

Oxford Handbooks Online

Computational Models of Sentence Production: A Dual-Path Approach

Franklin Chang and Hartmut Fitz

The Oxford Handbook of Language Production

Edited by Matthew Goldrick, Victor S. Ferreira, and Michele Miozzo

Print Publication Date: Apr 2014

Subject: Psychology, Personality and Social Psychology, Cognitive Neuroscience

Online Publication Date: Jul 2014

DOI: 10.1093/oxfordhb/9780199735471.013.019

Abstract and Keywords

Sentence production involves the complex interaction of meanings, words, and structures. These interactions are language-specific and to understand them, it is useful to build computational models of production that learn their internal representations. This chapter explores how a particular connectionist-learning model called the Dual-path model explains a range of sentence production behaviors, such as structural priming, heavy NP shift, and lexical-conceptual accessibility in structural choice. The model shows how learning can play an important role in explaining adult processing in different languages. This model is contrasted with other computational approaches to understand the strengths and weakness of each method.

Keywords: sentence processing, connectionism, incremental production, learning, structural priming

Talk is cheap. At least that is what we are told when action is required. But how do we do the action of talking? Is it as easy as the idiom suggests? Talking requires that we make decisions about words and structures. For example, if we find a big bag that belongs to Kris, we might say “Kris carries a big bag.” Or, we could have also said, “she always has her bag” or “that bag is owned by Kris.” To produce these sentences, we need to select language-specific words (e.g., “Kris” or “she”) and structures (e.g., active, passive), and ensure that these elements are accessed in time and are appropriate for each other (we cannot say “that bag is owned by she”). The rules for ordering these elements are also language-specific. In Japanese, we could convey similar meanings as above with the utterances “kurisu wa kaban wo motte itta,” “kaban wo kurisu ga motte itta,” or “kaban wo motte itta.” Are the same mechanisms used in English and Japanese? Or does learning a language change the way the sentence production system works? The goal of sentence production research is to explain how we make these language-specific word and structure choices, and how these separate decisions are integrated in time to create utterances.

To understand the complex interaction of factors in sentence production, researchers have developed computational models (Dell, 1986; Dell, Chang, & Griffin, 1999; Dijkstra & de Smedt, 1996). These models make explicit how different representations interact in the construction of sentences. As in other sciences, model building does not aim at replicating the complexity of the natural world, but focuses on simplifying a very complex mechanism into a system that can be understood. This simplification function of models can be seen in three broad approaches that have been taken in building models.

The first approach, which we call the *representational approach*, uses formal representations to embody the key features of the model. An influential exemplar of this group is the incremental procedural grammar developed by Kempen and Hoenkamp (1987). They developed a grammatical (p. 71) formalism that provided a way to build structures incrementally. In their account, meanings activated lexical entries called lemmas and these lemmas carried structural information. Combining the structural information from different lemmas allowed partial syntactic structures to be built. Thus, tightly linking lexical selection and structural choices provides a way to build trees incrementally and many representational theories have adopted similar structure-building mechanisms (e.g., tree-

adjoining grammar; Ferreira, 2000).

The second group of models, which we call the *empiricist approach*, argues that behavior is strongly guided by statistical relationships in the linguistic environment (Bresnan, Cueni, Nikitina, & Baayen, 2004; Chang, Lieven, & Tomasello, 2008; Gries, 2005; Jaeger & Snider, 2007; Kempen & Harbusch, 2004). These modelers collect statistical data from labeled corpora and use those data to evaluate the fit of different models. Although formal representations are used in these models, representations are selected based on their ability to fit the data and hence these models have a weaker commitment to their representations compared with the representational approach. Those that take the empiricist approach are interested in demonstrating how probabilistic and distributional information in the input can help to explain production behavior and how abstract computational principles can provide deeper insight into the nature of the human language system.

The third approach, which we call the *brain systems approach*, assumes that the properties of the sentence production system are partially caused by the properties of the neural systems that implement them. One example of this approach is by Gordon and Dell (2003), who developed a model of how semantic and syntactic information interact in word selection in sentence production. Their model had separate pathways, one for linking conceptual semantics to words and another for linking syntactic categories to words, and this separation of pathways was inspired by evidence of double dissociations in the brain areas that support these types of knowledge in patients with aphasia. The brain systems approach looks for ways that brain organization and function can constrain models of language. This is in contrast to representational and empiricist models where neuropsychology does not strongly constrain modeling.

In this chapter we will examine the behavior of a particular model that takes the brain systems approach called the Dual-path model (Chang, 2002). This model provides an explicit account of a wide range of production phenomena, and hence it is a useful baseline for comparing different modeling approaches. By examining this account in detail and comparing it with other models, we can see how these three approaches differ in the models that they create. This chapter is organized into five sections. The first section on generalization motivates some of the basic choices made for the Dual-path model's architecture. The second section on structural priming provides evidence that adult priming behavior can be explained by the same learning mechanism that is used to explain language acquisition. The third section on word order in different languages provides evidence that the model can acquire different languages and explain differences in adult processing in those languages. The fourth section focuses on the production of complex sentences in development. In the conclusion, the strengths and weaknesses of the three different modeling approaches are compared.

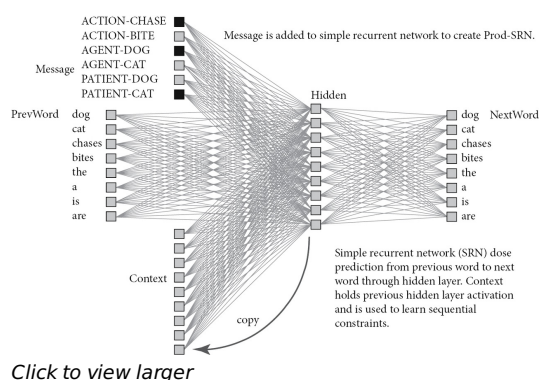
Learning to Generalize in Production

As with the Gordon and Dell (2003) model, the Dual-path model assumes that the brain's neural processing mechanisms play an important role in language production. It represents its language knowledge within an artificial neural network and this knowledge is learned through a connectionist-learning algorithm, called back-propagation of error (Rumelhart, Hinton, & Williams, 1986). Activation spreads forward along links in the network with the strength of the activation modulated by weights associated with each link. This spreading activation generates an activation pattern on the output units that represents the model's expectations. Back-propagation assumes that the difference between these expectations and the actual input, which is called error, is used to change the weights in the network (similar to theories of classical conditioning; Rescorla & Wagner, 1972). By gradually adjusting its internal representations to encode the structure of the input, the model becomes better and better at reproducing the input. Although back-propagation has some features that are not neurally plausible, it is similar to neural systems in that it updates weights between neurons using only local information in adjacent neurons and learns by making small gradual changes that can approximate the biological growth processes that support plasticity in the brain (Klintsova & Greenough, 1999). In addition, back-propagation has been very successful at modeling a wide (p. 72) range of linguistic and nonlinguistic phenomena (e.g., Botvinick & Plaut, 2004; Cleeremans & McClelland, 1991; Oppenheim, Dell, & Schwartz, 2010; Plaut, McClelland, Seidenberg, & Patterson, 1996), which suggests that the algorithm can learn representations that resemble those in human brains.

Connectionist models of sequence learning often use an architecture called a simple recurrent network (SRN) in conjunction with back-propagation of error (Christiansen & Chater, 1999; Elman, 1990). SRNs predict the words in

a sequence based on the previous word they receive as an input (Figure 4.1). In an SRN, the previous word input is linked to a hidden layer and then to the next word layer. In addition, the previous hidden layer activation is copied into a context layer (which functions as a memory buffer that enhances the ability to learn sequences) and this activation is fed back as input to the hidden layer at the next time step. The SRN learns how the previous word and the context representation could be used to predict the next word. The difference between the predicted next word and the actual next word is the error signal, and it is used to adjust the model's internal weights, so that in the future, the model is better able to predict the actual next word.

To allow this type of model to do production, Chang, Dell, Bock and Griffin (2000) augmented an SRN with a message (see also Dell et al., 1999). The message contained units that represented the combination of concepts and event roles, including a special role for the action (top part of Figure 4.1). For example, the sentence “The dog chased the cat” would have a message with three units: ACTION-CHASE, AGENT-DOG, PATIENT-CAT. When an SRN was augmented with this type of message, the resulting Prod-SRN could learn to generate sentences from meaning.



