target sentences for an individual model. For

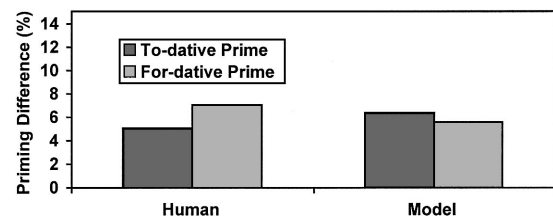


Figure 17. Transfer-benefactive dative priming (human results from Bock, 1989, Experiment 1).

Table 11  
*Prime and Targets for Bock (1989) Test*

Prime types	Prime sentence	Target message
Prepositional	sally send -ss the orange.	A=SELL X=BIRD Y=CUP Z=AUNT
Double object <i>to</i> dative	sally send -ss the mouse the orange.	EV=PAST (e.g., the bird sell -ed the cup to the aunt)
Prepositional <i>for</i> dative	sally carve -ss the orange for the mouse.	
Double object <i>for</i> dative	sally carve -ss the mouse the orange.	

each structure, the difference between the average X unit and the average Y unit activation was computed, creating the agent-bias score. Figure 18 shows that active sentences had higher average agent-bias scores during the production of the subject (e.g., “The brother”) than passive sentences had during the production of their subjects (e.g., “The grandpa”). The figure also shows that the agent-bias score is higher for passives when the *by* phrase is being produced (e.g., “by the brother”). These results suggest that the activation of the *where* units tracks the structure of the sentence. This is not surprising, because the *where* units encode the lexical semantics information that must be produced at these points in the sentences, and the model was designed to do incremental production in a way that is sensitive to lexical selection (Chang, 2002). Therefore, for priming to have the appropriate influence on the target, it has to influence subject selection at the relevant sentence positions (the first and second word positions, Tar0 and Tar1 in the figure).

To understand how the prime influences the activation of *where* units during target production, we need to understand how the *hidden* units activate the *where* units. Because the *where-what* weights are not set during the prime, the weights from the *hidden* units to the *where* units were not changed by back-propagation; so to record the influence of these weights, we created an influence score for each *hidden* unit by subtracting its weight for the patient link from its weight for the agent link. A positive value for the influence score meant that the particular *hidden* unit activated the agent *where* unit more than the patient unit, and vice versa for a negative value.

Knowing how much each *hidden* unit influenced the *where* units, we needed to see how much each *hidden* unit was changed by learning from the primes. Therefore we calculated, separately for each prime, the average *hidden* unit activations for each of the

40 *hidden* units during each position in the target sentence. From the average activation of the *hidden* units after an active prime we subtracted the average activation after a locative prime to create a prime difference score (positive values mean that the unit was more activated for actives than locatives).

Because the prime difference scores measure the change in the *hidden* units due to the primes, and the influence score records the influence of each *hidden* unit on the *where* units, multiplying them together indicates which *hidden* units influence the *where* unit activations. This is the weighted prime difference score. The weighted prime difference shows which *hidden* units influence priming during the production of the target subject. To understand the relationship between the weighted prime scores and priming behavior overall, 3 of the 20 instances of the model were tested. We chose three models because the models differ in how sensitive they are to a particular kind of priming. For locative-passive priming, there was considerable individual-model variability. The first model showed strong locative-passive priming (24.7%; this model was also used for Figure 18); the second showed an average level of priming (5.7%); and the last showed almost no priming (0.25%). For each model, the six *hidden* units with the strongest weighted prime scores during production of the subject in the target sentence (see Tar0, Tar1 in Figure 18) are displayed in Figure 19. The *hidden* units are listed in the legend of Figure 19 in terms of arbitrary identification numbers. Because most subjects are only two words (e.g., “the brother”), the influence from *hidden* units that are sensitive to priming is normally gone by Position 4 in the sentence (Tar3), and so the figure extends only to this position.

The weighted prime difference encodes how much a particular *hidden* unit contributes to locative-passive priming. When locative-passive priming was strong (see the top panel of Figure 19), there were two *hidden* units (4, 19) with strong activations during the subject position. When priming was close to average (middle panel), only one *hidden* unit (39) captured this relationship. Finally, when priming was close to zero (bottom panel), no *hidden* units exhibited a strong weighted prime difference.

This analysis shows that locative-passive priming is mainly due to changes in weights involving very few *hidden* units. Sometimes the priming is concentrated in a single unit. This concentration helps to explain why the model shows abstract structural priming. The learning process integrates semantic, lexical, and structural information from the whole sentence into individual units, so the units are no longer controlled by any single source of information (i.e., they are abstract). Of course, the existence of abstract units does not preclude the existence of units that respond primarily to lexical or semantic influences. Our claim is only that the model’s behavior implies that at least some of its components abstract over

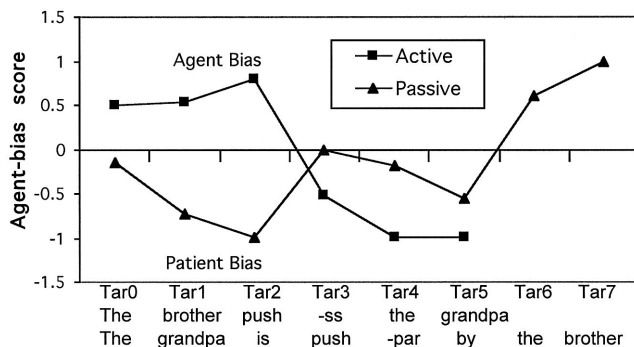


Figure 18. Agent-bias score in role units at each point during target (Tar) sentence production.



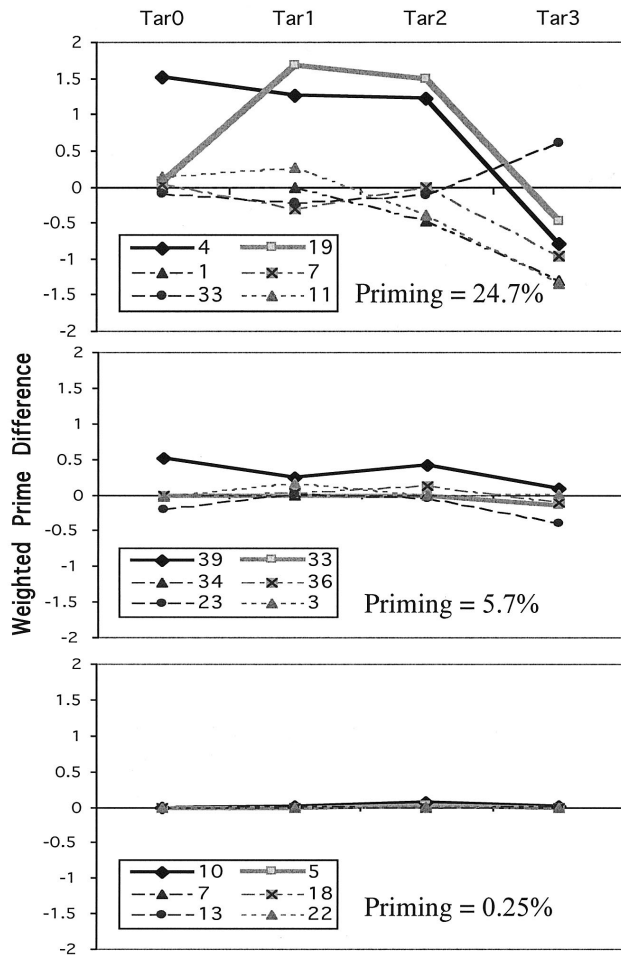


Figure 19. Weighted prime difference of *hidden* units early in sentence (Tar0 – Tar3). Tar = position in target utterance.

lexical and semantic information, helping to make the influence of syntax isolable. The *hidden* units 4, 19, and 39 are examples of such components.

### Structural Priming Conclusion

The dual-path model assumes that learning continues into adulthood (the learning-as-processing assumption). Learning depends on the difference between expectations and predictions (prediction error assumption) and, together with the model's architecture (the dual-pathway and sequencing-by-SRN assumptions), these assumptions allow syntactic representations to be learned that vary in their dependence on meaning. In short, the model explains abstract syntactic generalization.

To understand the present approach, it is useful to compare the dual-path model with the approach to structural priming introduced by Pickering and Branigan (1998). Their approach extends network approaches that have been used to explain word production (Dell, 1986; Levelt, 1989; Levelt et al., 1999) to structure selection. Pickering and Branigan's treatment of structure selection shares many features with the dual-path model. Both use spreading activation within a network to instantiate sentence planning and use mechanisms that create sensitivity to recent use of a structure

(e.g., nodes and links change with experience in Pickering and Branigan's, 1998, account). Both use abstract syntactic units that cannot be simply reduced to lexical entries (e.g., combinatorial nodes such as NP,NP or NP,PP in Pickering and Branigan's, 1998, theory). The main difference between the two approaches is that the dual-path model must learn its representations. Structural priming is a consequence of assumptions that are required for the acquisition of sentence production and processing skills. These assumptions yield three empirical consequences that do not directly follow from Pickering and Branigan's approach.

The first consequence is that the dual-path model's learning assumptions lead one to expect priming effects that, in certain circumstances, involve thematic role distinctions. Chang et al.'s (2003) demonstration that the order of theme and locative constituents can be primed constitutes a demonstration of such an effect (see also Griffin & Weinstein-Tull, 2003). Insofar as Pickering and Branigan's (1998) approach focuses on purely syntactic nodes and their links to lexical items, this kind of priming is unexplained.

The second consequence arises from the dual-path model's use of error-based learning. The weight changes that realize learning are greater to the extent that predictions are incorrect. Error-based learning, as opposed to Hebbian or correlation-based learning, appears to be needed for learning complex mappings such as those found in language (Dell, Schwartz, et al., 1997; Elman, 1990). Aside from the utility and (we argue) necessity of using error to guide learning, there is direct evidence for the prediction-error account of structural priming from the priming data themselves. It has been noted (e.g., Bock & Griffin, 2000; Ferreira, 2003; Hartsuiker & Kolk, 1998; Hartsuiker, Kolk, & Huiskamp, 1999; Hartsuiker & Westenberg, 2000; Scheepers, 2003) that prime structures that are less common are more effective primes than those that are more common. Particularly strong evidence for this comes from a study by Ferreira (2005), which used prime sentences where the main clause structure either demanded the normally disfavored reduced embedded clause or permitted it as a disfavored option. That is, the priming structure was the same, but it occurred in different contexts that made it more or less preferred. Ferreira (2005) found that the magnitude of the resulting priming was stronger when the prime's main clause structure allowed but did not require the reduced structure, demonstrating that it is not the structure itself but its context of occurrence that matters to the strength of priming. Similar results were obtained for unreduced structures in other contexts. Although our model does not have multiclaused embedded sentences and therefore cannot simulate these results directly, it provides a natural explanation for this pattern. Primes that contained the disfavored type of embedded clause would be associated with greater prediction error and hence more weight change. This leads to a relatively greater bias toward the disfavored structure when the target sentence is produced.

The final consequence of the model's assumptions in comparison to the Pickering and Branigan (1998) approach concerns the relationship between particular lexical items or morphemes and the syntax in structural priming. The empirical data show that repeating verb morphology and prepositions in the prime and target did not increase priming but that repeating a verb did. Additional studies have demonstrated that repeated verbs (Branigan et al., 2000) and nouns (Cleland & Pickering, 2003) increase structural priming. Pickering and Branigan's account of priming suggests that any morpheme with a lexical entry (a lemma) and a link to combinatorial nodes (e.g., NP,PP) should lead to an enhancement

of priming. That is, the structural effect should be greater if the prime and the target both include the same lexical item linked to the same combinatorial node.

The dual-path model, as we mentioned earlier, does not exhibit increased structural priming when there is lexical or morphological overlap. This is accurate with respect to data about overlap in function words and verb morphology, but it is at odds with the data concerning repeated verbs or nouns, which do enhance priming. How do we explain the model's discrepancy with the data? We hypothesize that lexical enhancement of priming is not due to the weight-change mechanisms that lead to long-lasting structural priming. Rather, they are due to explicit memory for the wording of the prime. When the target is being planned, the repeated content word serves as a cue to the memory of the prime and this biases the speaker to repeat its structure. This explicit memory component to priming is distinct from the model's weight-change mechanism.

We recognize that this account of the lexical enhancement effect is post hoc and, hence, requires additional justification. Consequently, we offer three sets of predictions concerning the relationship between the long-term priming mechanism and the hypothesized influence of explicit memory on priming. For two of the three sets of predictions, there exists at least some supportive data.