[Click to view larger](#)

Figure 4.1 Architecture of Prod-SRN model.

Analysis of the model revealed that it memorized the mapping between particular meanings and particular word sequences (Chang, 2002). This was because of the way that the model encoded the message. Because a single unit represented the binding of a role and concept (e.g., AGENT-DOG), the model could not use this unit to produce other agents or to produce the word “dog” in other event roles (e.g., PATIENT). Hence, a Prod-SRN that is trained with the message ACTION-CHASE, AGENT-DOG, PATIENT-CAT is able to produce the sentence “the dog chased the cat,” but not the sentence “the cat chased the dog,” from the message ACTION-CHASE, AGENT-CAT, PATIENT-DOG. This means that the model's syntactic knowledge did not have the property of systematicity, which is an important characteristic of human language syntax (Fodor & Pylyshyn, 1988; Hadley, 1994). The problem with the Prod-SRN was that the message did not have a separation between roles and concepts that would allow it to learn the right representations for mapping (p. 73) novel role-concept pairings. This means that the Prod-SRN was unable to generalize in a human-like manner.

The Prod-SRN did not generalize because it used a binding-by-space representation where different neurons represented the same concept in different roles (AGENT-DOG vs. PATIENT-DOG). This is the standard approach for representing slot/filler relations in connectionist models (Miikkulainen, 1996; St. John & McClelland, 1990). Because these bindings were represented with individual units, these models had trouble learning separate regularities over slots and fillers (Chang, 2002; Fodor & Pylyshyn, 1988; Marcus, 1998). Computers, however, can distinguish variable slots and fillers, because filler identity is represented with a binary code and the code has the same meaning regardless of its memory location. If the binary code 1001 represents the concept DOG, then it still means DOG regardless if it is in a memory location for agents or patients. In neural systems, however, copying the neural activation for DOG from one location in the brain to any other location does not preserve the same DOG meaning. There are ways to achieve variable-binding in neural networks (Pollack, 1990; Shastri & Aijjanagadde, 1993), but it is not clear if these mechanisms can implement the extensive variable binding that is needed in most linguistic theories. In linguistic theories, a sentence could require the dynamic binding of multiple words, heads, nouns, verbs, clause elements, phrasal elements, gaps, traces, and event roles. If the mental states that support language require extensive variable binding and neural systems do not implement such variable binding mechanisms, then it is not clear how the syntactic mind can be implemented by the neural cells in the brain (a syntactic mind-body problem).

Chang (2002) asked whether it was possible to provide a connectionist account of the combinatorial nature of syntax by giving a model a limited variable-binding mechanism. In this model, which became the Dual-path model, the message was instantiated with temporary bindings between a layer for roles and a layer for concepts (in Figure 4.2, AGENT in the role layer has a link to DOG in the concept layer).

This binding mechanism for roles and concepts was assumed to be related to the brain's spatial processing mechanisms. It is known that the brain represents object and location/action information in different neural pathways and this information needs to be bound together for spatial processing (Goodale & Milner, 1992; Mishkin & Ungerleider, 1982). The model hypothesized that a fast variable-binding mechanism evolved for spatial processing and, when language arose in humans, this mechanism was adopted for linking roles to concepts in messages. The model provided a test of whether this limited set of message variables was sufficient to explain syntactic behavior, which is normally explained in representational theories with a much larger range of linguistic variables.

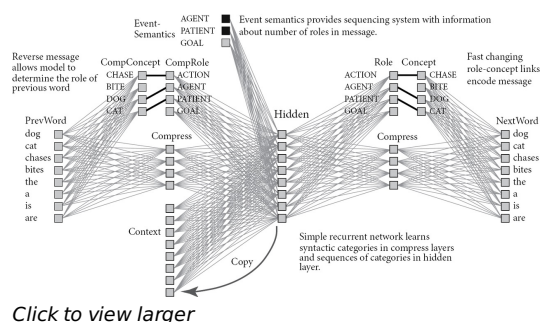


Figure 4.2 Architecture of Dual-path model.

To take advantage of this fast binding message representation, the Dual-path architecture was developed. This architecture used an SRN with Compress layers between the Hidden layer and the (p. 74) PrevWord and NextWord layers (bottom half of Figure 4.2; Elman, 1993). The Compress layers had only a small number of units and therefore the hidden layer could only influence the production of words through a small number of categories (these come to act like syntactic categories, such as nouns or verbs). To allow the model to learn the mapping between concepts and words, the Concept layer was linked to the NextWord layer. The Hidden layer in the SRN was linked to the Role layer. Because the Role layer had temporary links to the concepts in the Concept layer, this Role-Concept-NextWord pathway could learn how to activate roles to ensure that message-appropriate words were produced (top right part of Figure 4.2).

This architecture allowed the model to generalize words to novel positions and even generalize a noun as a verb. For example, “friend” is a noun and might be learned from such sentences as “The mother called a friend” paired with a message like ACTION=CALL AGENT=MOTHER PATIENT=FRIEND (= represents a temporary variable binding). The Dual-path model would learn that the concept FRIEND maps to the word “friend” and this sentence would also strengthen its ability to sequence the roles in active transitive sentences (AGENT -> ACTION -> PATIENT). This knowledge is in a format that allows the model to generalize. For example, “friend” has recently become a verb that refers to the act of adding someone to your page in a social networking site (e.g., “reluctantly, he friended his mother”). If we assume that the concept FRIEND is bound to the ACTION variable, then when the ACTION role becomes activated in a transitive sequence, the model can produce the word “friend” as a novel verb.

Putting the message into role-concept variables that are inaccessible to the SRN sequencing system has the desirable outcome that the syntactic representations in the SRN are independent of lexical content. The SRN still needs to have some knowledge about the message to know which construction to select. This information is provided to the SRN through a layer called the event-semantics (top middle of Figure 4.2), which encodes the number and type of roles that are present in the message (e.g., one role for “the dog sleeps,” two roles for “the girl chased the boy”). The activation of these units varied systematically with the structures in the input (e.g., AGENT might be less activated when a passive is to be produced) and the model could use this information to help select a structure when structural alternatives were available. Tense and aspect was also provided in the event-semantics, because this information was useful for planning the form of the verb.

With the role-concept message and the event-semantics, the model could learn different constructions with slots

for each role. For example, in the active sentence “the dog chased the cat,” the event-semantics would signal that there are two arguments (AGENT, PATIENT) and the model would learn that that event-semantics is associated with a sequence of role activations (AGENT → ACTION → PATIENT). However, two argument transitive utterances could also appear in the passive structure (“the cat is chased by the dog”). It has been found that the structural choice in this alternation is sensitive to the words that are in competition at the point where the structures diverge (the *choice point*). In the transitive alternation, the choice point is the subject noun phrase, where a speaker has to decide whether to say “dog” or “cat.” If these words are made more available, then speakers are more likely to select structures that place these words earlier (Bock, 1986a; Bock & Irwin, 1980; McDonald, Bock, & Kelly, 1993).

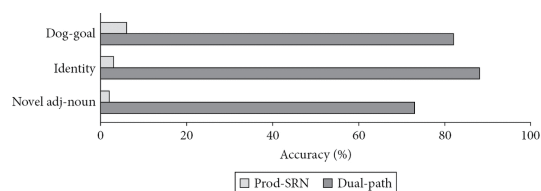
To model this lexical sensitivity, a reverse message network was created by adding a CompConcept and CompRole layer (Figure 4.2, top left). The PrevWord layer was linked to the CompConcept layer and the CompRole layer was linked to the Hidden layer in the SRN. The CompConcept and CompRole links contained a reverse copy of the message. With this reverse network, the word that was produced at the choice point (e.g., “cat”) could be mapped to its concept (e.g., CAT) and then to its role (e.g., PATIENT). This role information was then passed to the hidden layer where it combined with the event-semantics information (e.g., two arguments) and the information about the position in the sentence from the context (e.g., sentence-initial position). The model learned that activating the PATIENT role at this position with two arguments is associated with passive structures and therefore it would begin the production of this structure. Thus, even if the model was initially planning to produce an active sentence, lexical priming of “cat” could lead to the production of a passive.

An important feature of the model is the input language that the “child” model was trained on. The Dual-path model was trained on message-sentence pairs from an artificial language. The language had several types of verbs, such as intransitive, transitive, and dative verbs (“sleep,” “hit,” “give,” respectively) and several structural alternations, such as active/ (p. 75) passive transitive, and double object/prepositional dative. Although connectionist models can be trained on real corpora, it is computationally intensive to train them with a large vocabulary, because each word requires its own input/output node (a localist representation) and this leads to a large number of weights that have to be adjusted in training. When an SRN learns syntactic categories, it learns to activate a set of word units that belong to that category, so the localist representation of words is critical for these models to exhibit syntactic knowledge (a distributed output representation would not allow the SRN to activate the multiple members of each category). In addition, because corpora do not always provide information about meaning, it is difficult to use real corpora with models that require meaning input. Empiricist approaches, in contrast, often use statistical regularities from real corpora, but they reduce the computational load by only modeling particular structural decisions rather than the whole process of mapping meaning into word sequences (Bresnan et al., 2004). Thus, the complexity of sentence production can be reduced either by simplifying the production model or the language used.

To determine if the architectural assumptions of the Dual-path model allowed it to capture human syntactic behavior better than the Prod-SRN model, Chang (2002) examined the generalization ability of both models in three experiments. The first experiment asked whether the models could generalize the word “dog” to goal positions in dative sentence (e.g., “the man gave **the dog** the bone”). This was done by restricting the model’s input environment so that dog could occur in different roles, but not in the goal role. Then the models were given messages with “dog” in the goal role (ACTION=GIVE AGENT=MAN PATIENT=BONE **GOAL=DOG**) and the models were tested on whether they could correctly produce a sentence with “dog” in that position. An utterance was counted correct if it matched the target utterance word for word. The Dual-path model produced 82 percent correct *dog-goal* utterances, whereas the Prod-SRN only produced 6 percent (Figure 4.3). The Prod-SRN never trained the DOG-GOAL unit, so it was unable to generalize. But the Dual-path model separately learned to sequence the GOAL role and produce “dog” when the DOG concept was activated. Thus, when the GOAL role was linked to the DOG concept, the model was able to generalize appropriately.

Another test involved the identity construction (e.g., “a blicket was a blicket”; Marcus, 1998), where the model must generalize a novel word to two sentence positions. In training, only a subset of the possible nouns in the lexicon appeared in this construction and the remaining nouns were used for testing. The Dual-path model produced 88 percent of these novel utterances, whereas the Prod-SRN produced only 3 percent (Figure 4.3). A final generalization test used novel adjective-noun pairs. In training, one subset of adjectives was paired with animate nouns (e.g., “happy boy”) and the other subset was unrestricted by animacy (e.g., “good cake”). Messages were created that bound animate adjectives with inanimate nouns (e.g., “happy cake”). When the two models were given these messages, the Dual-path model produced 73 percent of these novel adjective-noun

phrases, whereas the Prod-SRN produced only 2 percent (Figure 4.3). For these three types of generalization, the Dual-path model performed better than the Prod-SRN and therefore it provides a closer match to human syntactic behavior.



[Click to view larger](#)

Figure 4.3 Generalization in Prod-SRN and Dual-path models. Adapted from Chang, F. (2002). Symbolically speaking: A connectionist model of sentence production. *Cognitive Science*, 26(5), 609–651.

The Dual-path model was able to generalize words to novel positions, and this ability was caused by the fact that syntactic knowledge in the SRN was independent of the lexical content in the message. This separation of lexicon and syntax was (p. 76) needed to ensure that abstract syntax was learned in the model. However, representational models often have built-in abstract language-specific syntactic structures and they often advocate a tight linking of syntax and lexicon (Reitter, Keller, & Moore, 2011). It seems clear that there should be a link between the two (Ferreira, 1996), but this link needs to be weak enough to allow nouns to be used as verbs (e.g., “I friended my mother”; Clark & Clark, 1979) or verbs to be used in novel constructions (e.g., “I sneezed the napkin across the table”; Goldberg, 1995). The learning of how verbs pair with particular structures is an important challenge for all theories of syntactic development and use (e.g., Baker’s paradox; Baker, 1979; Chang, 2002).

Structural Priming

Syntactic knowledge of language includes word order constraints across different structures. For example, both transitives and datives can appear in passive forms (e.g., “the dog was chased by the cat,” “the books were given by the man to the girl”). In addition, in English, verbs agree in number with their subjects for intransitive, transitive, and dative structures. These examples suggest that syntactic representations are shared across constructions and work in structural priming has provided experimental evidence for these links.

Structural priming is a tendency to reuse previously heard syntactic structures (Bock, 1986b; Pickering & Ferreira, 2008). For example, if participants hear a prime sentence that uses the prepositional dative structure (e.g., “the man gave the car to the church”), they are more likely to use the same structure to describe a picture (e.g., “the children showed the picture to their teacher”) than if they had heard a prime sentence in the double object dative structure (“the man gave the church the car”). Priming has been found between structures that have different roles and different function words, and this has been used to argue that priming involves an abstract structural representation that is shared across different constructions.

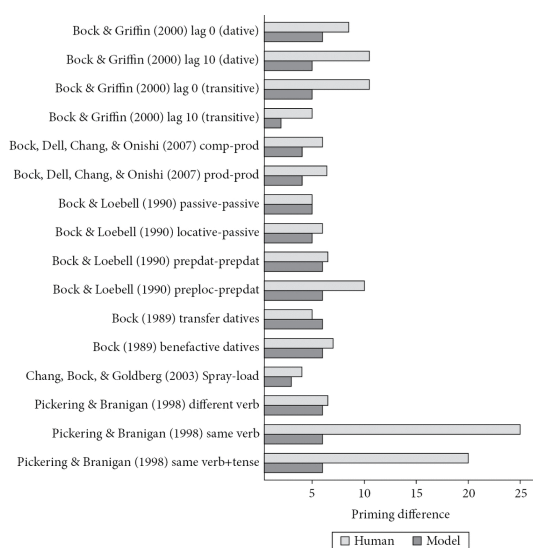
An important feature for understanding the mechanism behind structural priming is the duration of the effect. One mechanism that has been proposed to explain priming is residual activation on nodes that are used for planning structures (Pickering & Branigan, 1998). For example, hearing a prepositional dative prime could change the residual activation of an NP-PP structure node (say from an activation level of 0.2 to 0.3). If the node for the alternative double object NP-NP structure was activated at 0.25, then the speaker would be more likely to choose the prepositional dative structure after this prime ($0.3 > 0.25$). Because activation is used for structural planning at various sentence positions, it is necessary for activation to dissipate quickly to allow the system to produce other words or structures. Therefore, an activation-based account predicts that priming dissipates quickly.

An alternative to this account of priming is the idea that priming is caused by learning. In this approach, learning strengthens the representations of the structure and these changes could persist in the system. To test this, Bock and Griffin (2000) separated the prime and target by 10 intervening filler sentences and they found that priming persisted over the processing of these fillers. The magnitude of priming was the same when there were 10 intervening sentences as when there were none; this finding supported a learning-based account.

The persistence of priming suggested that long-term adaptation processes were taking place in adult speakers. One possibility is that these long-term changes in adults were caused by the same learning processes that were

used to learn the language initially (Chang, Dell, & Bock, 2006). Error-based learning in SRNs has been used to explain how adults learn and adapt to input sequences in various domains (Cleeremans & McClelland, 1991) and therefore these models might be able to explain structural priming in terms of adaptation.

Chang, Dell, and Bock (2006) examined whether error-based learning within the Dual-path architecture could model the persistence of priming. The model was first trained to map between meanings and English sentences. Once it had acquired the language, it was tested for priming by presenting the prime sentence with learning turned on. Thus, the prime sentence was treated the same way as the input sentences that the model had experienced to learn the language in the first place. For both dative and transitive prime-target pairs, the model processed a prime followed by 10 intransitive filler sentences with learning on, and then it was given a target message to produce. The model was more likely to describe the target message using the same structure as the prime and therefore it exhibited priming (see Bock & Griffin, 2000; Figure 4.4). Importantly, the magnitude of priming was the same regardless of the number of fillers, which suggests that immediate and long-term priming could be explained by a single mechanism.