First, we know that explicit memory for the wording of sentences decays very quickly (e.g., Levelt & Kelter, 1982; Sachs, 1967). If this is the mechanism for the lexical enhancement effect, the enhancement should be present for prime–target lags of 0 or perhaps 1, but not for longer lags. Konopka and Bock (2005) examined structural priming for verb–particle constructions (e.g., “lace the boot up” vs. “lace up the boot”). Primes contained the same or different verbs as the intended target sentences and were associated with lags of 0, 1, 2, or 3 sentences. At Lag 0, there was a strong lexical enhancement of priming; primes with the same verb led to increased use of the prime's structure. At longer lags, the lexical enhancement effect disappeared and all that remained was a structural priming effect. The structural priming was as large as that at Lag 0. These results are consistent with the long-lasting structural priming expected from the model, and a short-lived lexically based enhancement effect. Consequently, the lexical enhancement effect is likely not due to the mechanisms proposed for the model, and its existence is orthogonal rather than contrary to the model.

Second, one can ask whether explicit episodic memory is generally responsible for structural priming, contrary to our implicit learning hypothesis. Explicit and implicit memory data from Bock et al. (1992) analyzed by Bock (1990) showed that explicit memory for primes was uncorrelated with priming effects: The conditional probability of priming given later, explicit recognition of the form of the priming sentence was exactly the same as the overall conditional probability of priming, .29. Likewise, the conditional probability of explicitly recognizing a sentence's form given that it had elicited priming was almost identical to the overall conditional probability of form recognition (.66 and .67). Of course, these conditional probabilities are not particularly strong evidence about the association between structural priming and memory, because the relevant dependent measures are noisy at the individual level. Ferreira, Bock, Wilson, and Cohen (2005) provided stronger evidence that long-lasting structural priming is not due to explicit memory. Patients with amnesia who have little explicit memory

for any but the most recent events had normal levels of structural priming for up to 10 prime–target lags. This result is a direct prediction from the implicit learning hypothesis that forms the basis of the model.

Finally, the hypothesis that lexical enhancement of priming is due to explicit short-term memory for the prime can be tested by manipulating variables that affect explicit memory. We hypothesize that lexical enhancement occurs for verbs and nouns, but not function morphemes, because the latter are not particularly effective retrieval cues. A target sentence with “of” is not going to remind one of a previous sentence with “of”, but a target sentence that repeats “throw” or “ball” might. Only by making function morphemes into effective cues should there be lexical enhancement, at least at short prime–target intervals. For example, asking participants to detect repetition of prepositions (instead of repetition of whole sentences as in running recognition procedures) or using unusual prepositions (“The man sat athwart the chair”) could promote greater structural priming when the material is repeated in the prime and target.

The lexical enhancement of priming from content words is the only significant deviation between the model and priming data that we know of. Consequently, we have taken some pains to articulate an account of the effect, to make that account concrete by identifying predictions from it, and to make it plausible by citing data consistent with some of the predictions. We acknowledge, though, that there is a deep unresolved issue here. It is in the nature of the model to keep its learning about structures somewhat separate from particular lexical items. This is required to achieve a production system that generalizes effectively and creates structural priming. A large lexical enhancement of such priming suggests a system in which structural and lexical information are not so separate. For now, we suggest that the lexical–structural interaction reflects short-term bindings that are not as durably represented in the production system.

## Language Acquisition

The facts of structural priming suggest that abstract syntactic representations or processes are changed by experience with language. The model's account of these facts is based on language acquisition in two respects. First, the mechanism of priming is the same error-based learning algorithm that is used to acquire language in the first place. Second, the structural representations that are primed arise through the model's developmental process and thus reflect how its learning algorithm interacts with its other assumptions, most important its dual-pathway and what–where assumptions. In this section, we provide an explicit account of how learned syntactic representations can explain differences in different syntax acquisition tasks. We suggest that the absence of explicit task models has led to the appearance of conflict in the data and confusion about the implications of the data for theory.

To address these issues, we applied the dual-path model to a debate about how the transitive construction develops in young children. There is general agreement that by 3 years of age, children have an abstract transitive construction, as indexed by the ability to easily combine novel verbs with transitive structures in production. The question then is about the nature of the transitive construction before that time. Tomasello (2000) reviewed a variety of studies, mainly of production, suggesting that before the age of

3, children are conservative in their ability to generalize the transitive construction. In particular, they seem to be more willing to use verb-structure pairings that they had previously experienced, and less willing to use novel verb-structure pairings (Abbot-Smith, Lieven, & Tomasello, 2001; Akhtar, 1999; Akhtar & Tomasello, 1997). These results suggest that early item-specific transitive representations develop only gradually into abstract transitive representations and, more generally, that late-syntax approaches to language acquisition, which de-emphasize innate factors, are true (Bates & Goodman, 2001; Lieven et al., 2003, 1997; MacWhinney, 1987; Tomasello, 2003).

Although children seem to be conservative in these production tasks, they also seem to early exhibit knowledge of the transitive construction if their abilities are assessed in preferential-looking tasks (Hirsh-Pasek & Golinkoff, 1996; Naigles, 1990). Preferential-looking tasks measure which of two actions children prefer to look at when a particular sentence is heard. For example, a causative action might be presented on one of two video monitors, showing one individual acting on another individual. The other monitor displays an action in which the same individuals do something by themselves (called here the *noncausative action*). When children hear a transitive sentence with either a novel or a known verb and the individuals mentioned in the sentence match the individuals in the video, children as young as 25 months tend to look at the causative action video. Because the causative action video is the one that best matches the meaning of the transitive construction, children's preference for this video suggests that they have an abstract transitive construction that helps them to infer the relationship to the appropriate video. This task thus provides evidence for early-syntax approaches that posit some syntax-specific tendency to learn abstract structure (Fisher, 2002a; Gleitman, 1990; Naigles, 2002).

The early- and late-syntax approaches are clearly at odds. Early-syntax approaches view the preferential-looking results as providing evidence of abstract syntactic knowledge in children by around age 2 and argue that children are conservative with novel verbs in production because they must understand the meaning of the novel verb as well as link the verb to a syntactic structure (Fisher, 1996, 2002a; Fisher, Gleitman, & Gleitman, 1991; Naigles, 2002). The late-syntax approaches treat the production results as a more valid index of syntactic knowledge and regard the preferential-looking results as due to an earlier ability to use partial knowledge to bias looking in a forced-choice task (Tomasello & Abbot-Smith, 2002). Because both theories can explain elicited production and preferential looking, the disparities in the results do not selectively undermine either approach (Fisher, 2002a; Naigles, 2003; Tomasello & Abbot-Smith, 2002; Tomasello & Akhtar, 2003). This makes the theories hard to falsify with the kinds of tests that are available.

Our view is that the dual-path model can help to address this debate, because it can learn abstract representations using a general-purpose learning mechanism. Therefore, it can address both data sets in a way that might be compatible with late-syntax approaches. But it also develops representations in the SRN that record sentence-position-specific syntax-semantic links, and these links might allow it to exhibit preferences earlier in preferential looking than production. Thus, it has architectural constraints on learning that make it compatible with early-syntax approaches.

### *Modeling Language Acquisition Tasks*

To model the acquisition data, we needed to simulate the tasks. The elicited production task was easy. The dual-path model was simply asked to produce transitive sentences with novel verbs, which is basically what the children must do. For preferential looking, the model had to take the same novel-verb transitive sentences and see if they matched a causative meaning. The model was tested every 2,000 epochs to see how both of the measures changed over time. Furthermore, to equate the materials, the same transitive sentences were used in both tasks. Forty transitive progressive-aspect messages with proper name arguments were generated, and then the action information was replaced with a novel verb (e.g., *Marty is glorp-ing Mary*).

As a test of the novel-verb production, the model was given each of the 40 *glorp* messages to produce. Because *glorp* is a novel verb, the model cannot produce sentences with *glorp* unless it learns to map from the action meaning of *glorp* to the lexical representation of *glorp*. To do this, we set the weight between the *what* unit GLOP and the *word* unit *glorp* to a strong value of 15 (and the corresponding weight from *cword glorp* to *cwhat GLOP* was also set). These links only allow the novel word to be produced and recognized and do not influence syntax (just as in the human experiments, where learning the word *glorp* does not teach the child the syntax of the transitive). Rather, the model's syntactic knowledge has to come from the knowledge that it accumulated during learning of the language from sentences with real verbs. The dependent measure for this task was the percentage of correct transitive sentences produced. The same 20 models that were used for the structural priming simulations were tested on these sentences, and the average number of messages correctly produced can be seen by referring back to Figure 8.

The model's ability to produce transitive sentences with novel verbs grows gradually over time. The overall pattern of the model is similar to what has been found with children. In Tomasello (2000), productive transitives constituted approximately 30% of children's responses at 36 months and 65% at around 48 months (approximated from Tomasello's, 2000, Figure 3). The model reaches 32% correct at 12,000 epochs, and 70% at 20,000 epochs (see Figure 20). Given these results, we can use the production results to calibrate the model's "age." The model reaches its third birthday at Epoch 12,000 and its fourth birthday at 20,000, so a period of 12 human months approximates 8,000 epochs. This approximation is used later to compare development on other measures. The model's gradual development of an abstract transitive construction in production stems from its experience with specific sentence-message pairs.

To simulate preferential looking, it is necessary to understand the mechanisms behind performance of the task. One way to think of preferential looking is that the child looks at the event display that provides the best information for predicting the structure of the sentence. Prediction of sentence sequences is essentially what the dual-path model does. The model uses all of its internal representations to predict the next word in the sentence, and the difference between its prediction and the actual experienced word (the error difference) is used to adjust the weights. Therefore, prediction error represents the compatibility of the model's internal representations, including its message, with the actual sentence. Prediction error summed across the sentence is thus a measure of the model's preference for a particular event-sentence



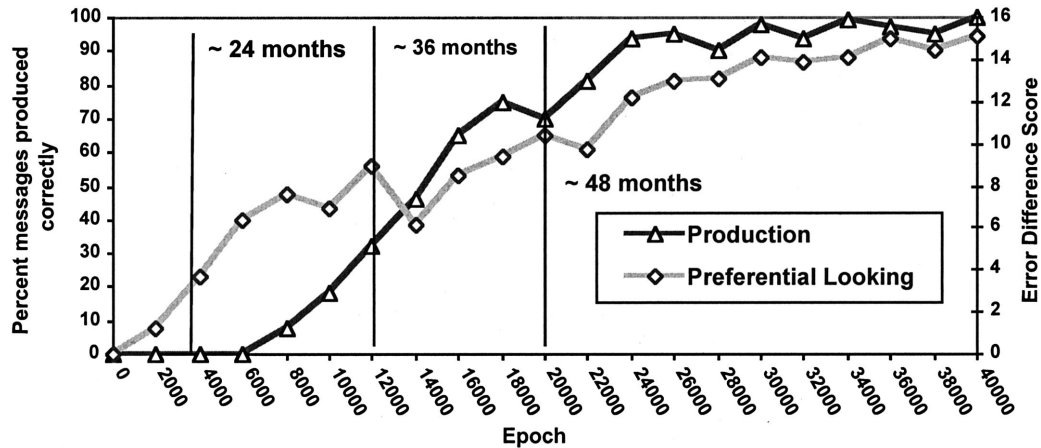


Figure 20. Production and preferential looking with novel transitives during development.

pairing, with lower error meaning better matching. Because the task uses two events, we presented the stimulus sentence twice (once with matching and once with mismatching meaning; see Table 12 for an example of the message–sentence pairings), recording the error for the sentence during the processing of the message–sentence pair. To determine the preference for a particular sentence, we subtracted the matching-pair error from the mismatching-pair error (because lower error equals higher preference). The subtraction yielded a positive number (the error difference score) when the matching pair had a lower error than the mismatching pair.

The preferential-looking results of the model for novel transitives are shown in Figure 20, compared with its production of the same novel transitives. In preferential looking, the transitive match preference score starts to rise from the beginning of learning, that is, before the model's second birthday at 4,000 epochs, using the assignment of epochs to ages that we previously described for the model's production behavior. (Note that Epoch 0 is not age 0, because the model begins training with input, e.g., already segmented into words.) This demonstrates that the model's meaning-based prediction approach to preferential looking exhibits data consistent with the observation that transitive knowledge in preferential looking precedes transitive knowledge in production. However, what the model knows when performing both tasks is exactly the same.

The model also sheds light on why children's performance on related structures advances unevenly. In preferential-looking tasks, conjoined-noun-phrase intransitives (e.g., *The duck and the rabbit*

*are daking*) and *with* intransitives (e.g., *The duck is daking with the rabbit*) have been used in addition to transitives. Whereas novel transitives are correctly and consistently associated with causative meanings by children after 23 months, performance with conjoined-noun-phrase intransitives and *with* intransitives by children around this age is much more variable.