[Click to view larger](#)

Figure 4.4 Structural priming in humans and Dual-path model. Adapted from Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, 113(2), 234–272.

The model assumes that error is generated during the prediction of the prime as it is comprehended. (p. 77) Thus, priming should be similar when primes are only comprehended and when they are produced. Support for this hypothesis was found in Bock, Dell, Chang, and Onishi (2007), where priming persisted when primes were only comprehended and when they were comprehended and produced (Figure 4.4). Although it is clear that production representations must be learned from comprehended input, the model does not predict that comprehension-to-comprehension priming has the same properties. This is because comprehension of meaning requires a way to map from structures to meaning, and the next word prediction mechanism in the model does not naturally explain how this mapping is trained. Experimental work has found comprehension-to-comprehension priming in eye-tracking, but it seems to have different timing properties from priming in production (Arai, van Gompel, & Scheepers, 2007; Thothathiri & Snedeker, 2008a, 2008b).

Structural priming has provided evidence that the syntactic representations that support production are abstract, in that they do not seem to depend on overlap in event-roles or in function words. For example, Bock and Loebell (1990) examined whether the description of transitive pictures (e.g., “the postman was chased by the dog,” “the dog chased the postman”) would be primed by locative primes like “the 747 was landing by the control tower,” passive primes like “the 747 was alerted by the control tower,” and active control sentences like “the 747 radioed the control tower.” If priming was caused by the order of agent and patient, then passives should prime differently from locatives, because only the passive has the same roles as the transitive. However, if priming was caused by the surface syntactic structure, then it is possible that passive and locatives would prime similarly, because both have a similar sequence of surface categories (something like DET NOUN AUX VERB PREP DET NOUN). In fact, they found that locatives and passives primed passives equally well relative to active primes. When the Dual-path

model was tested with stimuli like those in this study, there (p. 78) was no difference in the magnitude of priming from locatives and passives (see Bock & Loebell, 1990; Locative/Passive-Passive in Figure 4.4). Similar results were found for prepositional dative structures like “the wealthy widow gave the Mercedes to the church” and prepositional locative structures like “the wealthy widow drove the Mercedes to the church”) even though the “church” is a goal in the dative and a location in the locative. The model can reproduce these results (see Bock & Loebell, 1990, Preploc/Prepdat-Prepdat in Figure 4.4). Therefore as with humans, the model learned internal representations that encoded surface structural similarity between constructions with different meanings and structural priming in adults can be explained with error-based learning in an SRN.

Although locatives and passives shared similar surface orders, they also both used the word “by” and it is possible that priming was caused by the overlap in this function word. To examine whether function word overlap could influence priming, Bock (1989) tested whether the description of transfer dative pictures (e.g., “the girl gave the man the paintbrush,” “the girl gave the paintbrush to the man”) could be influenced by overlap in a function word in the prime. She compared benefactive dative primes that use the preposition “for” (e.g., “the cheerleader saved a seat for her boyfriend”) with transfer dative primes that use the preposition “to” (e.g., “the cheerleader offered a seat to her friend”). She found that the amount of priming for prepositional dative primes relative to double object primes was the same for both transfer and benefactive datives, even though the benefactive dative uses a preposition that differs from the one in the transfer dative. This result suggested that priming was not simply caused by function word overlap, a finding that is difficult to explain in lexicalized accounts of priming (Reitter et al., 2011). When the Dual-path model was tested on such stimuli, there was no difference in the magnitude of priming from to-datives and for-datives (see Bock, 1989; Figure 4.4). Therefore, the model’s syntactic representations are independent of function words and morphological overlap. Although priming is insensitive to function word overlap, it can be magnified by verb or noun overlap between prime and target (Cleland & Pickering, 2003; Pickering & Branigan, 1998). Because it also occurs without any lexical overlap, the minimal requirement for priming is structural similarity between prime and target.

In general, structural priming studies support the idea that syntactic representations are separable from the meaning representations that control them. Is this separation between meaning and structure universal or does it depend on the structure being tested? In a study that speaks to this issue, Chang, Bock, and Goldberg (2003) found that the choice of structure in the theme-location alternation (e.g., theme-location “the maid rubbed polish onto the table,” location-theme “the maid rubbed the table with polish”) could be primed by utterances that had the same order of event roles. Because the two orders in this alternation have similar sequences of syntactic categories (e.g., VERB NP PP), the fact that the order of theme and location influences priming shows that meaning is encoded in syntactic representations for this construction. The Dual-path model provides an error-based learning account of this result (see Chang et al., 2003; Spray-load in Figure 4.4). Initially, the SRN has a tendency to learn syntactic categories, because it does not have direct access to the message. Sequencing representations made of syntactic categories are sufficient to distinguish active and passive, but not the structures in the theme-location alternation because both are made up of the same sequence of syntactic categories. This generates error in learning and the model is forced to reorganize its representations to distinguish theme-location from location-theme structures by marking the order of roles in the syntax for this alternation. Thus, in contrast to representational theories that require syntactic knowledge to be of a consistent type across constructions (e.g., Ferreira, 2000), models that learn their syntactic representations can allow some constructions to be made up of abstract categories (e.g., passive), whereas others incorporate thematic roles (e.g., theme-location).

To understand how the model implements locative-passive priming, the hidden layer representations were analyzed. Connectionist models are often thought to make use of distributed representations, where knowledge is encoded by multiple units. Therefore, it is possible that the similarity between locatives and passives in structure would be distributed over multiple hidden units. In actuality, it was often the case that only one hidden unit was involved in locative-passive priming. One reason for this is that the model is trying to reduce error over all of the structures in the training corpus and hence it must organize the hidden units to best deal with the diverse structures in the input. This creates pressure to isolate each structural type to a small set of hidden units. Hence, the model suggests that abstractness of structural priming arises from (p. 79) the need to learn a large set of structures within a limited representational space.

The Dual-path model can account for a wide variety of priming effects, but does it make any predictions? One prediction arises from the model’s error-based learning mechanism. If the prime is different from the utterance that

the model expects, then error should be large and this should predict greater priming. Bernolet and Hartsuiker (2010) examined this prediction by looking at whether structural priming was influenced by verb bias. Verb bias is a tendency for particular verbs to be paired with particular structures, a bias that can influence sentence processing (Garnsey, Pearlmutter, Myers, & Lotocky, 1997; Wilson & Garnsey, 2009). Bernolet and Hartsuiker found that priming was greater when the verb's structural bias and the prime structure mismatched, supporting the idea that priming is caused by error-based learning. Jaeger and Snider (2007) offer an empiricist account of structural priming that can explain these effects. They argue that priming reflects the principles of surprisal and cumulativeness. Surprisal measures how unexpected words or structures are given previous information. Cumulativeness represents the idea that the effect of each prime structure accumulates within the language system. This model makes similar predictions to the Dual-path model, because surprisal can be implemented with error-driven learning and cumulativeness can be implemented with gradual accumulation of knowledge in the weights in a neural network.

Another prediction of the Dual-path model arises from the assumption that learning is instantiated by physical changes in the connectivity of neurons and hence learning must take place slowly. This assumption is implemented by setting a parameter that controls the rate of learning in the model to a small value. A small learning rate is also behaviorally important for the encoding of frequency in the model. Each training episode makes a small change to the model's weights, so frequent episodes are better represented in the model's weights than infrequent ones. For example, if the verb "give" appears in a double object structure (e.g., "give him a book") more often than the prepositional dative structure (e.g., "give a book to him"), then the weights between "give" and animate noun phrases like "him" have been changed by a large number of training events and this can create verb-structure regularities (see Chang, 2002 for evidence that the Dual-path model learns verb-structure associations in acquisition).

A small learning rate means that the Dual-path model cannot explain large magnitude priming effects. One example of a large effect is the lexical boost, where priming is increased when prime and target share a verb or a noun (Pickering & Branigan, 1998). Sometimes this boost can create priming effects that are huge (73 percent priming; Hartsuiker, Bernolet, Schoonbaert, Speybroeck, & Vanderelst, 2008). If the lexical boost was caused by learning, then these large changes caused by priming would be making large changes to the speaker's language system and it would even be possible for structures to be primed out of existence (catastrophic interference; McCloskey & Cohen, 1989). These theoretical issues and the fact that the model did not exhibit a lexical boost (see Pickering & Branigan, 1998; Same Verb line in Figure 4.4) led Chang et al. (2006) to suggest that the lexical boost was caused by a different short-lived mechanism and the prediction of this Dual-mechanism account has been confirmed experimentally. Hartsuiker et al. (2008) found that the lexical boost dissipates quickly, whereas structure-based priming persists longer. Further evidence comes from acquisition, where Rowland, Chang, Ambridge, Pine, and Lieven (2011) found that abstract structural priming had a similar magnitude in 3- to 4-year-old children, 5- to 6-year-old children, and adults, but the lexical boost grew over development. Bringing these results together, Chang, Janciauskas, and Fitz (2012) argued that the slow and fast learning in complementary memory systems accounts of cortical and hippocampal learning provides a unified way of explaining both the large magnitude of the lexical boost and its variability over delay and development.

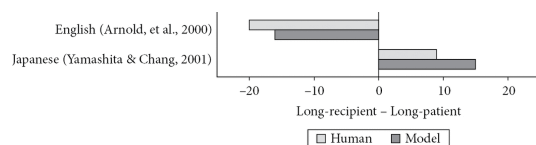
Word Order in Different Languages

The Dual-path model's account of syntactic generalization and priming suggests that it has the right properties for explaining English syntactic behavior. However, a language production system needs to be able to explain behavior in different languages, particularly those that are typologically different from English (Evans & Levinson, 2009; Jaeger & Norcliffe, 2009). Japanese differs from English in many ways. Japanese verbs occur at the ends of clauses, whereas English verbs tend to occur early in sentences. Japanese speakers can omit all arguments of a verb, where English speakers are required to use pronouns. For example, the English sentence "I gave it to them" could be expressed by the Japanese verb "ageta," because the arguments of the verb can often be inferred from the context. Also, whereas (p. 80) English has structural alternations like active and passive, Japanese speakers typically convey similar changes with scrambling of case-marked arguments ("I gave the book to the man" can be conveyed by the canonical "otoko-ni hon-o ageta" or scrambled as "hon-o otoko-ni ageta"). Because the Dual-path model is an account of how syntax is acquired, it is necessary to test it on typologically different languages to see if it can acquire different languages to the same degree (Chang et al., 2008).

To test the model, English and Japanese versions of the Dual-path model were created (Chang, 2009). Both models were given the same messages, but the sentences paired with the messages differed in verb position, argument omission, and scrambling in accordance with the language being learned. When tested on a novel set of message-sentence pairs, both models were able to show similarly high levels of accuracy (grammatical output: English model 93%, Japanese model 95%). Grammaticality was measured by labeling the model's output with syntactic categories and seeing if the produced sequence was in the set of sequences derived by labeling the model's input with syntactic categories. Although the message-sentence mapping in these two languages was quite different, the model was able to learn both languages equally fast.

A cross-linguistic model of production should also be able to explain differences in production biases between these languages. One such cross-linguistic difference comes from a phenomenon called heavy NP shift (Arnold, Losongco, Wasow, & Ginstrom, 2000; Hawkins, 1994, 2004; Ross, 1967). English speakers have a tendency to prefer configurations where long phrases are placed later in sentences. For example, speakers might change the sentence "the man gave the woman that he met last week the book" into "the man gave the book to the woman that he met last week," where the long phrase "the woman that he met last week" is at the end of the sentence. Although English speakers have a short-before-long bias, Japanese speakers have a long-before-short bias (Hawkins, 1994; Yamashita & Chang, 2001). A theory of sentence production needs to explain these cross-linguistic differences.

To examine the Dual-path model's heavy NP shift behavior, it was given dative messages where the patient or recipient phrase was made longer by modifying it with a relative clause (Chang, 2009). The model then produced the utterances and the order of short and long phrases was examined. The English version produced more recipient-before-patient orders when the patient was long than when the recipient was long (e.g., long patient "a man gave a girl a telephone that a dog touched" versus long recipient "a man gave a girl that a dog touched a telephone"). The Japanese model produced more recipient-before-patient orders when the recipient was long than when the patient was long. The model data are shown in Figure 4.5 (the Japanese human data come from the dative items in Yamashita & Chang, 2001, and the English human data were created by averaging the values in Figure 8 of Arnold, et al. 2000). The results show that the model can exhibit the English short-before-long and the Japanese long-before-short pattern.



Click to view larger

Figure 4.5 Heavy NP shift in English and Japanese in humans and Dual-path model (difference in use of recipient-before-patient order when recipient is long vs. when patient is long).

Analysis of the model's internal representations suggested that heavy NP shift behavior was caused by a difference in the relative importance of meaning and surface structural information in the two languages at the choice point where the two word orders diverge. In English, the choice point was after the verb and at that point, the model tended to use structural cues, such as the fact that short-before-long utterances are more similar to simple main clause structures than long-before-short utterances. For example, the simple utterance "a man gave a girl a telephone" is more similar to the long patient utterance "a man gave a girl a telephone that a dog touched" than it is to the corresponding long recipient utterance "a man gave a girl that a dog touched a telephone" and therefore the high frequency of simple main clauses in the input could bias the model toward the short-before-long utterance. (p. 81)

In Japanese, however, the choice point is at the beginning of the sentence, and at this position, it is difficult to use structural cues, because in Japanese the verb occurs late in the sentence and early structural configurations are highly variable because of argument omission and scrambling. Therefore, the Japanese model preferred to use meaning information at this position. Because the message signals that a relative clause should be produced and the model has learned that relative clauses go before their heads in Japanese, the model often prefers to start with a relative clause and that creates a long-before-short order.

Empiricist models have also addressed heavy NP shift by including phrase weight as a factor in predicting English word order (Arnold et al., 2000; Bresnan et al., 2004), but these models have not been extended to Japanese. One

question with these models is the stability of the parameters across different corpora or languages. If a similar model was built from Japanese data, would it have similar parameters to the English models? It is crucial to determine which aspects of these models are universal and which are language- or corpus-dependent.

Another important phenomenon that differs between English and Japanese is lexical/conceptual accessibility. Accessible words are words that are easier to produce in naming studies. Many studies have found that English speakers tend to place accessible words early in sentences and this sometimes requires later changes in the structure to maintain the same meaning. For example, McDonald, Bock, and Kelly (1993) found that participants preferred to put animate elements early in transitive sentences and this sometimes required them to use a generally dispreferred structure like a passive (e.g., “the students were frightened by the sound”). However, when the same manipulation was done with conjunctions (e.g., “the manager and the key,” “the key and the manager”), they found that animacy did not influence the word order. This difference seemed to align with the distinction between functional and positional levels in production theories (Garrett, 1988). Active and passive structures differ in the element that is assigned to the subject function and this assignment takes place at the functional level in the theory. The elements in conjunctions are assigned to the same syntactic function and the ordering of these two noun phrases takes place at the positional level. Therefore, the behavioral difference between transitives and conjunctions suggested that conceptual factors like animacy can have an influence on the functional but not on the positional level (similar results for other factors have been found, such as imageability; Bock & Warren, 1985).