The variability is well documented. Several studies have tested the transitive and the conjoined-subject intransitive with novel verbs (Hirsh-Pasek & Golinkoff, 1996; Naigles, 1990), although only three studies have used novel or low-frequency verbs with the *with* intransitive (Bavin & Growcott, 2000; Hirsh-Pasek & Golinkoff, 1996; Kidd, Bavin, & Rhodes, 2001). Hirsh-Pasek and Golinkoff (1996, pp. 144–148) found that 28-month-olds correctly preferred the noncausative event for intransitives, whereas in the 23-month-olds, the results were inconsistent: Boys actually preferred the *nonmatching* causative video. The other two studies used a within-subject design that allowed for comparison between transitives and *with* intransitives. Kidd et al. tested 30-month-old children and found a transitive bias for causative actions with novel verbs, but only a nonsignificant *with* intransitive bias for noncausative actions. Bavin and Growcott (2000) tested 27-month-olds and found both a significant transitive-causative and a *with* intransitive-noncausative bias. Thus, from 25 months, children know that transitives are associated with causative meanings but are less sure that *with* intransitives (and conjoined-subject intransitives) are associated with noncausative meanings. These results are important, because intransitives and transitives are structurally related in many linguistic frameworks, with intransitives (SUBJECT VERB) being a proper subset of transitives (SUBJECT VERB OBJECT). If children use preexisting structural sensitivities to learn about novel verbs, it is strange that they have more trouble with some kinds of structural knowledge than others.

To understand why the differences between transitives and intransitives might exist, we tested the model in its analog to the preferential-looking task with both kinds of structures. Because the model does not know conjoined noun phrases, only the *with* intransitives were used to represent intransitive performance. The transitive-*glorp* sentences from the production test were changed so that they would generate *with* intransitives. The model generates these structures when role Y represents the intransitive subject, Z represents *with*-adjunct information, and the D role (the *where* unit

Table 12  
Preferential Looking Message-Sentence Pairs

Structure match type	Message-sentence pair
Transitive match	A=GLORP X=MARTY Y=MARY Marty is glorp -ing Mary.
Transitive mismatch	A=GLORP Y=MARTY Z=MARY D=WITH Marty is glorp -ing Mary.
Intransitive match	A=GLORP Y=MARTY Z=MARY D=WITH Marty is glorp -ing with Mary.
Intransitive mismatch	A=GLORP X=MARTY Y=MARY Marty is glorp -ing with Mary.



linked to the semantics of the preposition) is set to the meaning of *with*. In this way, we created a matched set of transitive and *with* intransitive sentences paired with the same nouns and a novel verb. Each message was then paired with each sentence so that the message and the sentence either matched (e.g., causative message and transitive structure) or mismatched (e.g., *with* intransitive-noncausative message and transitive structure). Table 12 shows an example.

The preferential-looking results of the model for intransitives compared with transitives are shown in Figure 21 (the transitive results shown are the same ones displayed in Figure 20). Before the model reaches age 3, its intransitive-noncausative match preference is weaker than its robust transitive-causative match preference, just as in the human data. To understand why this is, we need to look carefully at how the model accomplishes the task.

### Explaining the Model's Behavior

The model's account of acquisition phenomena, like its account of priming, makes use of its incremental processing system. One key finding, the fact that the same representations in the model led to sensitivity in preferential looking earlier than in production, arises simply from the nature of these dependent measures. Production is coded at the level of a whole utterance, which requires a sequence of correct decisions. Preferential looking, on the other hand, can be seen as a series of forced choices between two monitors spread over an interval of time. Hence, partial form-meaning matches can be more useful in this task.

The other key finding is that robust preferential looking appears earlier for transitives than for intransitives. The model's account of this result is less transparent, and so we need to examine the model's preferences as it does incremental word prediction. To see how preferences change during the processing of a test sentence, a *message-bias score* for both the transitive and the *with* intransitive sentence was calculated at each position in a sentence. This message-bias score reflects the bias of the message toward a causative interpretation. It was computed at each position by taking the error that was generated when the sentence was paired with a noncausative message and subtracting the error that was

generated when the sentence was paired with a causative message. When error was lower, the message-bias score was positive, and the model preferred to pair the sentence with a causative message; when the message-bias score was negative, the model preferred to pair the sentence with the noncausative message.

It is useful to look at the message-bias scores at Epoch 18,000 first, where the model exhibits strong transitive and intransitive match preference (see Figure 22, top panel). The patterns for transitive and intransitive sentences are the same for the early part of the sentence, because these structures have the same initial word sequence ("Marty is glorp—ing. . ."). At the position after *glorp—ing*, the model prefers the causative message when it receives *Mary* (message-bias score becomes positive), and it prefers the noncausative message when it receives *with* (message-bias score becomes negative). Because these are also the cues that structurally distinguish the transitive and the intransitive sentences, the model seems to be sensitive to the structural features of the input at Epoch 18,000. This is not surprising, because the model is already able to produce transitives with novel nouns. We refer to the ability to use the words after the verb to bias toward one of the messages the *postverbal structural bias*.

The message-bias at Epoch 6,000 tells a different story (see Figure 22, middle panel). As with the model at Epoch 18,000, the model at this time has the postverbal structural bias. However, in addition, it has a bias for the causative message at the early points in the sentence (e.g., at *Marty* in Figure 22, middle panel). Because both the transitive and intransitive sentences have the same preverbal lexical material, this early bias for causative messages must be due to the difference in the messages. The likely cause is the model's greater facility with mapping the agent of causative messages ( $X = \text{MARTY}$ ) to subject position relative to its ability to map the agent of intransitives ( $Y = \text{MARTY}$ ). One reason that the model has this difference in its ability to map the  $X$  and  $Y$  roles to subject comes from the fact that the  $X$  role maps to subject in many constructions in the model's language (transitives, datives, locative alternation, change of state, benefactives), whereas the  $Y$  role maps to subjects only in intransitives and passives. Evidence from child-directed speech suggests that children receive more than 3 times as many transitives as intransitives (Cameron-Faulkner et al.,

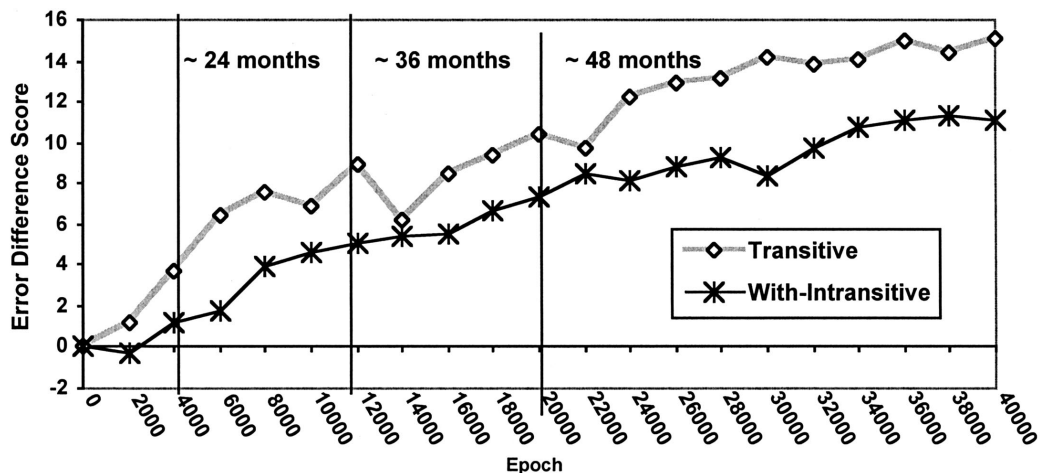


Figure 21. Preferential-looking results for transitives and *with* intransitives during development.

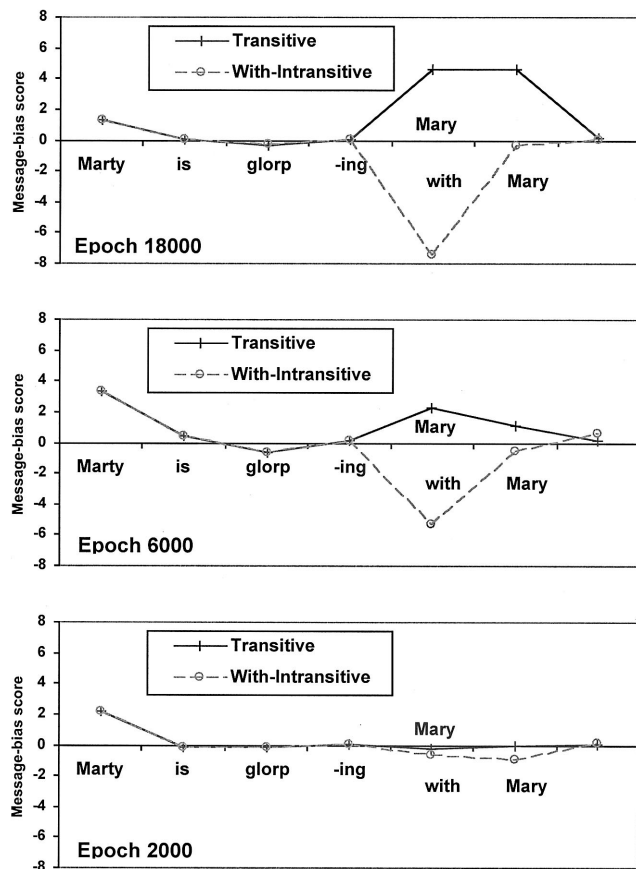


Figure 22. Message bias at each position in sentence (top, Epoch 18,000; middle, Epoch 6,000; bottom, Epoch 2,000).

2003), which suggests that children could use the frequency of causative agent subjects relative to noncausative agent subjects. Therefore, the transitive-causative preference at Epoch 6,000 is due to two factors, the *causative-agent-subject bias* and the postverbal structural bias at the position after the verb. The intransitive-noncausative preference is weaker at Epoch 6,000,

because the causative-agent-subject bias conflicts with the non-causative structure bias at the position after the verb.

To see how these biases develop, it is useful to look at the earliest point with the transitive-causative bias (Epoch 2,000). The bottom panel in Figure 22 shows that the postverbal structural bias for the match between the transitive and the causative meaning is nearly nonexistent at this point in development, and the model's preference is mainly due to the causative-agent subject bias. Therefore, the model's transitive preferential-looking behavior is initially due to causative-agent subject bias (Epoch 2,000) and only later comes to depend on the postverbal structural bias (Epoch 18,000). If the causative-agent-subject bias were stronger than the postverbal bias in some conditions with some children, then we would expect a *with* intransitive causative preference, and this is confirmed in the significant preference for the nonmatching causative video in 23-month-old boys in the Hirsh-Pasek and Golinkoff (1996) study. These boys looked at the nonmatching causative picture more (e.g., Big Bird bending Cookie Monster) when they heard "Big Bird is bending with Cookie Monster." It is significant that early in development, the model and these children both exhibit a nonmatching preference (causative-agent-subject bias with noncausative-intransitive pairs). Because early-syntax approaches assume early matching preferences, and late-syntax approaches assume that the input conspires to create matching rather than mismatching preferences, it is difficult for either of these theories to explain early mismatching preferences. Thus, the model's account may be superior in this respect.

The analysis of the error in preferential looking suggests two related reasons why transitives precede *with* intransitives in the model's simulation of preferential looking. First, causative agents are represented differently from noncausative agents in the XYZ message, and, second, constructions with causative agents are more frequent relative to those with noncausative agents. This predicts that models trained with a traditional-role message, where causative and noncausative agents are treated as the same role, would not be able to account for the weakness of intransitive preferences.

As a test of this prediction, the traditional-role models that were used in the structural priming message comparison were also tested on preferential looking. The traditional-roles message yielded a transitive preference (see Figure 23) that appeared early

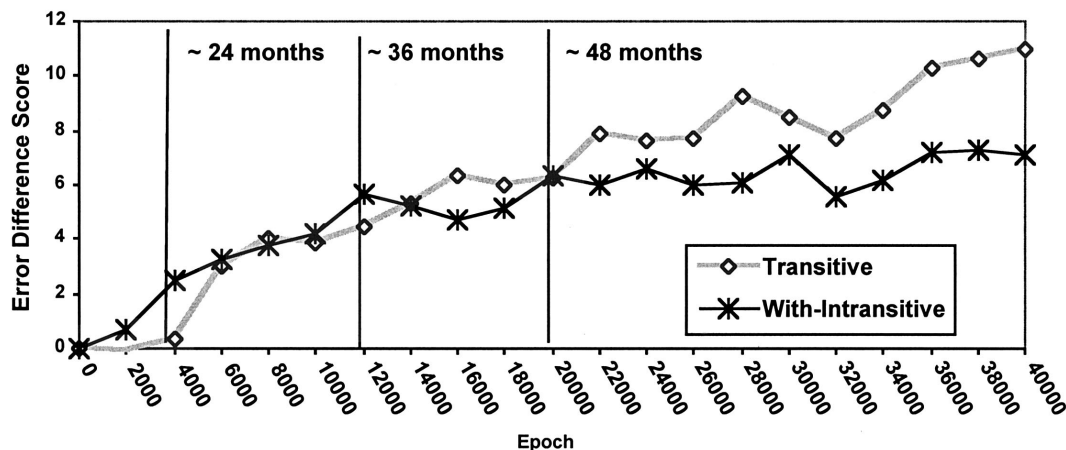


Figure 23. Preferential-looking results for transitives and *with* intransitives with traditional-roles message.

like the transitive preference with the XYZ message (see Figure 21). What differed was the *with* intransitive bias, which was stronger earlier on with the traditional-roles message than with the XYZ message and, in particular, seemed actually to precede the transitive bias in development. Thus, whereas the XYZ message had a weak *with* intransitive preference that resembled the weak preference of children, the traditional-roles message made the model's preferences for intransitives and transitives similar in strength and the intransitive preference emerged earlier. Neither of these patterns fit the existing data.