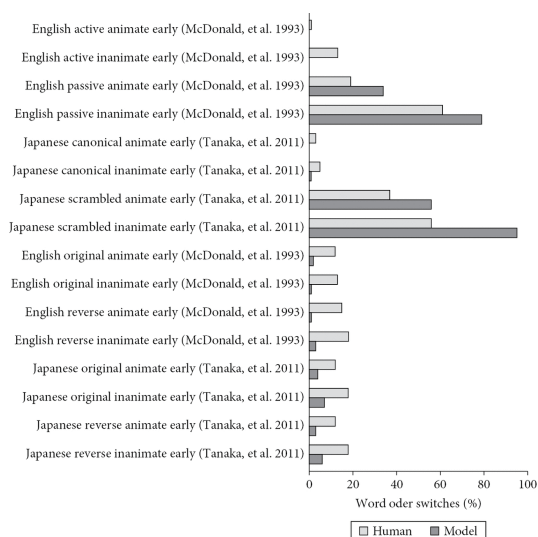
Although the functional/positional distinction has been useful for explaining results in English, it is a bit problematic for explaining behavior in Japanese. Syntactic functions in Japanese are signaled by case markers and the same case markers are used regardless of whether canonical or scrambled order is produced (e.g., “John eats rice” could be said in canonical order as “John-ga gohan-o taberu” or scrambled order as “gohan-o John-ga taberu”). Scrambling does not change syntactic functions and therefore it should be a positional-level effect. This would predict that conceptual factors do not influence scrambling in Japanese, but in fact it has been found that animacy and discourse status can influence Japanese scrambling (Ferreira & Yoshita, 2003; Tanaka, Branigan, McLean, & Pickering, 2011). To further complicate things, Tanaka et al. (2011) found that animacy does not influence the order of elements in Japanese conjunctions, which is similar to the behavior in English. Thus, production behavior for transitives and conjunctions differs in similar ways in English and Japanese, but in Japanese this distinction is difficult to explain in terms of functional and positional processing

Chang (2009) provided an alternative to the functional/positional account using the Dual-path model. One component of this account is to explain how lexical accessibility can influence structural choice or scrambling. The model has a reverse message system that maps from the previously produced word to its concept and then to its role in the particular message that is being expressed. For example, if a person started a sentence with “dog” and the dog was the patient in a transitive event, then the model could use the information in the PrevWord-CompConcept-CompRole system to produce a passive structure. Given this feature of the architecture, any factor that makes a word more likely to be produced early can influence the model’s structural choices and this influence will be felt regardless of whether the language uses syntactic function assignment or scrambling to make these words prominent.

Because the Dual-path model learns its internal representations from input utterances, the prominence of words must be learned from the mapping of words to sentence positions. By giving the model input where animate words tended to occur earlier in sentences than inanimate elements, the model learned stronger weights from animate concepts to animate words and this made them more (p. 82) prominent. The learned prominence of words and the reverse word-role system can work together to create the accessibility-sensitive nature of structure selection in English and Japanese transitives.

If animacy can influence word order, why does it not influence the order of words in conjunctions? The data in McDonald et al. (1993) provide a clue as to why conjunctions were different from transitives. They gave participants utterances to recall and then examined how often they recalled the original order. What is interesting is that the participants rarely switched the order of words in conjunctions. To explain this, Chang (2009) argued that the model could use the activation of the units in the event-semantics to guide the word order that was produced. However, because the influence of the event-semantics was learned, effective use of this information required that both orders were trained equally often and this seems to be the case for conjunctions. Although the event-semantics also signaled word order for transitives, the low frequency of structures like English passives or

scrambled Japanese utterances made it hard for the model to use this information and therefore it was more likely to switch the structure in response to other factors, such as animacy. Thus in this account, learning plays an important role in explaining the difference between transitives and conjunctions in their sensitivity to animacy.



[Click to view larger](#)

Figure 4.6 Accessibility in English and Japanese in humans and Dual-path model. Adapted from Chang, F. (2009). Learning to order words: A connectionist model of heavy NP shift and accessibility effects in Japanese and English. *Journal of Memory and Language*, 61(3), 374–397.

When given input with appropriate prominence and frequency information, the Dual-path model was able to reproduce the overall pattern of the human behavioral data (Figure 4.6). In the English model, passives were likely to switch to actives and these switches were sensitive to animacy. Meanwhile, the Japanese model was also likely to switch scrambled transitives back to the unscrambled version and the switch was also sensitive to animacy. Animacy had little effect on conjunctions, because in both languages the model rarely switched the order of nouns. Therefore, the Dual-path model's learning algorithm was (p. 83) dependent on distributional regularities and meaning in a way that simulated the difference between transitives and conjunctions. Without an architectural functional/positional distinction, the model nonetheless could explain the behavioral data that support this distinction.

The brain systems approach, as implemented with the Dual-path model, allows us to separate out aspects of production that are universal (e.g., architecture, learning algorithm) from the parts that are learned. The Dual-path model was able to learn English-like and Japanese-like languages equally well and it provides an explicit account of word ordering differences in these two languages. It is possible to build representational models for typologically different languages, but this sometimes requires changes to the formalism (e.g., scrambling in tree-adjoining grammar; Rambow & Lee, 1994). Empiricist approaches can also be applied to different languages, as long as similar tagged corpora exist in each language. More work is needed to disentangle the parameters in the model that are just caused by fitting a corpus and those that reflect the underlying nature of production.

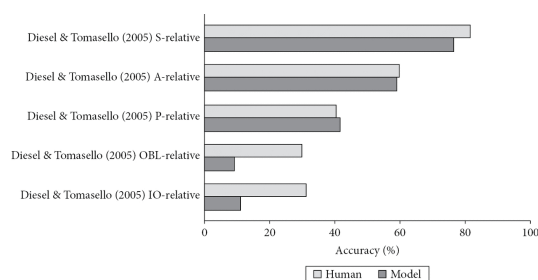
Production of Complex Syntax

In the previous section we saw that the Dual-path model could produce relative clauses. Relative clauses are theoretically important because they are a syntactic device that makes language structurally productive and they create long-distance dependencies between words. In the sentence “the dog that the cat chased loves the girl,” the main clause verb “loves” has to agree in number with the subject “dog” and this dependency spans the relative clause (Bock & Cutting, 1992). Relative clause constructions, such as the one above, have also been characterized as an instance of recursion where a transitive structure is embedded within itself and it has been argued that recursion is universal, innate, and uniquely human (Hauser, Chomsky, & Fitch, 2002, but see Evans & Levinson, 2009, for an alternative view). Several studies have explored the learnability of recursive languages with SRN-type models (Cartling, 2008; Christiansen & Chater, 1999; Elman, 1993). An important aspect of recursion that has not been examined in great detail, however, is the role of meaning (or the message) in the generation of these

utterances. For example, in the above sentence, “dog” is the subject and agent of the main clause verb “love,” but also the fronted object and patient of the embedded clause verb “chased.” If meaning helps to support recursion, then we can examine its influence within the Dual-path production model.

English relative clauses can be distinguished based on verb type and the grammatical function of the head noun in the relative clause. In S-relatives, for example, the relative clause has an intransitive verb and the head noun is the subject (e.g., “the boy that _ runs”; the underscore indicates the canonical position of the head noun “boy” in the relative clause). When the verb is transitive, the head noun can be the relative clause subject, as in “the boy that _ chased the dog” (A-relative), or the direct object, as in “the cat that the dog chased _” (P-relative). With dative verbs, three constituents can be relativized (e.g., the indirect object as in “the girl who the boy gave the apple to _”; IO-relative). The final type considered here is the case where the relative clause verb is intransitive and the head noun is an oblique argument as in “the boy who the girl played with _” (OBL-relative).

Diessel and Tomasello (2005) conducted an elicited production study with English and German children age 4;3 to 4;9, where subjects had to repeat sentences from an experimenter. They found that children were able to reproduce S-relatives the best, followed by A-relatives, then P-relatives, and finally OBL-relatives and IO-relatives (Figure 4.7). This order of acquisition also resembled results from adult production (Keenan & Hawkins, 1987), suggesting that the sources of difficulty might be similar for adults and children. A version of the Dual-path model was developed that could generate utterances with relative clauses (Fitz, 2009). The model had a message with multiple propositions and there were special units in the event-semantics that signaled the coreference of roles (similar to the message in Chang, 2009). During learning, the model was periodically tested on the structures used by Diessel and Tomasello (2005). In these structures, the five relative clause types defined previously were attached to a presentational main clause (e.g., “There is the boy that the dog was chasing”). The model eventually learned to produce all structures with more than 90 percent accuracy at the end of training.



[Click to view larger](#)

Figure 4.7 Elicited production of relative clauses in children and the Dual-path model. Adapted from Fitz, H., & Chang, F. (2008). The role of the input in a connectionist account of the accessibility hierarchy in development. In H. Chan, H. Jacob, & E. Kapia (Eds.), *Proceedings of the Boston University conference on language development* (Vol. 32, pp. 120–131). Somerville, MA: Cascadia Press.

It was found that early in development the model’s behavior approximated the child data (Figure 4.7). S-relatives were easier than A- or P-relatives, because S-relatives are unambiguous about which element is relativized — there is only one participant in intransitive events. Transitive utterances, however, can relativize either the subordinate clause subject or object and this ambiguity increases the difficulty in producing the A- or P-relatives. Another (p. 84) advantage of S-relatives over A- and P-relatives was that transitive relative clauses admitted passive alternations. Active and passive relative clauses had very similar messages in the event-semantics and the model had to learn both these structures over a single set of connection weights. This created competition between the syntactic alternations and complicated the meaning to form mapping for A- and P-relatives, but not for S-relatives. Transitive A-relatives were produced more accurately than the P-relatives, because A-relatives contained a sequence of surface syntactic categories that was shared with S-relatives (i.e., THAT VERB) and also occurred in subject-relativized obliques and datives (e.g., “that gave the toy to the dog”). For P-relatives this sequence was different (THAT ARTICLE NOUN VERB) and it was less frequent overall. Thus, although the Dual-path model was exposed to equal amounts of A- and P-relatives during learning, A-relatives had an advantage because of substructure overlap with other subject-relativized constructions in the input. The difference between P-relatives and OBL/IO-relatives was more complicated, but it was also caused in part by mapping and surface structure similarity. In summary, the Dual-path model was able to explain the performance ordering found in Diessel and Tomasello (2005) for English-speaking children, because the acquisition of each structure in the model depended

on surface similarity, frequency, and the complexity of the mapping from meaning into utterances.

Representational and empiricist models have also been applied to explain differential processing of relative clause types. Hale (2006) was able to account for the Keenan and Hawkins (1987) adult data by using a word-prediction uncertainty mechanism based on a probabilistic grammar. This mechanism is similar to word-based error in SRNs, except that connectionist approaches suggest that uncertainty as error has value as a learning signal. The main difference between the Dual-path model and Hale's models is the fact that the Dual-path model learns syntactic representations that can be used in production. Hale's model depends on a particular set of English syntactic representations and those representations have not been integrated with theories of sentence production. Empiricist approaches have also correlated processing difficulty in adults and children with frequency of occurrence, especially for the subject/object relative clause asymmetry (Kidd, Brandt, Lieven, & Tomasello, 2007; Reali & Christiansen, 2007). Future work is needed to distinguish these different approaches to complex syntax.

Conclusion

This chapter focuses on the behavior of the Dual-path model, because it provides a unified account of several important phenomena in production: generalization, structural priming, heavy NP shift, accessibility, syntax acquisition, and recursion. Error-based learning explains both the model's ability to acquire internal syntactic representations and the persistence of structural priming. The Dual-path architecture helps to explain why structural priming can be abstract and why words (p. 85) can generalize to novel positions. When learning is combined with this incremental production architecture, the result is a model that can account for the behavioral differences in the production of word order in English and Japanese. The combination of meaning, incremental production, and learned substructure sequences allowed the model to explain the order of relative clause acquisition in children, providing foundations for a usage-based theory of recursion. Error-based learning and the Dual-path architecture provide an account of the aspects of production that are universal and make explicit the link between syntax acquisition and production.

There are many areas of language processing that need to be integrated with models of language production. One area of active research is the relationship between comprehension and production. Representational models of comprehension and production have argued for strong homologies between the two in representation and procedures (Kempen & Harbusch, 2002; Vosse & Kempen, 2000). Empiricist approaches have used similar mechanisms to explain data from both domains. For example, surprisal has been a useful construct for explaining comprehension parsing behavior (Frank, 2009; Hale, 2006; Levy, 2008) and for priming in production behavior (Jaeger & Snider, 2007). Surprisal is similar to error in the Dual-path model, and Chang et al. (2006) use prediction error to explain structural priming and to model preferential looking behavior in development, which suggests that aspects of comprehension can also be explained by this expectation-based mechanism. What needs to be clarified in the future is whether these surprisal/error effects reflect the parsing process itself or learning that takes place when we comprehend utterances.

It is clear from this chapter that there are deep similarities between representational, empiricist, and brain-system approaches. At the same time, these approaches have their own unique advantages and disadvantages. Representational theories can be formulated in a way that closely matches verbal theories. For example, Performance Grammar directly encodes the distinction between functional and positional levels (Kempen & Harbusch, 2002), and therefore it can explain English accessibility data. However, as seen previously, the same distinction does not provide a good explanation of accessibility effects in scrambling in Japanese and in these cases, the tight link between representations and theory can be a disadvantage. Empiricist approaches can be easily applied to different languages as long as suitably labeled corpora are available. However, it can be difficult to determine which parameters in the model are fixed universal parts of the language processor and which can vary across corpora or languages. Therefore, one challenge for empiricist approaches is to create methods that allow them to make these distinctions.

One crucial feature of the brain systems approach is the importance of learning in explaining production behavior. Learning can help to explain variability between different structures in priming and variability across languages. However, the variability caused by learning can also lead to mismatches between verbal theories and the model (e.g., functional/positional distinction). It can be quite difficult to fit models with learned language representations to

human data. For example, because the Dual-path model was not able to explain the lexical boost, Chang et al. (2006) were forced to posit that the lexical boost was caused by a separate mechanism and this prediction was confirmed experimentally. This suggests that the limitation of the neural mechanisms in the brain may provide important constraints on theorizing about human behavior and brain system modeling can expose these constraints. Regardless of the approach that is taken, researchers should try to develop coherent integrated models of sentence production. Unified models make stronger predictions that allow for comparison and falsification (Newell, 1994). More such models are needed. As they say, talk is cheap.

References

- Arai, M., van Gompel, R. P. G., & Scheepers, C. (2007). Priming ditransitive structures in comprehension. *Cognitive Psychology*, 54(3), 218–250.
- Arnold, J. E., Losongco, A., Wasow, T., & Ginstrom, R. (2000). Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language*, 76(1), 28–55.
- Baker, C. L. (1979). Syntactic theory and the projection problem. *Linguistic Inquiry*, 10(4), 533–581.
- Bernolet, S., & Hartsuiker, R. J. (2010). Does verb bias modulate syntactic priming? *Cognition*, 114(3), 455–461.
- Bock, K. (1986a). Meaning, sound, and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12(4), 575–586.
- Bock, K. (1986b). Syntactic persistence in language production. *Cognitive Psychology*, 18(3), 355–387.
- Bock, K. (1989). Closed-class immanence in sentence production. *Cognition*, 31(2), 163–186.
- Bock, K., & Cutting, J. C. (1992). Regulating mental energy: Performance units in language production. *Journal of Memory and Language*, 31(1), 99–127.
- Bock, K., Dell, G. S., Chang, F., & Onishi, K. H. (2007). Persistent structural priming from language comprehension to language production. *Cognition*, 104(3), 437–458. (p. 86)
- Bock, K., & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General*, 129(2), 177–192.
- Bock, K., & Irwin, D. E. (1980). Syntactic effects of information availability in sentence production. *Journal of Verbal Learning and Verbal Behavior*, 19(4), 467–484.
- Bock, K., & Loebell, H. (1990). Framing sentences. *Cognition*, 35(1), 1–39.
- Bock, K., & Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, 21(1), 47–67.
- Botvinick, M., & Plaut, D. C. (2004). Doing without schema hierarchies: A recurrent connectionist approach to normal and impaired routine sequential action. *Psychological Review*, 111(2), 395–428.
- Bresnan, J., Cueni, A., Nikitina, T., & Baayen, R. H. (2004). Predicting the dative alternation. In G. Bouma, I. Kraemer, & J. Zwarts (Eds.), *Cognitive foundations of interpretation* (pp. 69–94). Amsterdam: Royal Netherlands Academy of Science.
- Cartling, B. (2008). On the implicit acquisition of a context-free grammar by a simple recurrent neural network. *Neurocomputing*, 71(7), 1527–1537.
- Chang, F. (2002). Symbolically speaking: A connectionist model of sentence production. *Cognitive Science*, 26(5), 609–651.
- Chang, F. (2009). Learning to order words: A connectionist model of heavy NP shift and accessibility effects in Japanese and English. *Journal of Memory and Language*, 61(3), 374–397.