The model's account of preferential looking for transitives and *with* intransitives also leads to predictions about data yet to be gathered. Specifically, the model predicts that English-speaking children should never show a preference for pairing an intransitive frame with a noncausative meaning if they do not also show a preference for pairing a transitive frame with a causative meaning. This arises from the model's causative-agent-subject bias, which in turn arises from the model's XYZ roles and the relative frequencies of these roles occurring in certain positions and grammatical functions. If a language frequently places objects in initial position (e.g., Japanese), however, the causative-agent-subject bias should be weaker, and one should expect that early transitive-causative and intransitive-noncausative development would develop more in parallel.

### Converging Evidence

To seek converging evidence about the nature of syntax in the developing model, we also tested whether the relationship between lexical and syntactic knowledge in the model mimics the relationship in development and whether the model's structural priming behavior is analogous to what has been observed in children.

Several studies of children have found structural priming. In these studies, dative and transitive structural priming have been found at age 4 and above (Huttenlocher et al., 2004; Savage, Lieven, Theakston, & Tomasello, 2003; Whitehurst, Ironsmith, & Goldfein, 1974) and can be found at younger ages with stronger manipulations (Brooks & Tomasello, 1999; Savage et al., 2003). Because priming is a relation between different sentences, rather than a preference for a certain meaning or the ability to generate a novel sentence, it is a different measure of the abstractness of the

model's syntax. Structural priming was tested in the developing model in the same way that it was tested in the adult model. Both transitive-transitive and dative-dative priming were tested with no lag between prime sentences and target messages. Because percentages can be skewed by low numbers of correct responses, the priming difference was assessed as the difference between the raw numbers of active transitives and prepositional datives after their respective prime structures. This difference as a function of epoch is given in Figure 24.

Structural priming in the model develops gradually over time. The increase in priming as the model gets older is due to the growing abstractness of the structural representations, allowing changes to transfer between prime and target. To compare these results with the percentage priming results in the adult data, we calculated priming differences expressed in percentages for Epoch 40,000. Dative priming was 4.1%,  $F(1, 19) = 22.15$ ,  $p < .001$ , and transitive priming was 5.8%,  $F(1, 19) = 23.25$ ,  $p < .001$ , which suggests that priming is adultlike by Epoch 40,000 (see, e.g., Figure 11 for comparison). The fact that priming grows over time in the model, rather than maintaining the same level throughout, mirrors the change in the techniques needed to find robust priming in children. Priming studies at earlier ages often require stronger manipulations (lexical overlap, blocks of primes) than studies at later ages. More generally, the results suggest that the word sequence prediction mechanism, which is the heart of the model's language learning algorithm, gradually acquires representations that are sufficiently abstract and syntactic in nature to create structural priming.

One criticism of sentence production models such as the dual-path model is that they assume an abstract set of thematic roles, which are not evident early on in child language (Bowerman, 1976; Tomasello, 1992). Children's early production often involves lexically specific frames, where thematic roles appear in different orders with different verbs. For example, a child studied by Tomasello (1992) for a period of time produced agent subjects for *take*, but not for *put*. Although the dual-path model has abstract roles, it has to learn to map them onto syntax, and its ability to do this depends on its experience with particular sentences with particular verbs. This suggests that the model's syntactic knowledge may be tied to specific lexical items at the earliest stages. To

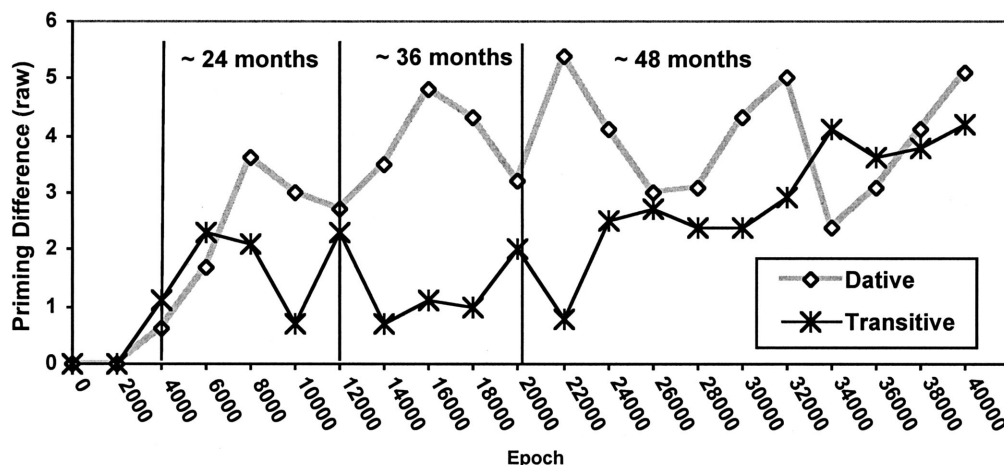


Figure 24. Structural priming in development.

examine whether early representations are verb-specific, we tested the model's developing ability to place verbs into the transitive construction. Only verbs that occurred in the transitive and benefactive constructions in the model (*kick*, *carry*, *push*, *hit*) were used. To examine the use of these verbs with transitive frames, occurrences of these verbs in the 2,000 test sentences that were used to evaluate the overall accuracy of the model were recorded. A single model was tested because averaging over models can obscure lexical variability in particular models.

The percentage of correct sentences out of all the messages that were supposed to yield a transitive sentence (NP V NP) with one of these verbs was recorded every 2,000 epochs (see Figure 25). If the model was using a common abstract transitive representation throughout development, there should be little variation in the correctness among the verbs. If, instead, abstract frames develop from lexically specific frames, we expect greater variation in earlier epochs. Figure 25 clearly shows that at the earliest epoch where there is some generalization (4,000), generalization varies from 15% (*kick*) to 63% (*carry*), a range of 48%. The range is consistently smaller for later epochs even when mean generalization is far from ceiling. The decrease in lexical variability with development in the model suggests that it starts with partially verb-specific transitive frames that later evolve into a fully abstract frame. This early lexical specificity arises from the model's SRN and localist word inputs and outputs. Because the model stores knowledge in weights from particular word units, its internal knowledge also tends to be strongly influenced by its experience with particular lexical elements (Chang, 2002; Marcus, 1998). Thus, even with abstract roles, the model can account for the lexical specificity of early sentence production.

We have shown that lexically specific syntactic patterns in child language are consistent with the model. This result demonstrates that early lexical specificity does not imply the early absence of abstract roles. The model, which has such roles from the start, nonetheless must learn to use them with particular lexical items. In this regard it is also worth noting that the model's assumption that learners know early on how to set these roles when processing situated messages has some support. Infants can extract verb/action-general information about role-correlated information such as intentionality and goal directedness (Behne, Carpenter,

Call, & Tomasello, 2005; Carpenter, Akhtar, & Tomasello, 1998; Carpenter, Call, & Tomasello, 2005). Therefore, we believe that the dual-path model provides an approach to acquisition that is consistent with much of what is known about lexical specificity and its implications for semantic and syntactic acquisition.

### Language Acquisition Conclusion

Early on, transitive development is gradual and lexically specific, because the model must make predictions from lexical inputs in order to learn (learning as processing), and its weight changes are related to the difference in its prediction and lexical outputs (prediction error). Nonetheless, the model can use its immature representations early in development in tasks that involve an implicit choice between two interpretations (e.g., preferential looking), provided that the construction being tested is associated with consistent structure-meaning mappings in the input, as is the case for the transitive in English.

As a theory of language acquisition, the dual-path model incorporates features of both late-syntax and early-syntax approaches. The sequencing system learns to sequence the XYZ roles, which leads to a preference for configurations that have stable mappings of arguments to sentence positions (Fisher, 2002b; Goldberg, 1995; Lidz, Gleitman, & Gleitman, 2003). This knowledge, however, is shaped by the types of lexical elements that occur in these configurations, with much syntactic knowledge arising from abstractions that are useful for predictions about word sequences (Bates & Goodman, 2001; Lieven et al., 1997; Tomasello, 2003). Given that the model has partial structural representations before it can produce whole transitive sentences in production, it is consistent with approaches that argue that structures are helpful for learning verb meaning (as in syntactic bootstrapping; Gleitman, 1990; Naigles, 1990). However, because syntactic bootstrapping claims that frames help to constrain the meaning of verbs, it does not explain how one learns to differentiate the meaning of verbs that occur in the same frames (e.g., "*eat*," "*drink*"). The model helps to explain how this occurs, because the model can use the noun concepts that are associated with particular verbs (e.g., "*eat*

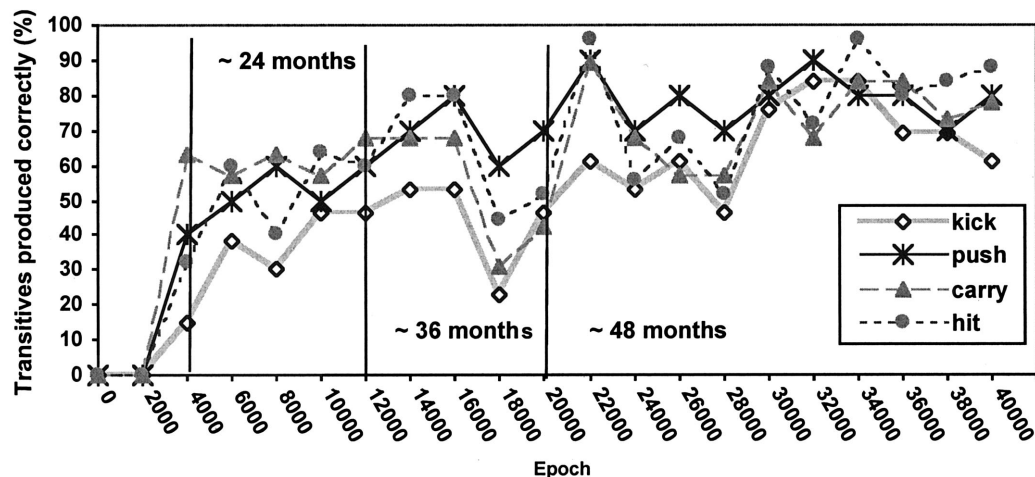


Figure 25. Verb-structure preferences during development.



a donut,” “drink some soda”) to help in distinguishing these verbs and selecting the correct verb in production.

The dual-path model differs from traditional generative approaches to language acquisition. An implicit assumption of these approaches is that grammaticality is a property of whole utterances (C. L. Baker, 1979; Chomsky, 1957; Gold, 1967; Pinker, 1984, 1989). Because a grammatical utterance will not inform the learner as to which parts of it are crucial for its grammaticality and because the amount of input is always a small subset of the indefinitely many possible utterances, generative approaches have argued that learners come preequipped to make hypotheses about experienced utterances. This is the “poverty of the stimulus” argument for innate knowledge. A variety of types of such knowledge is assumed to be necessary to learn main clause syntax: syntactic categories, X-bar templates, syntax-specific parameters (e.g., the head direction parameter), syntax-specific principles (e.g., the projection principle), syntax-specific relational operations (e.g., c-command), and syntax-semantics linking rules (e.g., agent → subject, object → noun). The dual-path model is able to learn without these types of innate knowledge, because it learns from its predictions at each word position. Incremental prediction increases the amount of information that can be used for learning and focuses learning on the particular representations that made the incorrect prediction. The prediction-based approach is thus a potential solution to the problem of learning from sparse input (Chang, Lieven, & Tomasello, 2005).

### Successes and Limitations of the Model

The dual-path model attempts to link language acquisition and adult sentence production. In Table 13, a summary of the model’s structural priming behavior is provided, showing that it is generally successful in reproducing what is known about priming. The model’s account of these data is unique because it must use its language acquisition mechanisms to account for priming in adults. At the same time, the model also directly accounts for a number of acquisition phenomena (see Table 14 for a summary). Here, assumptions that were needed for adult sentence production yielded results that resemble those found in children. Previous treatments

of the developmental phenomena did not explicitly model the relevant behavior, fueling disagreement about how to interpret the findings.

Although we believe the model provides the broadest implemented coverage of adult sentence production and child language acquisition data of any theory, theories and models are, in the end, just abstractions that tie together a complex set of results within a simpler framework. The simplifications in the dual-path model impose several limitations that are important for understanding its relationship to the totality of human language behavior.

One limitation is that the model only produces single-clause sentences and hence does not address issues of learnability, syntax, and production that relate to recursion, “moved” constituents, and the association of clauses with propositions. There are two problems related to recursion in connectionist models. One is whether it is possible to get humanlike recursion to emerge from learning in a SRN (as argued by Christiansen & Chater, 1999; Elman, 1993) or whether one needs a symbolic mechanism such as a stack (as in Miikkulainen, 1996; Miikkulainen & Mayberry, 1999). A second problem that has not been as clearly addressed is how meaning is related to recursive elements (but see a discussion of this issue in Griffin & Weinstein-Tull, 2003, and computational implementations in Miikkulainen, 1996; Miikkulainen & Dyer, 1991; Rohde, 2002). Presumably, the message must be able to encode role and propositional relationships to embedded clauses as well as to main clauses. For example, in the sentence “The boy that was chased by the dog climbed the tree,” the boy is the agent of climbing and the patient of chasing. Exploratory testing of the dual-path model with messages that encode more than one clause has been promising. What is not clear at this point, however, is whether this augmented model conforms to what we know about human production and learning of embedded clauses.

Another limitation is that the model does not truly extend to comprehension, in the sense of deriving an appropriate meaning from a sequence of words. Instead, it carries out the word prediction processes that we hypothesize accompany comprehension. Connectionist learning models that do both comprehension and

Table 13  
*Summary of Structural Priming Phenomena in Human and Model*

Phenomenon	Human	Model
Priming persists over lag?		
Lag 0 Dative-transitive	Yes-yes (Bock & Griffin, 2000)	Yes-yes
Lag 4 Dative-transitive	Yes-yes <sup>a</sup> (Bock & Griffin, 2000)	Yes-yes
Lag 10 Dative-transitive	Yes-yes (Bock & Griffin, 2000)	Yes-yes
Comprehension-based priming similar to production-based priming?	Yes (Bock et al., 2005)	Yes
Priming can be sensitive to meaning?		
Prepositional locative-dative	No (Bock & Loebell, 1990)	No
Locative-transitive	No (Bock & Loebell, 1990)	No
Locative alternation	Yes (Chang et al., 2003)	Yes
Priming can be sensitive to function morphemes?		
Preposition	No (Bock, 1989)	No
Past tense	No (Pickering & Branigan, 1998)	No
Priming can be sensitive to lexical overlap?		
Verb	Yes (Pickering & Branigan, 1998)	No

<sup>a</sup> Data suggested no transitive priming after 4 fillers, but priming returned after 10 fillers; priming from comprehension after 4 fillers showed no such decrement (Bock, 2002).

Table 14  
*Summary of Language Acquisition Phenomena in Human and Model*

Phenomenon	Human	Model
Gradual development of production?	Yes (e.g., Tomasello, 2000)	Yes
Preferential looking precedes production?	Yes (e.g., Naigles, 2003)	Yes
Transitive precedes intransitive in preferential looking?	Yes (e.g., Hirsh-Pasek & Golinkoff, 1996)	Yes
Structural priming during development?	Yes (e.g., Savage et al. 2003; Huttenlocher et al. 2004; Whitehurst et al., 1974)	Yes
Verb-based development of transitive?	Yes (e.g., Tomasello, 1992)	Yes

production have been built (Miikkulainen & Dyer, 1991; Rohde, 2002), but these models' message representations lack the features of the dual-path model that promote generalization, the need for which is certainly as acute in comprehension as it is in production.

Addressing these issues will be an important focus of future work. The present model is limited by the scope of our area of inquiry and not by inherent limitations in the ability of the model to do recursion or comprehension. Having clearer accounts of recursion and language comprehension would help to show whether and how the model can address some of the critical open questions in theories of syntactic knowledge and usage.

### Conclusion

Producing sentences from meaning and learning language from input have traditionally been studied in separate branches of psycholinguistics, but of course they are related, and a theory that recognizes the tight relationship is needed (MacDonald, 1999; Mazuka, 1998; Seidenberg & MacDonald, 2001). The model presented here represents an attempt to implement a theory of that relationship. The theory builds on three theoretical cornerstones, each of which is associated with two critical assumptions: an innate architecture, associated with sequencing by SRN and dual pathways; a structured message comprising *what–where* bindings and XYZ roles; and a domain-general learning algorithm using prediction error and implementing learning as processing. In this conclusion, we revisit these assumptions in turn.

The model's architecture incorporates an SRN to learn to sequence words. This network is needed to learn lexically specific structural preferences, as seen in the development of the transitive construction (see the *Converging Evidence* section) and more generally in adult sentence processing (e.g., Ferreira, 1996). The SRN developed representations that were appropriate for selecting words at different points in an utterance. These position-specific representations developed independently over time with language input (e.g., causative-agent-subject bias and postverbal structural bias in Figure 22). The changes in these states allowed the model to explain the overall differences between production and preferential looking, and the variable sensitivity to particular structures at different ages in preferential looking. More generally, although there is a lot of evidence that sentence production is incremental, it has not been clear whether an incremental system could explain structural priming. What this model shows is that an SRN can learn syntax in a way that allows its deployment in an incremental

fashion, while at the same time encoding aspects of the whole construction.

The dual-path architecture was designed originally to allow words to be placed into novel sentence positions (Chang, 2002). This kind of generalization resulted only when the lexical semantics and the sequencing system were located in different pathways and therefore each pathway could learn different components of the problem (akin to the division of labor in the triangle models of reading; Harm & Seidenberg, 2004; Plaut et al., 1996). Separating these pathways allowed the network to account for novel verb–structure generalization in acquisition and for double dissociations in aphasia (Gordon & Dell, 2003). The isolation of the sequencing system from lexical semantics also helps to explain why structural priming can occur without lexical or meaning overlap and why error from comprehension-input can create changes to priming in production. Finally, the dual-path architecture supports processing with novel verbs, which is necessary to explain the elicited production and preferential-looking results. Thus, an important claim of the present work is that children eventually develop abstract syntactic representations in part because of a preexisting separation in the brain between neurons that learn sequences and neurons that encode concepts (Chang, 2002; Ullman, 2001). The final result of syntactic development is an adult representation that distinguishes frames from their lexical fillers as in most theories of adult sentence production (Bock, 1982; Dell, 1986; Eberhard et al., 2005; Garrett, 1975; Gordon & Dell, 2003; Levelt, 1989).

Our use of a “*what–where*” system of bindings for semantic representations, like the dual-pathway architectural assumption, is motivated by the need for generalization. Chang (2002) specifically demonstrated that such a system of binding is required for a model to be able to produce sentences expressing meanings in which concepts occur in novel role configurations. Because structural priming can involve sentences with different meanings, it is crucial that the model have structural representations that can apply to many different meanings. Elicited production and preferential looking with novel verbs also require the ability to bind a novel word into a known syntactic structure. All of these behaviors depend on the *what–where* system of message binding.

The XYZ role representation is also an important factor in the model's behavior. Although similar representations were used in Chang (2002) and Chang et al. (2000), no experimental results were provided to support the use of those representations. Here, the XYZ role representation made the right prediction about the development of the intransitive construction in preferential-

looking studies; the traditional-role representation did not. The fact that XYZ roles are better than the traditional representation shows that children do not need a complex role representation in order to learn to map meanings into sentences. Rather they can learn language-specific and construction-specific mappings of these abstract roles (Croft, 2001), and this provides them with enough flexibility to organize their system in just the way that their language demands.

The model's learning mechanism uses prediction error to guide weight changes. This error-based learning within an incremental word-based prediction system is crucial for dealing with language acquisition. It changes the problem from an intractable whole-sentence-based hypothesis-testing problem to a simpler problem of constraining word transitions, and it thereby allows broad abstract grammars to be learnable from limited input. Error-based learning also implements structural priming, providing a common mechanism for syntactic change in children and adults. Because error can be used to implement the preferences in preferential-looking experiments, this learning algorithm motivates the existence of preferences in the first place. It is not just that children prefer to look at things that match but rather that they use their production abilities to support the anticipation of upcoming language and upcoming events, and they tend to look at the things that increase the validity of their expectations.

Perhaps the most critical assumption of this work is that language learning and adult processing are part of the same mechanism. To learn the language and to display structural priming, the model has to try to produce a string of words that can be compared with the input sentence. The process continues throughout life. This assumption is supported by many studies showing that children, including infants, learn experimentally induced linguistic regularities in the same way that adults do (Aslin, Saffran, & Newport, 1999; Chambers, Onishi, & Fisher, 2003; Gupta & Dell, 1999; Gupta & MacWhinney, 1997; Onishi, Chambers, & Fisher, 2002; J. R. Saffran, Aslin, & Newport, 1996). Connectionist models that instantiate this hypothesis have been developed in other domains (Plaut et al., 1996; Seidenberg & McClelland, 1989). Most pertinent to the present study is the fact that children and adults both show structural priming. These results forcefully argue for a unified approach to language acquisition and production, and the present model represents a step in that direction.

Modern generative linguistics emerged when language researchers started to take the abstractness of syntax seriously. The difficulty of explaining how people could learn abstract syntax from experience suggested that people must have innate syntax-specific biases in their genetic inheritance. Because it is nontrivial to imagine how abstract linguistic constraints can be implemented in a neural network (such as the brain), or why such an arrangement would have evolved in the first place, critics of generative grammar (Elman et al., 1996; MacWhinney, 1999; Seidenberg, 1997; Tomasello, 2003) and, more recently, generativists themselves (Culicover & Nowak, 2003; Hauser, Chomsky, & Fitch, 2002; Jackendoff, 2002), have placed their hopes on domain-general learning mechanisms, such as implicit sequence learning. Implicit sequence learning is evident in different tasks, different modalities, and different species (Conway & Christiansen, 2001; Gupta & Cohen, 2002; Seger, 1994), which suggests that it is domain general and computationally powerful, but also realizable in neural tissue and under selection pressures in evolution. What was missing, however, was a formal theory of how a domain-general

learning algorithm could yield the kinds of abstractions that motivated the generative enterprise in the first place. The model in this article, by showing how an error-based sequence learning mechanism within a particular architecture can yield adult syntax, provides a bridge between the biological evolution of learning mechanisms and the abstract products of our cultural evolution.

## References

- Abbot-Smith, K., Lieven, E., & Tomasello, M. (2001). What pre-school children do and do not do with ungrammatical word orders. *Cognitive Development, 16*, 1–14.
- Akhtar, N. (1999). Acquiring basic word order: Evidence for data-driven learning of syntactic structure. *Journal of Child Language, 26*, 339–356.
- Akhtar, N., & Tomasello, M. (1997). Young children's productivity with word order and verb morphology. *Developmental Psychology, 33*, 952–965.
- Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*, 247–264.
- Aslin, R. N., Saffran, J. R., & Newport, E. L. (1999). Statistical learning in linguistic and nonlinguistic domains. In B. MacWhinney (Ed.), *The emergence of language* (pp. 359–380). Mahwah, NJ: Erlbaum.
- Baker, C. L. (1979). Syntactic theory and the projection problem. *Linguistic Inquiry, 10*, 533–581.
- Baker, M. C. (2005). Mapping the terrain of language learning. *Language Learning and Development, 1*(1), 93–129.
- Bates, E., & Goodman, J. C. (2001). On the inseparability of grammar and the lexicon: Evidence from acquisition. In M. Tomasello & E. Bates (Eds.), *Language development: The essential readings. Essential readings in developmental psychology* (pp. 134–162). Oxford, England: Basil Blackwell.
- Bavin, E. L., & Growcott, C. (2000). Infants of 24–30 months understand verb frames. In M. Perkins & S. Howard (Eds.), *New directions in language development and disorders* (pp. 169–177). New York: Kluwer Academic.
- Behne, T., Carpenter, M., Call, J., & Tomasello, M. (2005). Unwilling versus unable: Infants' understanding of intentional action. *Developmental Psychology, 41*, 328–337.
- Bloom, P., Peterson, M. A., Nadel, L., & Garrett, M. F. (Eds.). (1996). *Language and space*. Cambridge, MA: MIT Press.
- Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review, 89*, 1–47.
- Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology, 18*, 355–387.
- Bock, K. (1989). Closed-class immanence in sentence production. *Cognition, 31*, 163–186.
- Bock, K. (1990, October). *Creating and remembering form in talk*. Paper presented at the Cognitive Lunch Group meeting, Department of Psychology, Duke University, Durham, NC.
- Bock, K. (1995). Sentence production: From mind to mouth. In J. L. Miller & P. D. Eimas (Eds.), *Handbook of perception and cognition: Speech, language, and communication* (Vol. 11, pp. 181–216). Orlando, FL: Academic Press.
- Bock, K. (2002, March). *Persistent structural priming from comprehension to production*. Paper presented at the 16th Annual CUNY Conference on Human Sentence Processing, New York, NY.
- Bock, K., Dell, G. S., Chang, F., & Onishi, K. H. (2005). *Persistent structural priming from language comprehension to language production*. Manuscript submitted for publication.
- Bock, K., & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General, 129*, 177–192.



- Bock, K., Irwin, D. E., Davidson, D. J., & Levelt, W. J. M. (2003). Minding the clock. *Journal of Memory and Language*, 48, 653–685.
- Bock, K., & Loebell, H. (1990). Framing sentences. *Cognition*, 35, 1–39.
- Bock, K., Loebell, H., & Morey, R. (1992). From conceptual roles to structural relations: Bridging the syntactic cleft. *Psychological Review*, 99, 150–171.
- Botvinick, M., & Plaut, D. C. (2004). Doing without schema hierarchies: A recurrent connectionist approach to normal and impaired routine sequential action. *Psychological Review*, 111, 395–429.
- Bowerman, M. (1976). Semantic factors in the acquisition of rules for word use and sentence construction. In D. Morehead & A. Morehead (Eds.), *Normal and deficient child language* (pp. 99–179). Baltimore: University Park Press.
- Boylund, J. T., & Anderson, J. A. (1998). Evidence that syntactic priming is long-lasting. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the 20th annual conference of the Cognitive Science Society* (p. 1205). Hillsdale, NJ: Erlbaum.
- Braine, M. D. (1976). Children's first word combinations. *Monographs of the Society for Research in Child Development*, 41(1), 104.
- Braine, M. D. (1992). What sort of innate structure is needed to "bootstrap" into syntax? *Cognition*, 45, 77–100.
- Branigan, H. P., Pickering, M. J., & Cleland, A. A. (1999). Syntactic priming in written production: Evidence for rapid decay. *Psychonomic Bulletin & Review*, 6, 635–640.
- Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, 75, B13–B25.
- Branigan, H. P., Pickering, M. J., Liversedge, S. P., Stewart, A. J., et al. (1995). Syntactic priming: Investigating the mental representation of language. *Journal of Psycholinguistic Research*, 24, 489–506.
- Branigan, H. P., Pickering, M. J., Stewart, A. J., & McLean, J. F. (2000). Syntactic priming in spoken production: Linguistic and temporal interference. *Memory & Cognition*, 28, 1297–1302.
- Brooks, P. J., & Tomasello, M. (1999). Young children learn to produce passives with nonce verbs. *Developmental Psychology*, 35, 29–44.
- Cameron-Faulkner, T., Lieven, E., & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science*, 27, 843–873.
- Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, 14(1), 177–208.
- Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen- through 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior and Development*, 21, 315–330.
- Carpenter, M., Call, J., & Tomasello, M. (2005). Twelve- and 18-month-olds copy actions in terms of goals. *Developmental Science*, 8, F13–F20.
- Chambers, K. E., Onishi, K. H., & Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experiences. *Cognition*, 87, B69–B77.
- Chang, F. (2002). Symbolically speaking: A connectionist model of sentence production. *Cognitive Science*, 26, 609–651.
- Chang, F., Bock, K., & Goldberg, A. E. (2003). Can thematic roles leave traces of their places? *Cognition*, 90, 29–49.
- Chang, F., Dell, G. S., Bock, K., & Griffin, Z. M. (2000). Structural priming as implicit learning: A comparison of models of sentence production. *Journal of Psycholinguistic Research*, 29, 217–229.
- Chang, F., Lieven, E., & Tomasello, M. (2005). Towards a quantitative corpus-based evaluation measure for syntactic theories. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the Cognitive Science Society* (pp. 418), Stresa, Italy.
- Chomsky, N. (1957). *Syntactic structures*. The Hague, the Netherlands: Mouton.
- Chomsky, N. (1959). A review of B. F. Skinner's Verbal Behavior. *Language*, 35, 26–58.
- Christiansen, M. H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science*, 23, 157–205.
- Christiansen, M. H., & Chater, N. (2001). Connectionist psycholinguistics: Capturing the empirical data. *Trends in Cognitive Sciences*, 5, 82–88.
- Clark, E. V., & Carpenter, K. L. (1989). The notion of source in language acquisition. *Language*, 65, 1–30.
- Clark, E. V., & Clark, H. H. (1979). When nouns surface as verbs. *Language*, 55, 767–811.
- Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General*, 120, 235–253.
- Cleland, A. A., & Pickering, M. J. (2003). The use of lexical and syntactic information in language production: Evidence from the priming of noun-phrase structure. *Journal of Memory and Language*, 49, 214–230.
- Cohen, N. J., & Eichenbaum, H. (1993). *Memory, amnesia, and the hippocampal system*. Cambridge, MA: MIT Press.
- Conway, C. M., & Christiansen, M. H. (2001). Sequential learning in non-human primates. *Trends in Cognitive Sciences*, 5, 539–546.
- Croft, W. (2001). *Radical construction grammar: Syntactic theory in typological perspective*. Oxford, England: Oxford University Press.
- Culicover, P. W., & Nowak, A. (2003). *Dynamical grammar: Minimalism, acquisition, and change*. Oxford, England: Oxford University Press.
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93, 283–321.
- Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review*, 104, 123–147.
- Dell, G. S., Chang, F., & Griffin, Z. M. (1999). Connectionist models of language production: Lexical access and grammatical encoding. *Cognitive Science*, 23, 517–542.
- Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, 104, 801–838.
- Diessel, H. (1999). *Demonstratives: Form, function and grammaticalization* (Vol. 42). Amsterdam: John Benjamins.
- Dixon, R. M. W. (1994). *Ergativity*. Cambridge, England: Cambridge University Press.
- Dowty, D. (1991). Thematic proto-rules and argument selection. *Language*, 67, 547–619.
- Eberhard, K. M., Cutting, J. C., & Bock, J. K. (2005). Making syntax of sense: Number agreement in sentence production. *Psychological Review*, 3, 531–559.
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14, 179–211.
- Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, 48, 71–99.
- Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development* (Vol. 10). Cambridge, MA: MIT Press.
- Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, 41, 469–495.
- Ferreira, V. S. (1996). Is it better to give than to donate? Syntactic flexibility in language production. *Journal of Memory and Language*, 35, 724–755.
- Ferreira, V. S. (2003). The persistence of optional complementizer production: Why saying "that" is not saying "that" at all. *Journal of Memory and Language*, 48, 379–398.
- Ferreira, V. S. (2005). *Maintaining syntactic diversity: Syntactic persistence, the inverse-preference effect, and syntactic affirmative action*. Manuscript submitted for publication.
- Ferreira, V. S., Bock, J. K., Wilson, M. P., & Cohen, N. J. (2005). *Anterograde amnesics learn the structures of sentences they cannot remember*. Manuscript submitted for publication.
- Fisher, C. (1996). Structural limits on verb mapping: The role of analogy in children's interpretations of sentences. *Cognitive Psychology*, 31, 41–81.



- Fisher, C. (2002a). The role of abstract syntactic knowledge in language acquisition: A reply to Tomasello. *Cognition*, 82, 259–278.
- Fisher, C. (2002b). Structural limits on verb mapping: The role of abstract structure in 2.5-year-olds' interpretations of novel verbs. *Developmental Science*, 5, 55–64.
- Fisher, C., Gleitman, H., & Gleitman, L. R. (1991). On the semantic content of subcategorization frames. *Cognitive Psychology*, 23, 331–392.
- Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28, 3–71.
- Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, 37, 58–93.
- Garrett, M. F. (1975). The analysis of speech production. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 133–177). London: Academic Press.
- Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition: A Journal of Developmental Linguistics*, 1(1), 3–55.
- Gleitman, L. R., Cassidy, K. W., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005). Hard words. *Language Learning and Development*, 1(1), 23–64.
- Gold, E. M. (1967). Language identification in the limit. *Information & Control*, 10, 447–474.
- Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
- Goldin-Meadow, S., & Mylander, C. (1998). Spontaneous sign systems created by deaf children in two cultures. *Nature*, 391, 279–281.
- Gordon, J. K., & Dell, G. S. (2002). Learning to divide the labor between syntax and semantics: A connectionist account of deficits in light and heavy verb production. *Brain and Cognition*, 48, 376–381.
- Gordon, J. K., & Dell, G. S. (2003). Learning to divide the labor: An account of deficits in light and heavy verb production. *Cognitive Science*, 27, 1–40.
- Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, 11, 274–279.
- Griffin, Z. M., & Weinstein-Tull, J. (2003). Conceptual structure modulates structural priming in the production of complex sentences. *Journal of Memory and Language*, 49, 537–555.
- Grimshaw, J. (1990). *Argument structure*. Cambridge, MA: MIT Press.
- Gropen, J., Pinker, S., Hollander, M., Goldberg, R., & Wilson, R. (1989). The learnability and acquisition of the dative alternation in English. *Language*, 65, 203–257.
- Gupta, P., & Cohen, N. J. (2002). Theoretical and computational analysis of skill learning, repetition priming, and procedural memory. *Psychological Review*, 109, 401–448.
- Gupta, P., & Dell, G. S. (1999). The emergence of language from serial order and procedural memory. In B. MacWhinney (Ed.), *The emergence of language* (pp. 447–481). Mahwah, NJ: Erlbaum.
- Gupta, P., & MacWhinney, B. (1997). Vocabulary acquisition and verbal short-term memory: Computational and neural bases. *Brain and Language*, 59, 267–333.
- Haegeman, L. (1994). *Introduction to government and binding theory*. Cambridge, MA: Blackwell.
- Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review*, 111, 662–720.
- Harris, A. C., & Campbell, L. (1995). *Historical syntax in cross-linguistic perspective*. Cambridge, England: Cambridge University Press.
- Hartsuiker, R. J., & Kolk, H. H. J. (1998). Syntactic persistence in Dutch. *Language and Speech*, 41(2), 143–184.
- Hartsuiker, R. J., & Kolk, H. H. J. (2001). Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology*, 42, 113–157.
- Hartsuiker, R. J., Kolk, H. H. J., & Huiskamp, P. (1999). Priming word order in sentence production. *Quarterly Journal of Experimental Psychology*, 52(A), 129–147.
- Hartsuiker, R. J., Pickering, M. J., & Velkamp, E. (2004). Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish–English bilinguals. *Psychological Science*, 15, 409–414.
- Hartsuiker, R. J., & Westenberg, C. (2000). Word order priming in written and spoken sentence production. *Cognition*, 75, B27–B39.
- Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002, November 22). The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298, 1569–1579.
- Hirsh-Pasek, K., & Golinkoff, R. M. (1996). *The origins of grammar: Evidence from early language comprehension*. Cambridge, MA: MIT Press.
- Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review*, 99, 480–517.
- Huttenlocher, J., Vasilyeva, M., & Shimpi, P. (2004). Syntactic priming in young children. *Journal of Memory and Language*, 50, 182–195.
- Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition*, 92, 101–144.
- Jackendoff, R. (1983). *Semantics and cognition*. Cambridge, MA: MIT Press.
- Jackendoff, R. (1990). *Semantic structures* (Vol. 18). Cambridge, MA: MIT Press.
- Jackendoff, R. (2002). *Foundations of language*. Oxford, England: Oxford University Press.
- James, W. (1890). *The principles of psychology*. Cambridge, MA: Harvard University Press.
- Jordan, M. (1986). *Serial order: A parallel distributed processing approach* (ICS Tech. Rep. No. 8604). La Jolla, CA: University of California, San Diego, Department of Cognitive Science.
- Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49, 133–156.
- Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research*, 32, 37–55.
- Karnath, H.-O. (2001). New insights into the functions of the superior temporal cortex. *Nature Reviews Neuroscience*, 2, 568–576.
- Keele, S. W., Ivry, R., Mayr, U., Hazeltine, E., & Heuer, H. (2003). The cognitive and neural architecture of sequence representation. *Psychological Review*, 110, 316–339.
- Kempen, G., & Harbusch, K. (2002). Performance grammar: A declarative definition. In A. Nijholt, M. Theune, & H. Hondorp (Eds.), *Computational linguistics in the Netherlands, 2001* (pp. 148–162). Amsterdam: Rodopi.
- Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science*, 11, 201–258.
- Kidd, E., Bavin, E. L., & Rhodes, B. (2001). Two-year-olds' knowledge of verbs and argument structures. In M. Almgren, A. Barreña, M.-J. Ezeizabarrena, I. Idiazabal, & B. MacWhinney (Eds.), *Research on child language acquisition: Proceedings of the 8th conference of the International Association for the Study of Child Language* (pp. 1368–1382). Somerville, MA: Cascadia Press.
- Knoefele, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition*, 95, 95–127.
- Konopka, A., & Bock, K. (2005, March). *Helping syntax out: How much do words do?* Paper presented at the 18th Annual CUNY Sentence Processing Conference, Tucson, AZ.
- Lakoff, G. (1987). *Women, fire, and dangerous things*. Chicago: University of Chicago Press.

- Lakusta, L., & Landau, B. (2005). Starting at the end: The importance of goals in spatial language. *Cognition*, 96, 1–33.
- Landau, B., & Jackendoff, R. (1993). "What" and "where" in spatial language and spatial cognition. *Behavioral and Brain Sciences*, 16, 217–265.
- Langacker, R. (1987). *Foundations of cognitive grammar*. Stanford, CA: Stanford University Press.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W. J., & Kelter, S. (1982). Surface form and memory in question answering. *Cognitive Psychology*, 14, 78–106.
- Levelt, W. J. M., Praamstra, P., Meyer, A. S., Helenius, P., & Salmelin, R. (1998). An MEG study of picture naming. *Journal of Cognitive Neuroscience*, 10, 553–567.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1–75.
- Levin, B., & Rappaport Hovav, M. (1995). *Unaccusativity: At the syntax–lexical semantics interface*. Cambridge, MA: MIT Press.
- Lidz, J., Gleitman, H., & Gleitman, L. (2003). Understanding how input matters: Verb learning and the footprint of universal grammar. *Cognition*, 87, 151–178.
- Lieven, E., Behrens, H., Speares, J., & Tomasello, M. (2003). Early syntactic creativity: A usage-based approach. *Journal of Child Language*, 30, 333–367.
- Lieven, E. V. M., Pine, J. M., & Baldwin, G. (1997). Lexically-based learning and early grammatical development. *Journal of Child Language*, 24, 187–219.
- Loebell, H., & Bock, K. (2003). Structural priming across languages. *Linguistics*, 41(5), 791–824.
- MacDonald, M. C. (1999). Distributional information in language comprehension, production, and acquisition: Three puzzles and a moral. In B. MacWhinney (Ed.), *The emergence of language* (pp. 177–196). Mahwah, NJ: Erlbaum.
- MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, 109(1), 35–54.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101(4), 676–703.
- MacKay, D. G. (1987). Asymmetries in the relationship between speech perception and production. In H. Heuer & A. F. Sanders (Eds.), *Perspectives on perception and action* (pp. 301–333). Hillsdale, NJ: Erlbaum.
- MacWhinney, B. (1987). The competition model. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 249–308). Hillsdale, NJ: Erlbaum.
- MacWhinney, B. (Ed.). (1999). *The emergence of language*. Mahwah, NJ: Erlbaum.
- Marcus, G. F. (1998). Rethinking eliminative connectionism. *Cognitive Psychology*, 37, 243–282.
- Marcus, G. F. (2001). *The algebraic mind: Integrating connectionism and cognitive science*. Cambridge, MA: MIT Press.
- Mazuka, R. (1998). *The development of language processing strategies: A cross-linguistic study between Japanese and English*. Mahwah, NJ: Erlbaum.
- McClelland, J. L., & Kawamoto, A. H. (1986). Mechanisms of sentence processing: Assigning roles to constituents in sentences. In J. L. McClelland, D. E. Rumelhart & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol 2. Psychological and biological models* (pp. 272–325). Cambridge, MA: MIT Press.
- Meijer, P. J. A., & Fox Tree, J. A. (2003). Building syntactic structures in speaking: A bilingual exploration. *Experimental Psychology*, 50(3), 184–195.
- Miikkulainen, R. (1996). Subsymbolic case-role analysis of sentences with embedded clauses. *Cognitive Science*, 20, 47–73.
- Miikkulainen, R., & Dyer, M. G. (1991). Natural language processing with modular PDP networks and distributed lexicon. *Cognitive Science*, 15, 343–399.
- Miikkulainen, R., & Mayberry, M. R., III. (1999). Disambiguation and grammar as emergent soft constraints. In B. MacWhinney (Ed.), *The emergence of language* (pp. 153–176). Mahwah, NJ: Erlbaum.
- Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action*. Oxford: Oxford University Press.
- Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, 90(1), 91–117.
- Mintz, T. H., Newport, E. L., & Bever, T. G. (2002). The distributional structure of grammatical categories in speech to young children. *Cognitive Science*, 26, 393–424.
- Mishkin, M., & Ungerleider, L. G. (1982). Contribution of striate inputs to the visuospatial functions of parieto–occipital cortex in monkeys. *Behavioral Brain Research*, 6, 57–77.
- Naigles, L. R. (1990). Children use syntax to learn verb meanings. *Journal of Child Language*, 17, 357–374.
- Naigles, L. R. (2002). Form is easy, meaning is hard: Resolving a paradox in early child language. *Cognition*, 86, 157–199.
- Naigles, L. R. (2003). Paradox lost? No, paradox found! Reply to Tomasello and Akhtar (2003). *Cognition*, 88, 325–329.
- Onishi, K. H., Chambers, K. E., & Fisher, C. (2002). Learning phonotactic constraints from brief auditory experience. *Cognition*, 83, B13–B23.
- Palmer, F. R. (1994). *Grammatical roles and relations*. Cambridge, England: Cambridge University Press.
- Petersson, K. M., Forkstam, C., & Ingvar, M. (2004). Artificial syntactic violations activate Broca's region. *Cognitive Science*, 28, 383.
- Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language*, 39, 633–651.
- Pine, J. M., & Lieven, E. V. M. (1997). Slot and frame patterns and the development of the determiner category. *Applied Psycholinguistics*, 18, 123–138.
- Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.
- Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT Press.
- Plaut, D. C., & Kello, C. T. (1999). The emergence of phonology from the interplay of speech comprehension and production: A distributed connectionist approach. In B. MacWhinney (Ed.), *The emergence of language* (pp. 381–415). Mahwah, NJ: Erlbaum.
- Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, 103, 56–115.
- Pollard, C., & Sag, I. A. (1994). *Head-driven phrase structure grammar*. Chicago: University of Chicago Press.
- Potter, M. C., & Lombardi, L. (1998). Syntactic priming in immediate recall of sentences. *Journal of Memory and Language*, 38, 265–282.
- Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, 107, 460–499.
- Redington, M., Chater, N., & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science*, 22, 425–469.
- Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., et al. (2004). Structure and deterioration of semantic memory: A neuropsychological and computational investigation. *Psychological Review*, 111, 205–235.
- Rohde, D. L. T. (1999). *LENS: The light, efficient network simulator* (Report No. CMU-CS-99–164). Pittsburgh, PA: Carnegie Mellon University, Department of Computer Science.
- Rohde, D. L. T. (2002). *A connectionist model of sentence comprehension*

- and production. Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh, PA.
- Rohde, D. L. T., & Plaut, D. C. (1999). Language acquisition in the absence of explicit negative evidence: How important is starting small? *Cognition*, 72, 67–109.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by back-propagating errors. *Nature*, 323, 533–536.
- Sachs, J. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics*, 2, 437–442.
- Saffran, E. M., & Martin, N. (1997). Effects of structural priming on sentence production in aphasics. *Language and Cognitive Processes*, 12, 877–882.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996, December 13). Statistical learning by 8-month-old infants. *Science*, 274, 1926–1928.
- Safire, W. (2003, May 11). On language: Jonesing. *New York Times*, p. 24.
- Savage, C., Lieven, E. V. M., Theakston, A. L., & Tomasello, M. (2003). Testing the abstractness of children's linguistic representations: Lexical and structural priming of syntactic constructions in young children. *Developmental Science*, 6, 557–567.
- Scheepers, C. (2003). Syntactic priming of relative clause attachments: Persistence of structural configuration in sentence production. *Cognition*, 89, 179–205.
- Seger, C. A. (1994). Implicit learning. *Psychological Bulletin*, 115, 163–196.
- Seidenberg, M. S. (1997, March 14). Language acquisition and use: Learning and applying probabilistic constraints. *Science*, 275, 1599–1603.
- Seidenberg, M. S., & MacDonald, M. C. (2001). Constraint satisfaction in language acquisition and processing. In M. H. Christiansen & N. Chater (Eds.), *Connectionist psycholinguistics* (pp. 281–318). Westport, CT: Ablex.
- Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96, 523–568.
- Shastri, L., & Ajjanagadde, V. (1993). From simple associations to systematic reasoning: A connectionist representation of rules, variables and dynamic bindings using temporal synchrony. *Behavioral and Brain Sciences*, 16, 417–494.
- Spiers, E. (2003). *Zeta-jonesing: A new word*. Retrieved Dec, 10, 2003, from <http://www.gawker.com/03/04/005480.html>
- St. John, M. F., & McClelland, J. L. (1990). Learning and applying contextual constraints in sentence comprehension. *Artificial Intelligence*, 46, 217–257.
- Talmy, L. (2000). *Toward a cognitive semantics: Vol. 1. Concept structuring systems*. Cambridge, MA: MIT Press.
- 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, 127–152.
- Tomasello, M. (1992). *First verbs: A case study of early grammatical development*. Cambridge, England: Cambridge University Press.
- Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition*, 74, 209–253.
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Tomasello, M., & Abbot-Smith, K. (2002). A tale of two theories: Response to Fisher. *Cognition*, 83, 207–214.
- Tomasello, M., & Akhtar, N. (2003). What paradox? A response to Naigles (2002). *Cognition*, 88, 317–323.
- Ullman, M. T. (2001). A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews Neuroscience*, 2, 717–726.
- Van Valin, R. D. J., & LaPolla, R. J. (1997). *Syntax: Structure, meaning and function*. Cambridge, England: Cambridge University Press.
- Vigliocco, G., & Hartsuiker, R. J. (2002). The interplay of meaning, sound, and syntax in sentence production. *Psychological Bulletin*, 128, 442–472.
- Vosse, T., & Kempen, G. (2000). Syntactic structure assembly in human parsing: A computational model based on competitive inhibition and lexicalist grammar. *Cognition*, 75, 105–143.
- Whitehurst, G. J., Ironsmith, M., & Goldfein, M. (1974). Selective imitation of the passive construction through modeling. *Journal of Experimental Child Psychology*, 17, 288–302.
- Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2003). Expecting gender: An event related brain potential study on the role of grammatical gender in comprehending a line drawing within a written sentence in Spanish. *Cortex*, 39, 483–508.
- Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*, 16, 1272–1288.
- Wilson, D. R., & Martinez, T. R. (2003). The general inefficiency of batch training for gradient descent learning. *Neural Networks*, 16, 1429–1451.

(Appendix follows)



## Appendix

### Details of the Model Simulations

The simulations were implemented using Version 2.6 of the LENS connectionist software package (Rohde, 1999). The model had 145 *word-cword* units, 20 *compress-ccompress* units, 40 *hidden-context* units, 8 *event-semantics* units (AA, DD, XX, YY, ZZ, COMMAND, PROG, PAST), 124 *what-cwhat* units, and 5 *where-cwhere* units (A, D, X, Y, Z). Unless specified otherwise, these units used the sigmoidal logistic activation function, with activation values running between 0 and 1. The learning rate started at 0.25 for 10,000 epochs and then was lowered gradually to 0.05 after an additional 30,000 epochs, where it stayed for the remaining 20,000 epochs. Weights were initially set to values uniformly sampled between -1 and 1. Steepest descent back-propagation was used for training the model during acquisition and for testing structural priming. The order of training sentences was permuted, so that all sentences were seen before repeating, but the order was random. Unless otherwise stated, default parameters of the simulator were used.

Normally, units in a connectionist model have a link from a bias unit, which helps to set the resting activation of the unit. Bias weights, however, complicated the interpretation of structural priming over lag, because some of the priming effect could be due to changes in the bias weight rather than the weight between representations. Thus, bias weights were left out, except for three layers: *event-semantics*, *what*, and *cwhat*. These layers needed bias terms to ensure that they had a low resting level of activation. The bias weight to the *event-semantics* units implemented the event semantics information (role prominence). This was done to make the setting of the event semantics information identical to the setting of the weights in the *what-where* links. The *event-semantics* units were linear units, and the connection from bias was either zero or one (and was set by the message, so it was not changed by learning). The *what* and *cwhat* units had a negative bias weight (weight = -3) to ensure that they had a low resting activation level when no input was present.

The *cwhat* units received a training signal from the previous activation of the *what* unit. Because the *cwhat* units were logistic units, the cross-entropy error measure was used. This helped to train the *cword-cwhat* mapping, because the *cword* activations were often related to the elements in the message (*where-what* links) during situated learning, and therefore the *what* unit activations were often helpful in mapping to appropriate concepts.

The *word* and *cwhere* units used a soft-max activation function, which passes the activation of the unit through an exponential function and then divides the activation of each unit by the sum of these exponential activations for the layer. In the *word* units, this leads to a single word's being activated at each time point. Error on the word units was measured in terms of divergence— $\sum t_i \log(t_i/o_i)$ —where  $o_i$  is the activation for the  $i$  output unit on the current word and  $t_i$  is its target activation because of the soft-max activation function. In the *cwhere* units, it is important to use this activation function because the *cword* layer, which feeds into the *cwhere* layer, often has multiple words activated (especially if the previous word out mismatches the external input word). Hence, it is the *cwhere* layer that decides which word was more active and which structure should be chosen.

The *context*, *cwherecopy*, and *cword* units were “elman” units, meaning that they took their activation from the previous activation of other units. The *context* units received a copy of the previous *hidden* unit activation (initially set to 0.5 to make it easier for the system to recognize the beginning of utterances). The *cwherecopy* units also averaged a copy of the *cwhere* state with the previous *cwherecopy* state, and this created a running sequence of the roles that had been processed by the model, helping the SRN to know which alternation was being produced. The *cword* units summed together the previous activation of the *word* units and the activation from external comprehended word input. The activation of this layer was normalized to make the activation pattern of particular units similar when they were activated by the *word* layer or by the external input.

Training began by randomizing the weights of the network (the seed was set to 100, and the same random weights were used for all models). In Chang (2002), batch weight update was used, that is, weight changes were collected during the processing of the entire training set (over one “epoch”), and then the weight changes were applied at once. In the present model, weights are updated after each message-sentence pattern, because this led to faster learning with fewer presentations (Wilson & Martinez, 2003). Because an epoch can be defined as the period of time between weight updates, the word “epoch” was used to refer to points in time during training. Unlike models where an epoch refers to a whole pass through the training set, here an epoch refers to the amount of time needed to train a single pattern (epoch 2,000 = after 2,000 patterns have been trained). The weights were saved every 2,000 epochs, and the model was tested on both the training and the test set (only 1,000 training patterns were tested). The training set allowed pronouns to replace full noun phrases, but in the testing set, no pronouns were used. This reduced the overlap between training and test sets (to provide a more stringent test of generalization), and the test set was a better sample of the possible sentences in the grammar (pronouns reduce the diversity of noun phrases). Grammaticality on the test sentences was initially higher than on training sentences, because there is an asymmetry in accuracy of noun phrases. If a pronoun was marked in the message, the model could still produce a full noun phrase, and early on in training, it was likely to make a mistake in grammaticality. If a full noun phrase was marked in the message, and the model decided to produce a pronoun, it was more likely to be grammatical. Thus, early in training, the test set had more switches to pronouns that were grammatical than the training set had switches to full noun phrases that were grammatical. Training ended when 60,000 patterns had been experienced, and the model was tested on structural priming with these weights. The raw output of the model was processed by a decoder program that yielded a sequence of words (both the target sequence and the actual sequence). If the most activated word unit was less than 0.3, the decoder left that position empty. The word sequence was then processed by a syntactic coder program that added the syntactic tags and the message coding (described in the “Training the Dual-Path Model” section of the main text of this article). This program also collected the statistics that were used in the figures.

### Input Environment Grammar

The next sections describe the generation of the input. Although the process of generating the input environment starts with the message and eventually yields a sequence of words, it is difficult to describe the grammar in that order, because the message is the most complex representation. Instead, we start with the lexicon, then talk about syntactic constraints related to those words, and then discuss the message structure.

### Lexicon

The model had 20 animate nouns (humans, pets), 24 inanimate nouns (vehicles, containers, places, food, drink, and plants), and 10 pronouns (*I, me, you, he, she, it, him, her, they, them*). There were 59 verbs in 14 verb classes; verbs could occur in more than one class (see the actions in Table A1). The other words included 6 adjectives, 2 determiners (*a, the*), 12 prepositions (*in, on, under, over, near, around, into, onto, by, to, for, with*), 5 forms of the auxiliary *be* (*is, are, was, were, am*) and an end-of-sentence marker (*.*). One unit was reserved for the novel test word “*glorp*”. There were also 5 inflectional morphemes: a plural noun marker (*-s*), a singular verb marker (*-ss*), a past tense marker (*-ed*), a progressive aspect marker (*-ing*), and a past participle marker (*-par*). Words, inflectional morphemes, and the end-of-sentence marker were all treated as separate lexical items and are hereafter just referred to as *words*.



Table A1  
*Message Templates for Grammar That Generates Model's Input*

Relation	Arguments	Actions
Animate intransitive	Y=ANIMAL	dance, sleep, laugh, play, go, walk, run, jump, crawl, bounce
Animate <i>with</i> intransitive	Y=ANIMAL D=WITH Z=ANIMAL	dance, sleep, laugh, play, go, walk, run, jump, crawl, bounce
Inanimate intransitive	Y=THING	open, close, break, smash, fall, disappear, float, appear
Locative	Y=ANIMAL D=PREP Z=GOAL	go, walk, run, jump, crawl, bounce
Transitive	X=ANIMAL Y=THING	open, close, break, smash, make, bake, build, mold, sculpt, shape, cook, carve
Transitive-motion	X=MOVER Y=GOAL	hit, carry, push, slide
Theme-experiencer	X=MOVER Y=ANIMAL	scare, surprise, bother, hurt
Cause-motion	X=ANIMAL Y=THING Z=GOAL D=PREP	put, hit, carry, push, slide
Transfer dative	X=ANIMAL Z=RECIPIENT Y=THING D=TO	give, send, throw, feed, trade, sell, lend, pass
Benefactive dative	X=ANIMAL Y=THING Z=RECIPIENT D=FOR	make, bake, build, mold, sculpt, shape, cook, carve
Benefactive transitive	X=ANIMAL Y=THING Z=RECIPIENT D=FOR	hit, carry, push, slide, open, close, break, smash, hurt, scare, surprise, bother
State-change	X=ANIMAL Z=CONTAINER Y=THING D=WITH	fill, cover, soak, bathe, plug, ring, flood, stain
Locative Alternation	X=ANIMAL Y=THING Z=CONTAINER D=PREP	spray, load, brush, heap, jam, rub, shower, pack

*Note.* ANIMAL refers to humans and pets; THING refers to plants, containers, foods, and drinks; CONTAINERS refers to containers; GOAL refers to animals, places, containers, and vehicles; MOVER refers to animals and vehicles; RECIPIENT refers to animals and places.

### Syntax

Table 2 in the main text presents the various sentence types and an example of each type. Some of the types are associated with syntactic alternations (active–passive voice, double-object/prepositional dative, location–theme/theme–location). The sentence structures were associated with meaning constraints as described in the next paragraph.

Sentences were generated by applying English-specific rewrite rules to the messages provided by the message generator. The order of arguments in the message used by the formal grammar was used to define the word order of the noun phrases in the sentence inventory. The verb was placed in the appropriate position, and the appropriate closed class elements were added to mark the syntactic structure. For example, if the patient preceded the agent in the message, then passive words and morphology would be added (*push* → “is push -par by”). Similar additions were made for the dative alternation and the locative alternation. The elements in each argument were then ordered in a noun-phrase appropriate way. Aspectual changes were also applied here (e.g., “is push -ing”). If a feature in the message (corresponding to the COMMAND unit in the event semantics) signaled an imperative, it caused the subject to be omitted (e.g., “give it to the girl”). A definiteness feature on arguments determined whether “*the*” or “*a*” was selected (indefinite mass nouns had no articles). The arguments in a message could have a pronoun marker that would cause it to be replaced by a pronoun with the appropriate gender and case. Then agreement rules were applied.

### Message

The message generator created messages from a message grammar. The message grammar did not include all the world knowledge that would be needed to ensure that only “meaningful” sentences would be created. Instead, it was designed to be easy to describe and to replicate. The message grammar had event templates for each sentence type and had constraints on the categories that could participate in that event (see Table 15 in the main text). Roles were labeled with the capital letters (A, D, X, Y, Z). In the table, each constraint is associated with a particular role. The letters X, Y, and Z represent arguments in the event. For example, “X=ANIMAL” means that only humans and pets could be linked to the X

role. The A role was linked to information about the action of the event (lexical semantics for the verb). The D role was linked to semantic information about the preposition used. There were six conceptual categories used for content selection: ANIMAL, THING, GOAL, RECIPIENT, MOVER, and CONTAINER (see Table 15 in the main text for more information about these categories). Event templates had frequencies associated with them, so that shorter templates were more likely to be selected than longer templates. This difference reflects our belief that adjective-modified arguments are less common than pronominal or simple noun phrase arguments. In addition, intransitive templates were half as frequent as other structures like transitives, because this mirrors the input better (Cameron-Faulkner et al., 2003). Benefactive dative templates were also half as frequent because the model had two benefactive templates (benefactive and benefactive transitive).

Roles by themselves do not fully determine the mapping into syntactic structure. Rather, it is assumed that the relative prominence of different roles within a particular message determines the mapping of roles into sentence positions (Dowty, 1991; Goldberg, 1995). Two aspects of the model were designed to deal with this assumption. The first aspect had to do with the message generation system. The message generator creates messages that have a particular default order of the roles (as in Table 15 in the main text). The linear order in the message is used only by the formal message generator and is not encoded in the message representation in the model. To represent the relative frequencies of the particular syntactic alternations, the message generator permuted the default order of the roles a certain percentage of the time for the five sentence types associated with alternations. For the transfer and benefactive dative messages, half of the messages had their Y and Z argument order switched. The rewrite rules were sensitive to these changes, so that if the Y role preceded the Z role, a prepositional dative was output by the sentence generator, but if Z preceded Y, a double-object dative occurred. The fact that half of these dative messages had permuted orders was intended to capture the fact that double-object and prepositional datives are both about equally likely in the human input environment. The locative alternation also was associated with permuted Y and Z roles half of the time, leading to equal chances of location–theme and theme–location structures. The active–passive alternation (transitive and theme-experiencer) was different. The default order X

then Y led the sentence generator to create a transitive active sentence. If the order was Y then X, it would generate a passive complete with the relevant passive morphology. Because passives are relatively uncommon, the X and Y role order was changed only 20% of the time, leading to four times as many transitive actives as passives in the collection of message–sentence pairs.

The second aspect related to alternations had to do with the way that the information about prominence of arguments was given to the model. One assumption of the dual-path architecture is that the sequencing system, which learns about syntactic frames, is relatively isolated from much of lexical semantics, allowing it to learn representations that generalize to novel content. However, the sequencing system does need some knowledge about the prominence of the arguments in the message, and so it uses three *event-semantics* units XX, YY, and ZZ (corresponding to X, Y, and Z roles in the *where* units) to encode this prominence information. For example, an intransitive sentence would only have an argument in the Y role, and hence only the YY unit would be activated in the *event semantics*. For a transitive sentence, both the XX and the YY *event-semantics* units would be active, signaling to the syntax that the X and Y roles have content. In addition, the relative activation of these two units would signal whether an active or passive was to be produced. Recall that the message generator represented the relative prominence of the X and Y roles with their linear order; an X–Y order was associated with actives and a Y–X order with passives. When messages are actually represented in the model itself, this linear order is gone and is replaced by activations of the XX and YY *event-semantics* units. If YY is more active than XX, then Y is more prominent than X and the model would be trained to produce a passive output. The reverse situation typically leads to an active output. Correspondingly, the dative and locative alternations depended on the activation of the YY and ZZ units. If the ZZ unit was more active than the YY unit, a double-object or a location–theme structure was preferred; otherwise, a prepositional dative or a theme–location was preferred. In testing situations where the model was supposed to choose between two alternatives, the difference between the relevant units was reduced, making it more difficult to use the *event-semantics* information to select the appropriate structure.

In addition to the XX, YY, and ZZ *event-semantics* units, the *event semantics* also included units for tense and aspect. There were units for past tense (PAST) and progressive aspect (PROG). Simple present tense was considered to be unmarked. There were also *event-semantics* units that marked the presence of an action (AA) and direction (DD). In general, all of these *event-semantics* units provide the sequencing system with the kind

of nonlexical information that normally affects the choice of sentence structures.

The *event-semantics* units and the material that is bound to the *where* units (X, Y, Z, A, and D) constitute the model's message. However, there is one additional property of the messages that were actually given to the model. This concerns features that determine whether a noun phrase is expressed as a pronoun or not, and if not, whether it has a definite or indefinite article. For each noun phrase, the formal grammar simply randomly generates meaning units that correspond to articles (DEFINITE) and pronouns (PRON). However, because languages of the world often use the same words for pronouns and articles (Diessel, 1999; Harris & Campbell, 1995), it did not seem appropriate to treat these as two independent features. Instead, an overlapping representation was used, which made use of the *what–where* links for the main concept (corresponding to the noun or the referent of the pronoun) and a single DEFPRO feature. If a full noun phrase was intended, then the main concept *where–what* link was set to its normal level. When a pronoun was intended, then the main concept link was set to half of this normal level. By reducing the activation of the lexical concept, this representation helps to explain why languages use pronouns (e.g., English) or null ellipsis (e.g., Japanese) rather than simply repeating a recently accessed content word. The DEFPRO unit was at the normal level for pronouns, but at 66% for definite noun phrases and 33% for indefinite noun phrases. Using one feature to represent these distinctions leads to more competition between pronouns and articles within the model. By making it harder for the model to use the message to predict the form of noun phrases, this semantic representation helped the model to develop syntactic constituents that would treat different sequences of words (e.g., “I”, “the boy”, “the old apple”) as the same type of unit (e.g., a noun phrase). This feature is crucial for developing structural representations that used these syntactic constituents and for generalizing across these structures in structural priming.

The input environment grammar was only used to generate the message–sentence pairs used in training and testing the model. The abstract syntactic knowledge that was used to generate the input was not available to the model. The model had to construct its own representations from its input experience, and therefore its representations differ from the input environment grammar.

Received August 19, 2004

Revision received August 23, 2005

Accepted August 24, 2005 ■