- Chang, F., Bock, K., & Goldberg, A. E. (2003). Can thematic roles leave traces of their places? *Cognition*, 90(1), 29–49.
- Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, 113(2), 234–272.
- Chang, F., Dell, G. S., Bock, J. K., & Griffin, Z. M. (2000). Structural priming as implicit learning: A comparison of models of sentence production. *Journal of Psycholinguistic Research*, 29(2), 217–230.
- Chang, F., Janciauskas, M., & Fitz, H. (2012). Language adaptation and learning: Getting explicit about implicit learning. *Language and Linguistics Compass*, 6, 259–278.
- Chang, F., Lieven, E., & Tomasello, M. (2008). Automatic evaluation of syntactic learners in typologically-different languages. *Cognitive Systems Research*, 9(3), 198–213.
- Christiansen, M. H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science*, 23(2), 157–205.
- Clark, E. V., & Clark, H. H. (1979). When nouns surface as verbs. *Language*, 55(4), 767–811.
- Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General*, 120(3), 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(2), 214–230.
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93(3), 283–321.
- Dell, G. S., Chang, F., & Griffin, Z. M. (1999). Connectionist models of language production: Lexical access and grammatical encoding. *Cognitive Science*, 23(4), 517–542.
- Diessel, H., & Tomasello, M. (2005). A new look at the acquisition of relative clauses. *Language*, 81(4), 1–25.
- Dijkstra, A., & de Smedt, K. (1996). *Computational psycholinguistics: AI and connectionist models of human language processing*. London: Taylor & Francis.
- Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, 14(2), 179–211.
- Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, 48(1), 71–99.
- Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, 32(5), 429–492.
- Ferreira, V. S. (1996). Is it better to give than to donate? Syntactic flexibility in language production. *Journal of Memory and Language*, 35(5), 724–755.
- Ferreira, F. (2000). Syntax in language production: An approach using tree-adjoining grammars. In L. Wheeldon (Ed.), *Aspects of language production* (pp. 291–330). Cambridge, MA: MIT Press.
- Ferreira, V. S., & Yoshita, H. (2003). Given-new ordering effects on the production of scrambled sentences in Japanese. *Journal of Psycholinguistic Research*, 32(6), 669–692.
- Fitz, H. (2009). *Neural syntax*. Amsterdam: University of Amsterdam. Institute for Logic, Language, and Computation dissertation series.
- Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28(1–2), 3–71.
- Frank, S. L. (2009). Surprisal-based comparison between a symbolic and a connectionist model of sentence processing. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st annual conference of the cognitive*

science society (pp. 1139–1144). Austin, TX: Cognitive Science Society.

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(1), 58–93.

Garrett, M. F. (1988). Processes in language production. In F. J. Newmeyer (Ed.), *Linguistics: The Cambridge survey, III. Language: Psychological and biological aspects* (pp. 69–96). Cambridge, UK: Cambridge University Press.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, 15(1), 20–25.

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), 1–40.

Gries, S. T. (2005). Syntactic priming: A corpus-based approach. *Journal of Psycholinguistic Research*, 34(4), 365–399.

Hadley, R. F. (1994). Systematicity in connectionist language learning. *Mind & Language*, 9(3), 247–272.

Hale, J. (2006). Uncertainty about the rest of the sentence. *Cognitive Science*, 30(4), 643–672.

Hartsuiker, R. J., Bernolet, S., Schoonbaert, S., Speybroeck, S., & Vanderelst, D. (2008). Syntactic priming persists while the lexical boost decays: Evidence from written and spoken dialogue. *Journal of Memory and Language*, 58(2), 214–238.

Hauser, M. D., Chomsky, N., & Fitch, W. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298(5598), 1569–1579.

Hawkins, J. A. (1994). *A performance theory of order and constituency*. Cambridge, UK: Cambridge University Press.

Hawkins, J. A. (2004). *Efficiency and complexity in grammars*. New York: Oxford University Press. (p. 87)

Jaeger, T. F., & Norcliffe, E. J. (2009). The cross-linguistic study of sentence production. *Language and Linguistics Compass*, 3(4), 866–887.

Jaeger, T. F., & Snider, N. (2007). Implicit learning and syntactic persistence: Surprisal and cumulativity. *University of Rochester Working Papers in the Language Sciences*, 3(1), 26–44.

Keenan, E. L., & Hawkins, S. (1987). The psychological validity of the accessibility hierarchy. In E. L. Keenan (Ed.), *Universal grammar: 15 essays* (pp. 60–85). London: Croon Helm.

Kempen, G., & Harbusch, K. (2002). Performance Grammar: A declarative definition. *Language and Computers*, 45, 148–162.

Kempen, G., & Harbusch, K. (2004). Generating natural word orders in a semi-free word order language: Treebank-based linearization preferences for German. In A. Gelbukh (Ed.), *Computational linguistics and intelligent text processing* (pp. 350–354). Berlin: Springer.

Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science*, 11(2), 201–258.

Kidd, E., Brandt, S., Lieven, E., & Tomasello, M. (2007). Object relatives made easy: A cross-linguistic comparison of the constraints influencing young children's processing of relative clauses. *Language and Cognitive Processes*, 22(6), 860–897.

Klitssova, A. Y., & Greenough, W. T. (1999). Synaptic plasticity in cortical systems. *Current Opinion in*

Neurobiology, 9(2), 203–208.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126–1177.

Marcus, G. F. (1998). Rethinking eliminative connectionism. *Cognitive Psychology*, 37(3), 243–282.

McCloskey, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In B. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 24, pp. 109–165). San Diego, CA: Elsevier.

McDonald, J. L., Bock, J. K., & Kelly, M. H. (1993). Word and world order: Semantic, phonological, and metrical determinants of serial position. *Cognitive Psychology*, 25(2), 188–230.

Miikkulainen, R. (1996). Subsymbolic case-role analysis of sentences with embedded clauses. *Cognitive Science*, 20(1), 47–73.

Mishkin, M., & Ungerleider, L. G. (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behavioural Brain Research*, 6(1), 57–77.

Newell, A. (1994). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition*, 114(2), 227–252.

Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language*, 39(4), 633–651.

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: A critical review. *Psychological Bulletin*, 134(3), 427–459.

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(1), 56–115.

Pollack, J. B. (1990). Recursive distributed representations. *Artificial Intelligence*, 46(1), 77–105.

Rambow, O., & Lee, Y. S. (1994). Word order variation and tree adjoining grammar. *Computational Intelligence*, 10(4), 386–400.

Reali, F., & Christiansen, M. H. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language*, 57(1), 1–23.

Reitter, D., Keller, F., & Moore, J. D. (2011). A computational cognitive model of syntactic priming. *Cognitive Science*, 35(4), 587–637.

Rescorla, R. A., & Wagner, A. R. (1972). Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning. II: Current research and theory*. New York: Appleton-Century-Crofts.

Ross, J. R. (1967). *Constraints on variables in syntax*. (Unpublished doctoral dissertation). MIT, Cambridge, MA.

Rowland, C., Chang, F., Ambridge, B., Pine, J. M., & Lieven, E. V. (2011). *The development of abstract syntax: Evidence from structural priming and the lexical boost*. Presented at the Conference on Architectures and Mechanisms for Language Processing, September 1–3, 2011, Paris.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536.

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(3), 417–451.

- St. John, M. F., & McClelland, J. L. (1990). Learning and applying contextual constraints in sentence comprehension. *Artificial Intelligence*, 46(1–2), 217–257.
- Tanaka, M. N., Branigan, H. P., McLean, J. F., & Pickering, M. J. (2011). Conceptual influences on word order and voice in sentence production: Evidence from Japanese. *Journal of Memory and Language*, 65(3), 318–330.
- Thothathiri, M., & Snedeker, J. (2008a). Syntactic priming during language comprehension in three- and four-year-old children. *Journal of Memory and Language*, 58(2), 188–213.
- Thothathiri, M., & Snedeker, J. (2008b). Give and take: Syntactic priming during spoken language comprehension. *Cognition*, 108(1), 51–68.
- Vosse, T., & Kempen, G. (2000). Syntactic structure assembly in human parsing: A computational model based on competitive inhibition and a lexicalist grammar. *Cognition*, 75(2), 105–143.
- Wilson, M. P., & Garnsey, S. M. (2009). Making simple sentences hard: Verb bias effects in simple direct object sentences. *Journal of Memory and Language*, 60(3), 368–392.
- Yamashita, H., & Chang, F. (2001). “Long before short” preference in the production of a head-final language. *Cognition*, 81(2), B45–B55.

Franklin Chang

Franklin Chang, School of Psychology, The University of Liverpool

Hartmut Fitz

Hartmut Fitz, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands

