

Language adaptation and learning: Getting explicit about implicit learning

Franklin Chang^{1*}, Marius Janciauskas¹ and Hartmut Fitz²

¹University of Liverpool and ²Max Planck Institute for Psycholinguistics Nijmegen

Abstract

Linguistic adaptation is a phenomenon where language representations change in response to linguistic input. Adaptation can occur on multiple linguistic levels such as phonology (tuning of phonotactic constraints), words (repetition priming), and syntax (structural priming). The persistent nature of these adaptations suggests that they may be a form of implicit learning and connectionist models have been developed which instantiate this hypothesis. Research on implicit learning, however, has also produced evidence that explicit chunk knowledge is involved in the performance of these tasks. In this review, we examine how these interacting implicit and explicit processes may change our understanding of language learning and processing.

Introduction

Human languages have rules that must be learned. If an English speaker says ‘I coffee want’, they might get a confused look, rather than a drink. This is explained by assuming that speakers have internal rules or constraints that they apply during language processing (e.g. ‘I want coffee’ is more acceptable than ‘I coffee want’). Speakers must learn these linguistic constraints, because they differ across languages. For example in Japanese, the equivalent sentence for ‘I want coffee’ would place the Japanese word for ‘want’ after the word for ‘coffee’. Despite the fact that linguistic constraints are learned in language acquisition and used in language processing, research in each of these domains has traditionally operated independently. Language processing theories have assumed that the representations in adult language are static and unchanging (van Gompel and Pickering 2007; Levelt 1993), while language acquisition theories have focused on how representations change over time (Pinker 1984; Tomasello 2003). A general theory of language needs to explicitly link these two aspects.

One phenomenon that links acquisition and processing is linguistic adaptation, where the representations that support language processing change in response to language input. For example, speech segmentation mechanisms in infants can be influenced by exposing them to syllable sequences (e.g. *tupirogolabubidakupadoti*), where statistical regularities between syllables provide cues for word boundaries (Aslin et al. 1998; Saffran et al. 1996). Likewise, both infant and adult linguistic knowledge of permissible combinations of sounds (phonotactic constraints) can be changed by experience with sound streams where the phonotactic regularities have been experimentally manipulated (Chambers et al. 2003, 2010; Dell et al. 2000; Goldrick 2004; Goldrick and Larson 2008; Onishi et al. 2002; Seidl et al. 2009; Conrad F. Taylor and Houghton 2005; Warker et al. 2008, 2009). In addition to these phonological adaptation effects, similar results have been found within the word production system. For example, repetition priming (faster naming of the same picture) and semantic interference (inhibited naming of semantically-related pictures)

have been found to persist (Cave 1997; Howard et al. 2006; Mitchell and Brown 1988) and it has been argued that these effects are due to learning within the word production system (Oppenheim et al. 2010; for a theoretical analysis of learning and priming, see Gupta and Cohen 2002). Further evidence for linguistic adaptation comes from a phenomenon known as syntactic or structural priming, where speakers tend to reuse previously heard sentence structures in their own utterance generation (Bock 1986; Pickering and Ferreira 2008). For example, the same picture of a man handing a book to a woman could be described using two different sentence structures: a prepositional dative structure like ‘the man gave the book to the woman’ or a double object dative structure like ‘the man gave the woman the book’. However, when speakers hear another prepositional dative sentence like ‘A child threw a ball to his friend’ before describing the picture, they are more likely to describe it with the prepositional dative (‘the man gave the book to the woman’). Importantly, structural priming occurs between sentences that differ in words, semantic roles, and prosody (Bock 1989; Bock and Loebell 1990), which has been used to argue that priming is abstract and occurs purely on the basis of structural similarity independent of meaning. In addition, structural priming has been found to persist over time (Bock and Griffin 2000; Bock et al. 2007). Here, linguistic adaptation is a term that labels behavioral phenomena without specifying the underlying mechanism. It can describe both short-term (Branigan et al. 1999) and long-term changes (Kaschak 2007; Kaschak et al. 2011; Wells et al. 2009), and applies to the acquisition of new structures (Kaschak and Glenberg 2004) or changes in the processing fluency or accessibility of existing representations (Smith and Wheeldon 2001).

The fact that some linguistic adaptation effects persist suggests that learning has taken place. But since language learning theories have focused on the acquisition of abstract linguistic knowledge such as the position of verbs (Mazuka 1998; Pinker 1984), it is difficult for them to explain these adaptation phenomena in adults. Instead, theories of adaptation have drawn inspiration from theories of memory and learning, in particular from work on artificial grammar learning (AGL). As pioneered by Reber (1967), these studies involve arbitrary symbols (e.g. letters, shapes, words) that are combined into sequences called items. The configuration of symbols in an item is governed by a set of rules. For example, a string of letters like MVST is consistent with a rule that V can be followed by S but never by T. A set of such rules specifies a grammar, which determines whether strings are grammatical or ungrammatical. A simple type of grammar that is often used in AGL is a finite state grammar (Figure 1).

In these studies, people are exposed to a large number of such items and are later asked to categorize new items based on regularities that they have induced from the training set. Many studies have found that participants’ performance on a variety of AGL tasks indicates the acquisition of representations that reflect rule-like behavior (for a review, see Pothos 2007). Since these studies show how adults can learn regularities, they provide a model that can help us to understand linguistic adaptation phenomena.

To explore the nature of the mechanism that supports the learning of these regularities, Cleeremans and McClelland (1991) conducted a grammar learning article and then fitted the data with a computational model (for a recent review, see Cleeremans and Dienes 2008). Their article used a variant of Reber’s paradigm, where instead of categorizing whole strings, participants had to reproduce sequences by typing them out on a keyboard. The sequences were signaled by dots at different positions on the screen that represented the symbols in the grammar. As participants produced these sequences, their speed improved and this indicates that the participants were able to acquire the regularities in the finite state grammar that generated the dot sequences. Furthermore, the continual

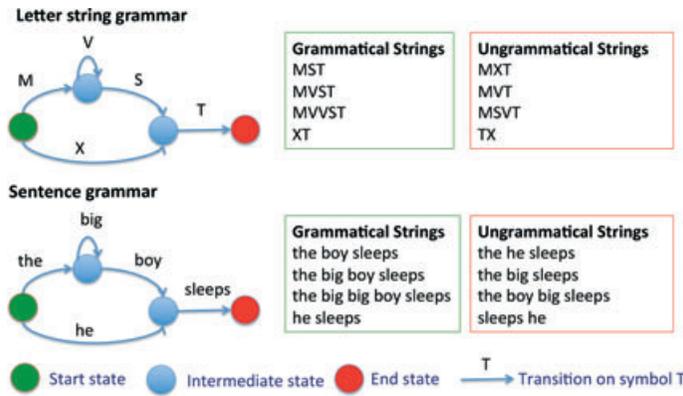


Fig 1. Finite state grammar that generates letter strings and an equivalent sentence grammar for a subset of English.

growth in the knowledge of the grammar over training is broadly similar to the changes that are seen in adult linguistic adaptation studies.

To explain this data, Cleeremans and McClelland employed a computational model developed by Elman (1990) called the simple recurrent network (SRN). An SRN is a connectionist model that learns to predict the next letter in a sequence from the previous letter (for more about these models, see Cleeremans et al. 1989; Christiansen and Chater 1999). Their model had three layers (input, hidden, and output) where the input layer projected to the output layer through the hidden layer. The input and output layers represented the letters from the artificial language by means of distinct units (Figure 2). Sequences were presented to the input layer one letter at a time by switching on the corresponding unit. Each of the input units was connected to all of the hidden units by a link with an initially-random weight (blue arrows in Figure 2). The activation of the input layer (1 for the input letter and 0 for the rest of the units) was multiplied by these weights to create the activation pattern for the hidden layer. The hidden layer activation was spread to the output layer where an activation pattern was produced. In addition, the

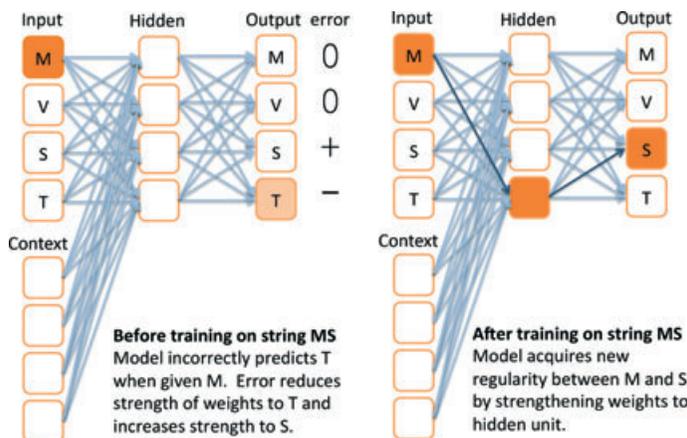


Fig 2. A simple recurrent network learns to predict the letter S from the input letter M.

model had a context representation, which held a copy of the hidden layer's activation pattern generated by the previous input. This provided a memory for the earlier parts of the sequence and allowed the model to learn longer distance dependencies. Thus, the model was able to produce sequences by generating symbols one at a time using the previous symbol and context to guide those choices.

The model learned the regularities in the input by means of an error-based learning algorithm, where the difference between the model's predicted output and the actual next letter (the error signal), was used to change the connection weights to strengthen the network's predictive abilities. After a number of sequences, the model was able to acquire an internal representation of the input regularities. To illustrate, let's assume that the model is exposed to a sequence where S follows M (Figure 2). When M is given to the input layer, activation spreads to the hidden layer where it combines with the context representation and a letter prediction is generated at the output layer. Since the weights that modulate the spreading of activation are initially random, the model might not predict the next letter correctly (e.g. T is active). The error is then calculated and is used to change the input-hidden and hidden-output weights to ensure that S is a more likely outcome following letter M on future trials (black arrows in Figure 2). Gradually, sequences like MS become more predictable (less error) than ungrammatical sequences like MT.

This way the network acquires knowledge about the relationship between input items as a by-product of processing those items. Cleeremans and McClelland (1991) showed that learning in an SRN could reproduce the development of their participants' knowledge of the grammar. Since the model both acquired the grammar and exhibited continual improvement due to learning, this model provides a template for thinking about how language acquisition could be related to linguistic adaptation.

Implicit Sequence Learning and Sentence Production

The SRN was originally devised to learn linguistic constraints from language-related sequences (Elman 1990, 1993). Instead of predicting the next letter, the model would predict the next word in sentences based on the previous word and the context. For example, if the model was given the word 'the', it would learn that a particular set of words tends to follow that word (e.g. 'boy') and these word-to-word constraints in the connection weights approximate a kind of syntactic knowledge that could be interpreted as a rule that articles tend to be followed by nouns (Mintz et al. 2002). The fact that SRNs can acquire syntactic constraints suggests that they might be able to account for the adult adaptation of syntactic knowledge in structural priming studies.

To make this link, Chang and colleagues developed a connectionist model of sentence production and syntactic development (Chang 2002, 2009; Chang et al. 2006; Fitz 2009; Fitz and Chang 2008). The model had a dual pathway architecture (*Dual-path model*) that combined an SRN with a meaning network containing the message that the model was attempting to convey (Figure 3). The message represented the meaning of the sentence and different sentences could have similar messages as in 'The child gave Sally a book' and 'The child gave a book to Sally' (see message in Figure 3). The Dual-path model learned a language by attempting to predict sentences, word-by-word, based on the sequential constraints in the SRN and message related information in the meaning network. Before it could exhibit priming, the model had to acquire the appropriate syntactic representations. The SRN in the Dual-path model developed syntactic categories in a distributional learning process (Elman 1990). The model then learned how to use meaning to sequence these abstract syntactic categories in sentence generation. These language

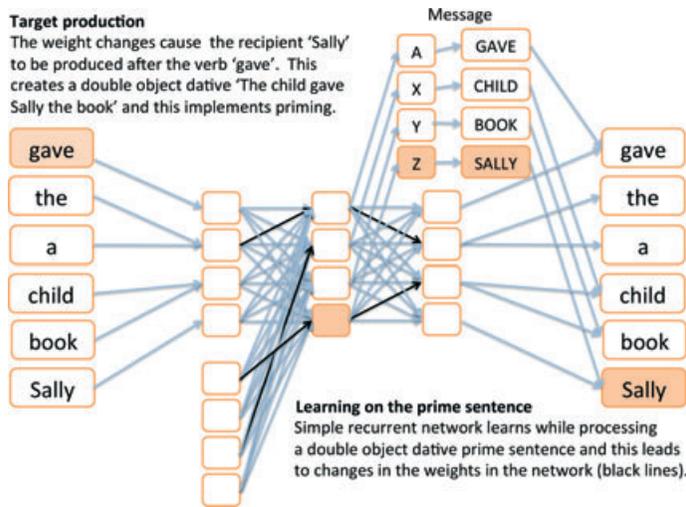


Fig 3. Simplified Dual-path model: Structural priming as implicit sequence learning.

learning processes are powerful, allowing the model to learn both English-like and Japanese-like languages (Chang 2009). To enable priming, the model was given a prime sentence like 'John threw the man a ball' with error-based learning turned ON and the error between the predicted and the actual sentence led to weight changes in the model. Then a new target message, which could be described in two ways (like the example with 'gave' above), was placed into the model's meaning system. Changes in the connection weights, although small, biased the model's description of the target message making it more likely to use the structure of the prime sentence (e.g. 'The child gave Sally a book', Figure 3).

To examine the persistence of priming in the model, Chang et al. (2006) tested it on a task used by Bock and Griffin (2000). In this article, the prime and target sentences were separated by a different number of structurally-unrelated filler sentences. Since learning was activated during the processing of the prime and fillers, it was possible that learning from the filler sentences would interfere with the prime-target priming. Instead, it was found that the priming effect was the same regardless of whether there were 0 or 10 fillers separating the target from the prime. This showed that the model's learning mechanism was able to account for the persistence of priming. In addition, although the model started off with no knowledge of English syntax, the same learning mechanism acquired appropriate syntactic representations to explain a wide range of structural priming results (Bock 1989; Bock and Loebell 1990; Bock et al. 2007; Chang et al. 2003; Pickering and Branigan 1998). These results show that structural priming behavior can be explained as implicit language learning.

Since the Dual-path model provided a computational account of structural priming, its mechanism also generated novel predictions about the nature of priming. The model used a form of error-based learning called back-propagation of error (Rumelhart et al. 1986) and this algorithm is critical for acquiring the abstract syntactic representations that support adult priming. Error-based learning predicts that adaptation should only occur if the prime sentence mismatches the system's expectation. Consequently, less expected structures should result in stronger priming effects. One phenomenon that can influence expectations is verb bias (Garnsey et al. 1997; Michael P. Wilson and Garnsey 2009).

Verb bias reflects the tendency of particular verbs to appear in particular structures. For example, the verb ‘threw’ occurs in the prepositional dative structure (e.g. ‘the man threw the ball to the girl’) more often than in the double object dative structure (e.g. ‘the man threw the girl the ball’). Therefore, if a prime contained a verb that had a bias towards a different structure it would produce a greater amount of error and stronger weight changes, resulting in stronger priming effects. Support for this prediction has been found in experimental and corpus-based studies (Bernolet and Hartsuiker 2010; Jaeger and Snider 2007). In addition, verb bias itself is acquired in the model by the same mechanism that learns which verbs occur in which structures (subcategorization) during language acquisition (see Juliano and Tanenhaus (1994) and Rohde (2002) for evidence that an SRN can learn verb biases; Chang (2002, pp. 637–40), provides evidence that learned verb-structure associations in the Dual-path model influence generalization to a novel construction).

There is also a phenomenon related to verb overlap between the prime and the target called the lexical boost, where the magnitude of structural priming increases when the prime and target utterances share the same verb (Pickering and Branigan 1998; see Cleland and Pickering 2003; for noun-based boost effects). One account of the lexical boost is provided by Pickering and Branigan’s (1998) Residual Activation theory (Figure 4). In this theory, verb nodes are linked to combinatorial nodes that determine which structure will be used to express a message. For example, the verb node for ‘throw’ can be linked to NP-NP and NP-PP combinatorial nodes. If the NP-NP node is more activated than the NP-PP node, a double object dative will be produced, otherwise a prepositional dative is produced. Priming in this model is due to the fact that when a node is activated, some of that activation remains as residual activation. In addition, the link between the verb and the combinatorial node also has residual activation associated with its recent use. Thus, all sentences that are experienced leave residual activation in combinatorial nodes and the verb links. If prime and target sentence have different verbs, the residual activation in the combinatorial node increases the likelihood that the same structure is used and this creates abstract structural priming. If they share the verb, the residual activation

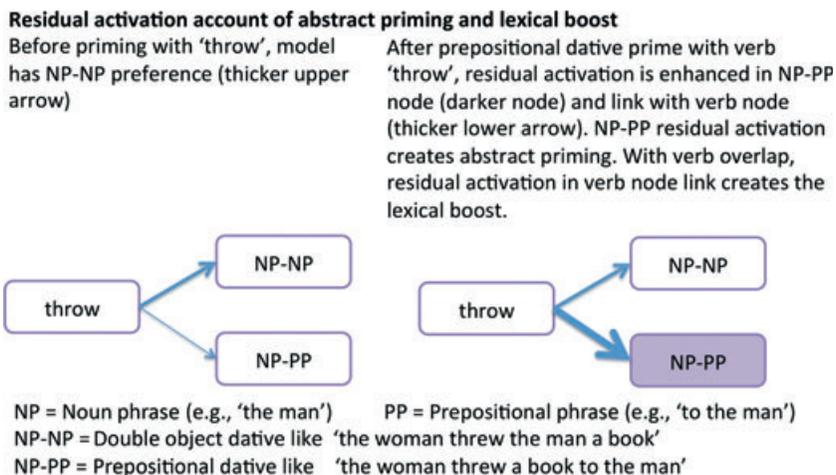


Fig 4. Residual Activation account of Pickering and Branigan (1998).

in the link further increases the likelihood of activating the combinatorial node, which generates the lexical boost (Figure 4).

It is natural to assume that a common verb-structure mechanism supports the lexical boost and verb bias. Verb bias learning requires a slow learning mechanism that updates a frequency representation after each verb-structure pairing. Coyle and Kaschak (2008) presented 10 consecutive same-structure primes with the same verb and they were able to modulate verb bias by only 15%, which means that each verb-structure prime influenced structural choice by about 1.5%. In contrast, lexical boost studies have found massive priming effects, where structural choices can be modulated by 73% (Hartsuiker et al. 2008; Exp. 1). The SRN in the Dual-path model can learn that certain verbs are associated with certain internal nodes that represent verb classes (Figure 5, top) and the magnitude of these associations depends on the model's learning rate. A small learning rate will create links between verbs and internal nodes that encode the frequency of verb-structure pairs in the input (Figure 5, bottom). But a large learning rate, which is needed to explain large lexical boost effects, will cause verb-structure associations to fluctuate violently and eventually knowledge about the frequency of verb-structure associations will be lost. This is an example of catastrophic interference, where newly learned knowledge overwrites old knowledge (McCloskey and Cohen 1989). For these reasons, Chang et al. (2006) argued that the lexical boost was due to a separate mechanism that is different from the implicit learning mechanism that supports abstract priming. The lexical boost mechanism should create large short-term effects that would not persist in the language system to avoid large changes in verb biases. Support for this dual-mechanism account of priming has been growing. Hartsuiker et al. (2008) found that structural priming was persistent, but the lexical boost dissipated quickly. Rowland et al. (2011) have found that while abstract priming has a consistent magnitude in both children and adults, the lexical boost is non-significant in 3- to 4-year-old children and very large in adults. This indicates that the two priming effects can be dissociated.

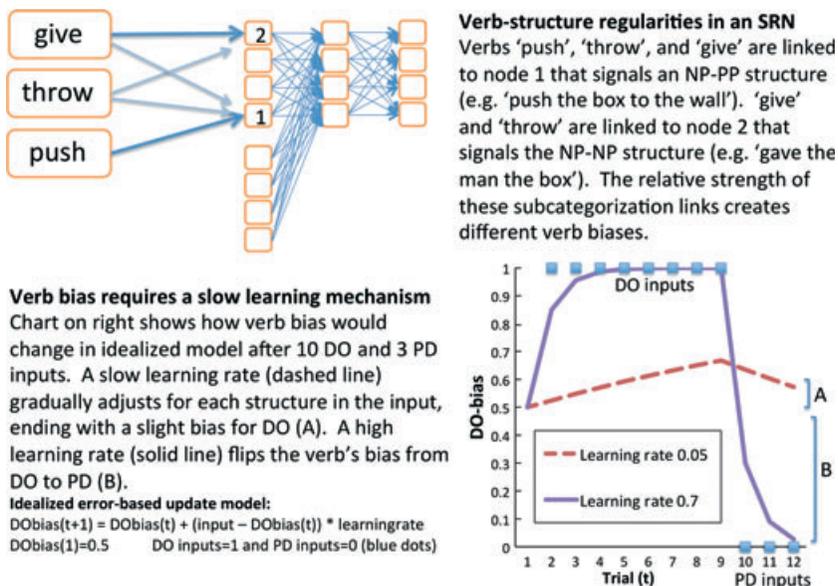


Fig 5. Verb-structure learning in a simple recurrent network.

To summarize, evidence of the existence of linguistic adaptation requires a way to link language learning and language processing. Work on implicit sequence learning suggested that learning in an SRN could provide a mechanism for explaining linguistic adaptation. The Dual-path model confirmed this hypothesis by using the same learning algorithm for both syntax acquisition (learning new structures) and structural priming (tuning existing structures). Furthermore, the computational assumptions of the model, namely its implementation as slow error-based learning, made correct predictions about verb bias and lexical boost effects in structural priming studies. In contrast to traditional accounts where one set of verb-structure links support all verb-based phenomena (e.g. Pickering and Branigan 1998), the Dual-path model suggests that multiple verb-structure association mechanisms are needed: an implicit mechanism which gradually learns subcategorizations and biases, and a separate explicit mechanism which can yield the large changes that are evident in the lexical boost.

Memory, Learning, and Processing

The dual mechanism account of priming, where different mechanisms support abstract and lexical priming effects, was influenced by theories concerned with the nature of multiple memory systems (Eichenbaum and Cohen 2004). These theories arose from the clinical history of a patient called HM (Corkin 2002; Squire 2009). Due to removal of the medial temporal lobe, he developed anterograde amnesia and lost his ability to recall facts and events that happened after the operation, while retaining his ability to learn new behaviors that were only visible in the performance of some action (e.g. learning to trace figures in a mirror). His memory dissociations were important for drawing the distinction between explicit memory, characterised by memories that can be accessed intentionally (everyday events and facts), and implicit memory, which is expressed as experience-induced change in performance (Eichenbaum and Cohen 2004; Schacter 1987; Squire 2004). When memories are changed by experience and these changes persist, then learning can be said to have occurred. Just like memory, learning can also be implicit (learning to ride a bike) or explicit (learning some facts) (Seger 1994).

Two paradigmatic tasks for studying implicit learning are AGL and serial reaction time (SRT) tasks. In a typical AGL experiment, participants are exposed to letter sequences (e.g. MVVST) that unbeknownst to them are constructed from a set of rules that specify which letter can follow which other letter (Reber 1967). In the test phase, they are presented with a new set of strings, half of which obey the rules, while the other half violate them. Their task is to distinguish grammatical strings from ungrammatical ones (see Figure 1 for more examples). Participants usually perform significantly above chance suggesting that they have learned some of the regularities in the input. In SRT studies (e.g. Nissen and Bullemer 1987), participants are given a reaction time task, where they have to press buttons corresponding to locations on a computer screen where stimuli appeared. The sequence of stimulus locations is generated either by a fixed rule or a finite state grammar (Cleeremans and McClelland 1991). The general finding here is that participants become faster at pressing the corresponding buttons as they process blocks of sequences (for reviews, see Clegg et al. 1998; Abrahamse et al. 2010). The increase in speed indicates that participants have acquired some knowledge about the rules that generated the sequences. While AGL and SRT are different in their manner of stimulus presentation, both paradigms have found that humans can incidentally learn regularities that they exhibit in their behavior without necessarily being able to fully verbalize the acquired knowledge.

Both AGL and SRT learning involve similar formal grammars, but there is little theoretical agreement about the mechanisms that support this learning. Some theories favor a type of mechanism that records the *transitional probabilities* between symbols (e.g. letters, words) in the language or, in other words, are sensitive to the frequency of symbols co-occurring together (e.g. if only V or S can follow M and both occur equally often, then the transitional probability of V following M is 0.5). In case of language, the transitional probability for ‘boy’ after ‘the’ should be higher than the transitional probability of ‘oak’ after ‘the’, because ‘boy’ follows ‘the’ more often than ‘oak’. SRNs are an example of a mechanism that is sensitive to these types of probabilities. Other theories favor chunking mechanisms that record how often particular sets of symbols co-occur across different strings (Perruchet and Pacteau 1990). For example, people may judge string grammaticality based on the knowledge of frequently occurring chunks like MV or VST. Likewise, in natural language, if people represent ‘the boy’ as a frequent chunk, they should prefer that sequence over a less frequent chunk like ‘the oak’.

Formally, these mechanisms are related. The transitional probability of ‘boy’ after the word ‘the’ (or V after M) is the chunk frequency of ‘the boy’ divided by the frequency of ‘the’ (middle of Figure 6). Transitional probabilities are objective properties of symbol sequences, but they only approximate the *transition probabilities* that guide state changes in a finite state grammar or an SRN. The transition probability is the same as transitional probability if each word in the language is a unique state within the grammar. But if the same word is associated with multiple states, then transitional probabilities will not match transition probabilities. For example, the likelihood of the word ‘boy’ and ‘oak’ should differ in subject and object position (e.g. ‘the boy cut the oak’ sounds better than ‘the oak cut the boy’). To explain this, we can posit two ‘the’ states, one for subject and one for object and the probabilities can be set accordingly (bottom of Figure 6). If language requires a state-based mechanism, then the transition between two words is the result of internal states that are not available to conscious inspection and hence implicit. In contrast, chunk representations involve simple conjunctions of explicit words (top of Figure 6) and hence, these representations are thought to be available to conscious awareness or declarative expression (Anderson et al. 2004; Perruchet and Pacteau 1990). Thus, computational mechanisms that support chunking or transition probabilities make different

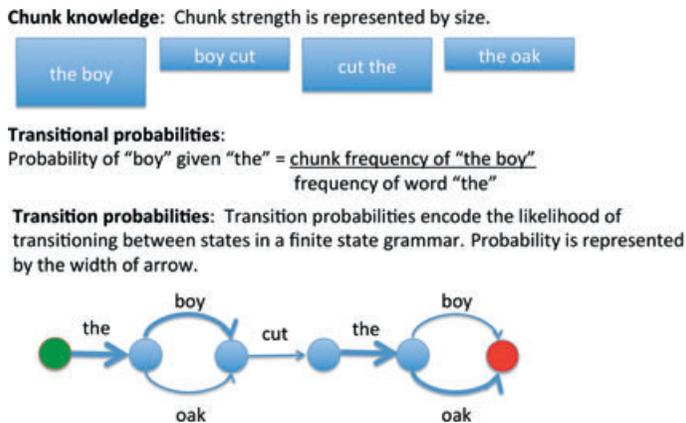


Fig 6. Chunks, transitional probabilities, and transition probabilities for a finite state grammar for the sentence “the boy cut the oak”.

predictions about the degree to which these representations are available to explicit awareness.

Some have argued that AGL could be influenced by explicit knowledge of fragments or chunks of symbols (Dulany et al. 1984; Gómez 1997) or that this knowledge is sufficient to explain people's performance in grammaticality judgment tasks (Perruchet and Pacteau 1990; Perruchet et al. 1997). While there is some evidence for explicit fragment knowledge in SRT studies (Buchner et al. 1998; Shanks and Johnstone 1999), SRT grammar learning seems to depend less on explicit chunk knowledge (Jiménez 2008). For example, Jiménez et al. (1996) found that grammatical knowledge was expressed in reaction times in an SRT task, but not in a sequence generation task indicating that the knowledge was not explicit enough to generalize to a new task. This suggests that both explicit chunks and implicit transition probabilities are used in both AGL and SRT studies and that the learning of these formally-similar grammars cannot be reduced to a single mechanism. Furthermore, dissociations in amnesic patients between chunk knowledge and abstract rule learning support the view that both of these components are needed (Knowlton and Squire 1996; Knowlton et al. 1992).

Some have attempted to disentangle the relationship between these two types of knowledge. One task that has been used in the memory literature to distinguish implicit and explicit processing is the process dissociation procedure (Jacoby 1991). Destrebecqz and Cleeremans (2001) created a variant of this procedure within an SRT task. After training, they instructed participants to produce sequences that were similar to the training sequences (inclusion condition) and those that did not appear in the training set (exclusion condition). If sequence learning is due to explicit knowledge, then participants in the exclusion condition should be able to consciously avoid producing the training sequences. They found that participants in some conditions were unable to use their explicit fragment knowledge, which confirms the implicit nature of SRT behavior. Wilkinson and Shanks (2004) attempted to replicate these results, but found that participants performed successfully in both inclusion and exclusion conditions (see also Dienes et al. 1995). Finally, Fu et al. (2008) found that motivation was part of the reason for the difference in these two studies suggesting that implicit knowledge can dominate when participants are not highly motivated.

Another way to distinguish different components of learning is to manipulate attentional resources by including a distractor task (e.g. Curran and Keele 1993). Early work suggested that implicit learning was mostly insensitive to attentional distraction (Cohen et al. 1990; Jiménez and Méndez 1999), but later work has found that it was possible to interfere with this type of learning (Jiang and Chun 2001; Shanks et al. 2005). One way to reconcile these seemingly incompatible results has been to argue that there are multiple types of attention and some of these can interfere with implicit components of learning. For example, Rowland and Shanks (2006) found that visual distractors in an SRT task did not influence sequence learning. They explained this dissociation by arguing that there are two kinds of attention; input attention (e.g. when one has to ignore other stimuli) and central attention (when working memory is loaded by an additional task like counting stimuli). They argued that only central attention influenced implicit sequence learning. Other studies have suggested that attention has selective effects. Cohen et al. (1990) demonstrated that attentional manipulations did not affect sequence learning if elements in the sequence could be uniquely associated with their neighbors (as in the number sequence 15243) but it interfered with sequences where symbols appeared in different orders in different parts of the sequence (e.g. 132312). They argued that processing the latter strings required hierarchical knowledge that would be used to distinguish different

exemplars of the same symbol and that the acquisition of such knowledge required attention. The idea that learning these different types of sequences requires different mechanisms is supported by the finding that amnesics could learn relations between adjacent symbols, but had more trouble with higher order associations, where items could only be predicted based on a combination of previous symbols (Curran 1997). In sum, although there is no unified account of how attention influences sequence learning, it does seem that attention can be used to differentiate processing components.

There are still many basic issues that have not been resolved within the AGL and SRT literature. Different tasks seem to involve different mixtures of implicit and explicit knowledge. The difficulty in distinguishing these components in non-linguistic tasks suggests that language knowledge might also be composed of tightly interacting explicit and implicit systems. The reviewed AGL-SRT studies suggest that these systems can be dissociated in amnesic patients and are differentially sensitive to conscious and attentional manipulations when learning complex stimuli (e.g. multiple languages, languages with hierarchical structure). Similar techniques might also be useful for fractionating language into its components.

Linking Language with Memory and Learning Theories

To recapitulate, linguistic adaptation phenomena indicate that learning and processing are closely intertwined. Studies using AGL-SRT tasks have found that the same mechanisms seem to be involved in both the acquisition and tuning of grammatical knowledge. Here, we will review neuropsychological and behavioral evidence that suggests that the mechanisms used in AGL-SRT may also be involved in language learning. In addition, computational models of linguistic adaptation and AGL-SRT studies have suggested that explicit and implicit components may be at work in the learning of grammatical constraints. In this section, we will point to evidence that a similar dissociation may exist in language as well.

There is a growing body of evidence that AGL and SRT behaviors recruit similar brain regions as those used for language processing (Conway and Pisoni 2008; de Vries et al. 2011). Brain stimulation showed that Broca's area, which has been classically associated with language (Geschwind 1970; Grodzinsky and Santi 2008), is also involved in AGL (de Vries et al. 2010). Functional brain imaging showed that common areas are activated in grammatical violations in a serial AGL task and natural language processing (Pettersson et al. 2012). Agrammatic aphasics display an impairment in AGL suggesting that aphasia affected both language and sequence learning (Christiansen et al. 2010). Behavioral data showed that children's performance in non-linguistic SRT tasks directly predicted the persistence of structural priming effects (Kidd 2012). Finally, Misyak et al. (2010a,b) demonstrated that the ability to learn non-linguistic and linguistic non-adjacent dependencies correlated within individuals. These studies strengthen the view that the mechanisms that support AGL-SRT learning may also be involved in linguistic adaptation.

The behavioral paradigm that provides the strongest link between AGL-SRT phenomena and language is statistical learning (SL), where participants extract linguistic regularities from speech samples made up of symbols (e.g. syllables or words) that exhibit particular distributional properties (Conway and Christiansen 2001; Gómez and Gerken 2000; Perruchet and Pacton 2006). It has been shown that preverbal infants could discover word boundaries by using transitional probabilities in an artificial speech stream (Saffran et al. 1996, 1997) and in samples of real speech (Pelucchi et al. 2009). The same

mechanisms are also thought to govern the learning of word sequences (Gómez and Gerken 1999; Saffran and Wilson 2003), word classes (Hunt and Aslin 2010; Mintz 2002), and rudimentary phrase structure (Saffran 2001). Similar paradigms have been used to study speech segmentation (Frank et al. 2010) and the acquisition of syntactic constraints in adults (Morgan and Newport 1981; Thompson and Newport 2007; Wonnacott et al. 2008). These paradigms demonstrate that children and adults can acquire, mostly without conscious awareness, rich statistical information about structure at multiple linguistic levels.

Although SL phenomena have often been explained in terms of transitional probabilities, they can also be explained by chunking mechanisms (French et al. 2011; Perruchet and Pacton 2006), Bayesian learning (Frank et al. 2010), or SRNs (Mirman et al. 2010). Chunking mechanisms in particular have provided accounts of language development using statistical information from child language corpora (Chang et al. 2008; Freudenthal and Pine 2005; Freudenthal et al. 2007). It has been argued that unique evidence in support of these accounts is provided by studies showing that children and adults have stored four-word sequences as chunks (Arnon and Snider 2010; Bannard and Matthews 2008). But these findings are also explainable within other mechanisms like SRNs (Rodríguez 2003, e.g. showed that an SRN could learn 5–6 word sequences). The difficulty in differentiating these learning mechanisms from one another indicates that they capture similar sets of regularities from the input. If multiple mechanisms are at work, then it might be more fruitful to differentiate them in terms of the explicitness of their representations.

Knowledge acquired in SL is usually characterized as implicit, but recent studies have begun to explore how explicit factors like attention might be involved. Toro et al. (2011) found that adults could learn vowel-matching rules in both adjacent (AAB) and non-adjacent (ABA) syllables when exposed to syllable streams. But a secondary distractor task impaired learning of non-adjacent regularities more than adjacent ones, which suggests that attention plays a role in SL of rules. This mirrors the findings of Cohen et al. (1990) in an SRT task showing that attentional distraction impaired participants' abilities to learn hierarchical structures. If SL has components that differ in their sensitivity to conscious processes like attention, then the process dissociation paradigm may be useful for distinguishing these components. Franco et al. (2011) applied this paradigm in an SL article where people had to learn two artificial languages. They found that participants were consciously aware of the acquired knowledge and were able to control the language that they responded with. This task required that participants applied cognitive control in their language use and it has been hypothesized that cognitive control is regularly involved in language processing (Novick et al. 2010) and learning (Davidson and Indefrey 2011). Critically, the implicit SL mechanisms discussed thus far do not specify how cognitive control, conscious awareness, and attention influence learning, and that suggests that these effects are supported by separate systems or mechanisms.

Since infants and children presumably have different attention and cognitive control abilities compared to adults, we should expect to see similar dissociations in language development. Some SL studies of segmentation have found similar behavior in children and adults (Saffran et al. 1997), while other studies have found that adults are better than 7-year-old children in the SL of phrase structure (Saffran 2001). This is consistent with the idea that adults may be engaging in more explicit strategies for higher-level representations. These developmental dissociations can also be seen structural priming. Rowland et al. (2011) found that abstract structural priming had a consistent magnitude in 3–4 year olds, 5–6 year olds, and children and adults. The lexical boost, however, was not present in the youngest children and it grew over development. Although these

dissociations could be specific to language, similar patterns have been observed in a visual SRT task. Thomas and Nelson (2001) found that 7- and 10-year olds' speed at tracking an object's location improved over blocks as they implicitly acquired sequential regularities. While implicit learning was similar over both ages, the ability to explicitly generate the sequence was higher in the 10-year olds than the 7-year olds. Taken together, these studies suggest that there are mechanisms that are relatively constant over development (e.g. segmentation, abstract priming, visual tracking) and those that grow in this period (e.g. phrase structure learning, lexical boost, visual prediction).

In addition to these developmental dissociations, there are also neuropsychological dissociations. Traditionally, amnesia has been thought to spare language abilities as patients like HM were able to generate fairly complex sentences (Kensinger et al. 2001; Skotko et al. 2005). There are also reports of amnesics acquiring the vocabulary and syntax of a new language (Hirst et al. 1988). Furthermore, amnesics exhibit abstract structural priming (Ferreira et al. 2008), which suggests that they have intact mechanisms for changing abstract grammatical knowledge. However, other studies have documented a range of language-related deficits in HM, suggesting that the medial temporal lobe may be critical for maintaining some language functions (MacKay and Hadley 2009; MacKay and James 2002; MacKay, Burke, and Stewart 1998a; MacKay et al. 2011, 1998b). Of particular interest is evidence that HM showed impairments in the ability to update his idiomatic/clichéd language knowledge (MacKay et al. 2007). Furthermore, intracranial ERP studies where electrodes were placed directly on the medial temporal lobe have found sensitivity to syntactic and semantic operations (Patric Meyer et al. 2005). Together, these findings suggest that the medial temporal lobe is involved in language processing and that it may partially be responsible for dissociations between idiomatic and abstract knowledge.

Further support for the idiomatic-abstract dissociation can be found in adult production. Idioms like 'kicked the bucket' (meaning 'died') act like chunks, in that they do not allow standard syntactic operations on their components (e.g. 'John kicked the ball and the bucket' cannot mean that 'John kicked the ball and died'). Even though idioms have these chunk-like properties, work in sentence production has found that speech errors are sensitive to their internal syntactic structure (Cutting and Bock 1997). Thus idioms have a dual nature in that they act both as chunks and as syntactically-structured sequences of words.

If the lexical boost involves more explicit memory than abstract structural priming, then we might predict that the chunk nature of idioms should be evident in the lexical boost, while the structural nature of idioms should be involved in abstract priming. Konopka and Bock (2009) provided data that support this hypothesis. They found that an utterance like 'The New York Mets brought up the rear' with an idiomatic verb-particle combination like 'brought up' (meaning 'occupied') increased the likelihood of speakers to use a post-verb order like 'the toddler threw away one of his toys' rather than the post-object order 'the toddler threw one of his toys away'. Since different verbs and prepositions were used in this article, this finding can only be explained as abstract priming based on the idiom's internal syntactic structure (e.g. verb preposition). They also tested the lexical boost with these particle verbs and found that a lexical boost effect only occurred when both the verb and particle were repeated, but not when just the verb or particle was repeated. Thus the lexical boost seems to depend on the chunk nature of the idiom, while abstract priming across verbs seems to depend on internal structure.

One way to make the idea of multiple interacting systems more concrete is to combine the Dual-path model with ideas from the complementary cortical/hippocampal systems theory (McClelland et al. 1995; O'Reilly and Rudy 2001). In these theories, the hippocampus in the medial temporal lobe has bidirectional connections with various cortical

areas and the hippocampus represents sparse conjunctive codes that encode associations between these areas (Figure 7). If we assume that these conjunctive code units are connected to lexical or semantic units for verbs and structural units such as the hidden layer in the SRN, then they will encode verb-structure associations. An important assumption in these theories is that cortical systems have a slow learning rate and hippocampal links with cortical systems have a fast learning rate. The fast learning in the hippocampal links means that verb-structure associations will be quickly changed as new associations are stored, and this could explain the decay of the lexical boost (Hartsuiker et al. 2008). Hippocampal units are thought to be involved in encoding the task context (Komorowski et al. 2009) and this could help to explain variation in the magnitude of the lexical boost (Hartsuiker et al. 2008; reported same verb + structure priming of 73% in Exp. 1 and priming of 42% in Exp. 4). If the ability to maintain a constant task context across an experiment grows in development, then the links between task context and hippocampal associations could also help to explain the growth of the lexical boost in Rowland et al.'s (2011) article. Finally, if the meaning of 'brought up' is linked to the conjunctive codes, then the whole idiom will be needed to get the lexical boost. Meanwhile, the slow learning in the SRN can explain why different verb-particle combinations prime each other. This complementary systems account is speculative, but it does suggest some avenues that can be explored in future research: Which language representations are linked to the hippocampus? How does task context influence the lexical boost? What is the role of the medial temporal lobe in the boost?

Conclusion

Traditionally, language acquisition and language processing have been treated as distinct domains. Linguistic adaptation phenomena indicate that representations in the language processing system are being changed by some type of learning mechanism. The AGL-SRT

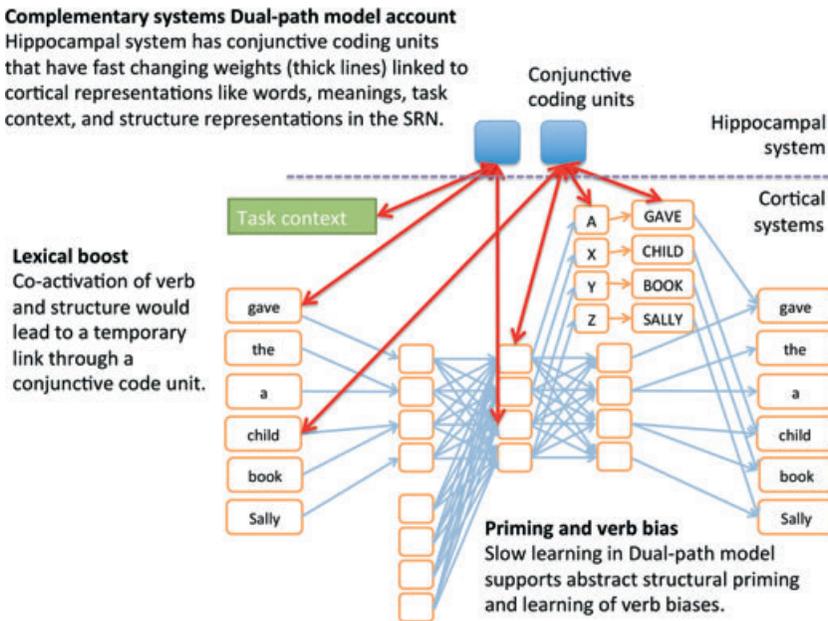


Fig 7. Complementary systems Dual-path model account of lexical boost and abstract priming.

literature suggests that the same mechanisms that govern the acquisition of new grammatical information could also be responsible for the tuning of existing representations. Computational models like the Dual-path model extended this idea to natural language by demonstrating that acquisition mechanisms like error-based learning can explain adult linguistic adaptation. These models also suggested that a single mechanism may not be able to account for the full complexity of adaptation phenomena, which pointed to the existence of multiple mechanisms. Likewise, AGL-SRT studies have found evidence for multiple mechanisms that differ in the explicitness of their representations, and we have reviewed findings from neuropsychological and developmental literatures which support a similar distinction in language. We suggest that further progress can be made by using the dissociation methods developed in the AGL-SRT literature to articulate the multifaceted nature of linguistic representations. In addition, models of language need to go beyond linguistic representations and integrate themselves with the multiple systems in the brain.

Short Biographies

Dr Franklin Chang is a lecturer at the University of Liverpool. He is interested in the relationship between language learning and sentence production. He has been a post-doctoral research fellow at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany and at the NTT Communication Sciences Laboratories near Kyoto, Japan. He was also a lecturer at Hanyang University in Seoul, South Korea. He received his BS in Math-Computer Science from Carnegie Mellon University and his PhD from the University of Illinois at Urbana-Champaign.

Marius Janciauskas is a new PhD student at the University of Liverpool and he is interested in structural priming in comprehension and production, as well as connectionist modeling of language processing.

Dr Hartmut Fitz is currently Research Staff at the Max Planck Institute for Psycholinguistics in Nijmegen, on a fellowship from the Netherlands Organization for Scientific Research. His research areas are language acquisition and sentence processing with a focus on SL and computational modeling. Previously, he taught at the University of Groningen. He holds an MA in Philosophy and Mathematics from Free University Berlin and a PhD in Cognitive Science from the University of Amsterdam.

Acknowledgement

Hartmut Fitz was funded by the Netherlands Organization for Scientific Research (NWO) under grant no. 275-89-008. Marius Janciauskas was funded by an ESRC Ph.D. studentship (grant number ES/J500094/1).

Note

* Correspondence address: Franklin Chang, School of Psychology, The University of Liverpool, Eleanor Rathbone Building, Bedford Street South, Liverpool L69 7ZA, UK. Email: franklin.chang@liverpool.ac.uk

Works Cited

- Abrahamse, Elger L., Luis Jiménez, Willem B. Verwey, and Benjamin A. Clegg. 2010. Representing serial action and perception. *Psychonomic Bulletin & Review* 17. 603–23.
- Anderson, John R., Daniel Bothell, Michael D. Byrne, Scott Douglass, Christian Lebiere, and Yulin Qin. 2004. An integrated theory of the mind. *Psychological Review* 111. 1036–60.

- Arnon, Inbal, and Neal Snider. 2010. More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language* 62. 67–82.
- Aslin, Richard N., Jenny R. Saffran, and Elissa L. Newport. 1998. Computation of conditional probability statistics by 8-month-old infants. *Psychological Science* 9. 321–4.
- Bannard, Colin, and Danielle Matthews. 2008. Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science* 19. 241–8.
- Bernolet, Sarah, and Robert J. Hartsuiker. 2010. Does verb bias modulate syntactic priming? *Cognition* 114. 455–61.
- Bock, Kathryn. 1986. Syntactic persistence in language production. *Cognitive Psychology* 18. 355–87.
- . 1989. Closed-class immanence in sentence production. *Cognition* 31. 163–86.
- , Gary S. Dell, Franklin Chang, and Kristine H. Onishi. 2007. Persistent structural priming from language comprehension to language production. *Cognition* 104. 437–58.
- , and Zenzi M. Griffin. 2000. The persistence of structural priming: transient activation or implicit learning? *Journal of Experimental Psychology: General* 129. 177–92.
- , and Helga Loebell. 1990. Framing sentences. *Cognition* 35. 1–39.
- Branigan, Holly P., Martin J. Pickering, and Alexandra A. Cleland. 1999. Syntactic priming in written production: Evidence for rapid decay. *Psychonomic Bulletin & Review* 6. 635–40.
- Buchner, Axel, Melanie C. Steffens, and Rainer Rothkegel. 1998. On the role of fragmentary knowledge in a sequence learning task. *The Quarterly Journal of Experimental Psychology Section A* 51. 251–81.
- Cave, Carolyn B. 1997. Very long-lasting priming in picture naming. *Psychological Science* 8. 322–5.
- Chambers, Kyle E., Kristine H. Onishi, and Cynthia Fisher. 2003. Infants learn phonotactic regularities from brief auditory experiences. *Cognition* 87. B69–77.
- , —, and —. 2010. A vowel is a vowel: Generalizing newly learned phonotactic constraints to new contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 36. 821–8.
- Chang, Franklin. 2002. Symbolically speaking: A connectionist model of sentence production. *Cognitive Science* 26. 609–51.
- . 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. 374–97.
- , Kathryn Bock, and Adele E. Goldberg. 2003. Can thematic roles leave traces of their places? *Cognition* 90. 29–49.
- , Gary S. Dell, and Kathryn Bock. 2006. Becoming syntactic. *Psychological Review* 113. 234–72.
- , Elena Lieven, and Michael Tomasello. 2008. Automatic evaluation of syntactic learners in typologically-different languages. *Cognitive Systems Research* 9. 198–213.
- Christiansen, Morten H., and Nick Chater. 1999. Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science* 23. 157–205.
- , M. Louise Kelly, Richard C. Shillcock, and Katie Greenfield. 2010. Impaired artificial grammar learning in agrammatism. *Cognition* 116. 382–93.
- Cleeremans, Axel, and Zoltán Dienes. 2008. Computational models of implicit learning. *The Cambridge handbook of computational modeling*, ed. by Ron Sun, 396–421. Cambridge, UK: Cambridge University Press.
- , and James L. McClelland. 1991. Learning the structure of event sequences. *Journal of Experimental Psychology: General* 120. 235–53.
- , David Servan-Schreiber, and James L. McClelland. 1989. Finite state automata and simple recurrent networks. *Neural Computation* 1. 372–81.
- Clegg, Benjamin A., Gregory J. DiGirolamo, and Steven W. Keele. 1998. Sequence learning. *Trends in Cognitive Sciences* 2. 275–81.
- Cleland, Alexandra A., and Martin J. Pickering. 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. 214–30.
- Cohen, Asher, Richard I. Ivry, and Steven W. Keele. 1990. Attention and structure in sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 16. 17–30.
- Conway, Christopher M., and Morten H. Christiansen. 2001. Sequential learning in non-human primates. *Trends in Cognitive Sciences* 5. 539–46.
- , and David B. Pisoni. 2008. Neurocognitive basis of implicit learning of sequential structure and its relation to language processing. *Annals of the New York Academy of Sciences* 1145. 113–31.
- Corkin, Suzanne. 2002. What's new with the amnesic patient H.M.? *Nature Reviews Neuroscience* 3. 153–60.
- Coyle, Jacqueline M., and Michael P. Kaschak. 2008. Patterns of experience with verbs affect long-term cumulative structural priming. *Psychonomic Bulletin & Review* 15. 967–70.
- Curran, Tim. 1997. Higher-order associative learning in amnesia: Evidence from the serial reaction time task. *Journal of Cognitive Neuroscience* 9. 522–33.
- , and Steven W. Keele. 1993. Attentional and nonattentional forms of sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19. 189–202.

- Cutting, J. Cooper, and Kathryn Bock. 1997. That's the way the cookie bounces: syntactic and semantic components of experimentally elicited idiom blends. *Memory & Cognition* 25. 57–71.
- Davidson, Doug J., and Peter Indefrey. 2011. Error-related activity and correlates of grammatical plasticity. *Frontiers in Psychology* 2. 1–16.
- Dell, Gary S., Kristopher D. Reed, David R. Adams, and Antje S. Meyer. 2000. Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26. 1355–67.
- Destrebecqz, Arnaud, and Axel Cleeremans. 2001. Can sequence learning be implicit? New evidence with the process dissociation procedure. *Psychonomic Bulletin & Review* 8. 343–50.
- Dienes, Zoltán, Gerry T. M. Altmann, Liam Kwan, and Alastair Goode. 1995. Unconscious knowledge of artificial grammars is applied strategically. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21. 1322–38.
- Dulany, Don E., Richard A. Carlson, and Gerald I. Dewey. 1984. A case of syntactical learning and judgment: How conscious and how abstract? *Journal of Experimental Psychology: General* 113. 541–55.
- Eichenbaum, Howard, and Neal J. Cohen. 2004. *From conditioning to conscious recollection: Memory systems of the brain*. New York, NY: Oxford University Press.
- Elman, Jeffrey L. 1990. Finding structure in time. *Cognitive Science* 14. 179–211.
- . 1993. Learning and development in neural networks: The importance of starting small. *Cognition* 48. 71–99.
- Ferreira, Victor S., Kathryn Bock, Michael P. Wilson, and Neal J. Cohen. 2008. Memory for syntax despite amnesia. *Psychological Science* 19. 940–6.
- Fitz, Hartmut. 2009. *Neural syntax*. Amsterdam: University of Amsterdam. Institute for Logic, Language, and Computation dissertation series.
- , and Franklin Chang. 2008. The role of the input in a connectionist account of the accessibility hierarchy in development. *Proceeding of the 32nd Annual Boston University Conference on Language Development*, ed. by Harvey Chan, Heather Jacob and Enkeleida Kapia, 120–31. Somerville, MA: Cascadia Press.
- Franco, Ana, Axel Cleeremans, and Arnaud Destrebecqz. 2011. Statistical learning of two artificial languages presented successively: how conscious? *Frontiers in Psychology* 2. 1–12.
- Frank, Michael C., Sharon Goldwater, Thomas L. Griffiths, and Joshua B. Tenenbaum. 2010. Modeling human performance in statistical word segmentation. *Cognition* 117. 107–25.
- French, Robert M., Caspar Addyman, and Denis Mareschal. 2011. TRACX: A recognition-based connectionist framework for sequence segmentation and chunk extraction. *Psychological Review* 118. 614–36.
- Freudenthal, Daniel, and Julian M. Pine. 2005. On the resolution of ambiguities in the extraction of syntactic categories through chunking. *Cognitive Systems Research* 6. 17–25.
- , —, Javier Aguado-Orea, and Fernand Gobet. 2007. Modeling the developmental patterning of finiteness marking in English, Dutch, German, and Spanish using MOSAIC. *Cognitive Science* 31. 311–41.
- Fu, Qiufang, Xiaolan Fu, and Zoltán Dienes. 2008. Implicit sequence learning and conscious awareness. *Consciousness and Cognition* 17. 185–202.
- Garnsey, Susan M., Neal J. Pearlmutter, Elizabeth Myers, and Melanie A. Lotocky. 1997. The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language* 37. 58–93.
- Geschwind, Norman. 1970. The organization of language and the brain. *Science* 170. 940–4.
- Goldrick, Matthew. 2004. Phonological features and phonotactic constraints in speech production. *Journal of Memory and Language* 51. 586–603.
- , and Meredith Larson. 2008. Phonotactic probability influences speech production. *Cognition* 107. 1155–64.
- Gómez, Rebecca L. 1997. Transfer and complexity in artificial grammar learning. *Cognitive Psychology* 33. 154–207.
- , and LouAnn Gerken. 1999. Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition* 70. 109–35.
- , and —. 2000. Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences* 4. 178–86.
- van Gompel, Roger P. G., and Martin J. Pickering. 2007. Syntactic parsing. *The Oxford handbook of psycholinguistics*, ed. by Gareth Gaskell, 289–307. Oxford, UK: Oxford University Press.
- Grodzinsky, Yosef, and Andrea Santi. 2008. The battle for Broca's region. *Trends in Cognitive Sciences* 12. 474–80.
- Gupta, Prahlad, and Neal J. Cohen. 2002. Theoretical and computational analysis of skill learning, repetition priming, and procedural memory. *Psychological Review* 109. 401–48.
- Hartsuiker, Robert J., Sarah Bernolet, Sofie Schoonbaert, Sara Speybroeck, and Dieter Vanderelst. 2008. Syntactic priming persists while the lexical boost decays: evidence from written and spoken dialogue. *Journal of Memory and Language* 58. 214–38.
- Hirst, William, Elizabeth A. Phelps, Marcia K. Johnson, and Bruce T. Volpe. 1988. Amnesia and second language learning. *Brain and Cognition* 8. 105–16.

- Howard, David, Lyndsey Nickels, Max Coltheart, and Jennifer Cole-Virtue. 2006. Cumulative semantic inhibition in picture naming: Experimental and computational studies. *Cognition* 100. 464–82.
- Hunt, Ruskin H., and Richard N. Aslin. 2010. Category induction via distributional analysis: Evidence from a serial reaction time task. *Journal of Memory and Language* 62. 98–112.
- Jacoby, Larry L. 1991. A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language* 30. 513–41.
- Jaeger, T. Florian, and Neal Snider. 2007. Implicit learning and syntactic persistence: Surprisal and cumulativity. *University of Rochester Working Papers in the Language Sciences* 3. 26–44.
- Jiang, Yuhong, and Marvin M. Chun. 2001. Selective attention modulates implicit learning. *The Quarterly Journal of Experimental Psychology Section A* 54. 1105–24.
- Jiménez, Luis. 2008. Taking patterns for chunks: Is there any evidence of chunk learning in continuous serial reaction-time tasks? *Psychological Research* 72. 387–96.
- , and Castor Méndez. 1999. Which attention is needed for implicit sequence learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition* 25. 236–59.
- , —, and Axel Cleeremans. 1996. Comparing direct and indirect measures of implicit learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22. 948–69.
- Juliano, Cornell, and Michael K. Tanenhaus. 1994. A constraint-based lexicalist account of the subject/object attachment preference. *Journal of Psycholinguistic Research* 23. 459–71.
- Kaschak, Michael P. 2007. Long-term structural priming affects subsequent patterns of language production. *Memory & Cognition* 35. 925–37.
- , and Arthur M. Glenberg. 2004. This construction needs learned. *Journal of Experimental Psychology: General* 133. 450–67.
- , Timothy J. Kutta, and Christopher Schatschneider. 2011. Long-term cumulative structural priming persists for (at least) one week. *Memory & Cognition* 39. 381–8.
- Kensinger, Elizabeth A., Michael T. Ullman, and Suzanne Corkin. 2001. Bilateral medial temporal lobe damage does not affect lexical or grammatical processing: Evidence from amnesic patient HM. *Hippocampus* 11. 347–60.
- Kidd, Evan. 2012. Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology* 48. 171–84.
- Knowlton, Barbara J., Seth J. Ramus, and Larry R. Squire. 1992. Intact artificial grammar learning in amnesia: Dissociation of classification learning and explicit memory for specific instances. *Psychological Science* 3. 172–9.
- , and Larry R. Squire. 1996. Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22. 169–81.
- Komorowski, Robert W., Joseph R. Manns, and Howard Eichenbaum. 2009. Robust conjunctive item–place coding by hippocampal neurons parallels learning what happens where. *The Journal of Neuroscience* 29. 9918–29.
- Konopka, Agnieszka E., and Kathryn Bock. 2009. Lexical or syntactic control of sentence formulation? Structural generalizations from idiom production. *Cognitive Psychology* 58. 68–101.
- Levelt, Willem J. M.. 1993. *Speaking: from intention to articulation*. Cambridge, MA: MIT Press.
- MacKay, Donald G., Deborah M. Burke, and Rachel Stewart. 1998a. HM's language production deficits: Implications for relations between memory, semantic binding, and the hippocampal system. *Journal of Memory and Language* 38. 28–69.
- , and Christopher B. Hadley. 2009. Supra-normal age-linked retrograde amnesia: Lessons from an older amnesic (H.M.). *Hippocampus* 19. 424–45.
- , and Lori E. James. 2002. Aging, retrograde amnesia, and the binding problem for phonology and orthography: A longitudinal study of “hippocampal amnesic” H.M. *Aging, Neuropsychology, and Cognition* 9. 298–333.
- , —, Christopher B. Hadley, and Kethera A. Fogler. 2011. Speech errors of amnesic H.M.: Unlike everyday slips-of-the-tongue. *Cortex* 47. 377–408.
- , —, Jennifer K. Taylor, and Diane E. Marian. 2007. Amnesic HM exhibits parallel deficits and sparing in language and memory: Systems versus binding theory accounts. *Language and Cognitive Processes* 22. 377–452.
- , Rachel Stewart, and Deborah M. Burke. 1998b. H.M. revisited: Relations between language comprehension, memory, and the hippocampal system. *Journal of Cognitive Neuroscience* 10. 377–94.
- Mazuka, Reiko. 1998. *The development of language processing strategies: A cross-linguistic study between Japanese and English*. Mahwah, NJ: Lawrence Erlbaum Associates.
- McClelland, James L., Bruce L. McNaughton, and Randall C. O'Reilly. 1995. Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review* 102. 419–57.
- McCloskey, Michael, and Neal J. Cohen. 1989. Catastrophic interference in connectionist networks: The sequential learning problem. *The psychology of learning and motivation: Advances in research and theory*, ed. by Brian H. Ross, 24:109–65. San Diego, CA: Elsevier.
- Meyer, Patric, Axel Mecklinger, Thomas Grunwald, Juergen Fell, Christian E. Elger, and Angela D. Friederici. 2005. Language processing within the human medial temporal lobe. *Hippocampus* 15. 451–9.

- Mintz, Toben H. 2002. Category induction from distributional cues in an artificial language. *Memory & Cognition* 30. 678–86.
- , Elissa L. Newport, and Thomas G. Bever. 2002. The distributional structure of grammatical categories in speech to young children. *Cognitive Science* 26. 393–424.
- Mirman, Daniel, Katharine Graf Estes, and James S. Magnuson. 2010. Computational modeling of statistical learning: effects of transitional probability versus frequency and links to word learning. *Infancy* 15. 471–86.
- Misyak, Jennifer B., Morten H. Christiansen, and J. Bruce Tomblin. 2010a. On-line individual differences in statistical learning predict language processing. *Frontiers in Language Sciences* 1. 1–9.
- , —, and —. 2010b. Sequential expectations: The role of prediction-based learning in language. *Topics in Cognitive Science* 2. 138–53.
- Mitchell, David B., and Alan S. Brown. 1988. Persistent repetition priming in picture naming and its dissociation from recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14. 213–22.
- Morgan, James L., and Elissa L. Newport. 1981. The role of constituent structure in the induction of an artificial language. *Journal of Verbal Learning and Verbal Behavior* 20. 67–85.
- Nissen, Mary J., and Peter Bullemer. 1987. Attentional requirements of learning: evidence from performance measures. *Cognitive Psychology* 19. 1–32.
- Novick, Jared M., John C. Trueswell, and Sharon L. Thompson-Schill. 2010. Broca's area and language processing: Evidence for the cognitive control connection. *Language and Linguistics Compass* 4. 906–24.
- Onishi, Kristine H., Kyle E. Chambers, and Cynthia Fisher. 2002. Learning phonotactic constraints from brief auditory experience. *Cognition* 83. B13–23.
- Oppenheim, Gary M., Gary S. Dell, and Myrna F. Schwartz. 2010. The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition* 114. 227–52.
- O'Reilly, Randall C., and Jerry W. Rudy. 2001. Conjunctive representations in learning and memory: Principles of cortical and hippocampal function. *Psychological Review* 108. 311–45.
- Pelucchi, Bruna, Jessica F. Hay, and Jenny R. Saffran. 2009. Statistical learning in a natural language by 8-month-old infants. *Child Development* 80. 674–85.
- Perruchet, Pierre, Emmanuel Bigand, and Fabienne Benoit-Gonin. 1997. The emergence of explicit knowledge during the early phase of learning in sequential reaction time tasks. *Psychological Research* 60. 4–13.
- , and Chantal Pacteau. 1990. Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General* 119. 264–75.
- , and Sébastien Pacton. 2006. Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences* 10. 233–8.
- Petersson, Karl-Magnus, Vasiliki Folia, and Peter Hagoort. 2012. What artificial grammar learning reveals about the neurobiology of syntax. *Brain and Language* 120. 83–95.
- Pickering, Martin J., and Holly P. Branigan. 1998. The representation of verbs: evidence from syntactic priming in language production. *Journal of Memory and Language* 39. 633–51.
- , and Victor S. Ferreira. 2008. Structural priming: a critical review. *Psychological Bulletin* 134. 427–59.
- Pinker, Steven. 1984. *Language learnability and language development*. Cambridge, MA: Harvard University Press.
- Pothos, Emmanuel M. 2007. Theories of artificial grammar learning. *Psychological Bulletin* 133. 227–44.
- Reber, Arthur S. 1967. Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior* 6. 855–63.
- Rodriguez, P. 2003. Comparing simple recurrent networks and n-grams in a large corpus. *Applied Intelligence* 19. 39–50.
- Rohde, Douglas L.T. 2002. A connectionist model of sentence comprehension and production. Technical Report. Pittsburgh, PA: Carnegie Mellon University.
- Rowland, Caroline, Franklin Chang, Ben Ambridge, Julian M. Pine, and Elena Lieven. 2011. The development of abstract syntax: Evidence from structural priming and the lexical boost. Paper presented at the Conference on Architectures and Mechanisms for Language Processing, Paris.
- Rowland, Lee A., and David R. Shanks. 2006. Sequence learning and selection difficulty. *Journal of Experimental Psychology: Human Perception and Performance* 32. 287–99.
- Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams. 1986. Learning representations by back-propagating errors. *Nature* 323. 533–6.
- Saffran, Jenny R. 2001. The use of predictive dependencies in language learning. *Journal of Memory and Language* 44. 493–515.
- , Richard N. Aslin, and Elissa L. Newport. 1996. Statistical learning by 8-month-old infants. *Science* 274. 1926–8.
- , Elissa L. Newport, Richard N. Aslin, Rachel A. Tunick, and Sandra Barrueco. 1997. Incidental language learning: listening (and learning) out of the corner of your ear. *Psychological Science* 8. 101–5.
- , and Diana P. Wilson. 2003. From syllables to syntax: Multilevel statistical learning by 12-month-old infants. *Infancy* 4. 273–84.

- Schacter, Daniel L. 1987. Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 13. 501–18.
- Seger, Carol Augart. 1994. Implicit learning. *Psychological Bulletin* 115. 163–96.
- Seidl, Amanda, Alex Cristià, Amelie Bernard, and Kristine H. Onishi. 2009. Allophonic and phonemic contrasts in infants' learning of sound patterns. *Language Learning and Development* 5. 191–202.
- Shanks, David R., and Theresa Johnstone. 1999. Evaluating the relationship between explicit and implicit knowledge in a sequential reaction time task. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 25. 1435–51.
- , Lee A. Rowland, and Mandeep S. Ranger. 2005. Attentional load and implicit sequence learning. *Psychological Research* 69. 369–82.
- Skotko, Brian G., Edna Andrews, and Gillian Einstein. 2005. Language and the medial temporal lobe: Evidence from HM's spontaneous discourse. *Journal of Memory and Language* 53. 397–415.
- Smith, Mark C., and Linda R. Wheeldon. 2001. Syntactic priming in spoken sentence production—an online study. *Cognition* 78. 123–64.
- Squire, Larry R. 2004. Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory* 82. 171–7.
- . 2009. The legacy of patient H.M. for neuroscience. *Neuron* 61. 6–9.
- Taylor, Conrad F., and George Houghton. 2005. Learning artificial phonotactic constraints: Time course, durability, and relationship to natural constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31. 1398–416.
- Thomas, Kathleen M., and Charles A. Nelson. 2001. Serial reaction time learning in preschool and school-age children. *Journal of Experimental Child Psychology* 79. 364–87.
- Thompson, Susan P., and Elissa L. Newport. 2007. Statistical learning of syntax: The role of transitional probability. *Language Learning and Development* 3. 1–42.
- Tomasello, Michael. 2003. *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Toro, Juan M., Scott Sennett, and Salvador Soto-Faraco. 2011. Generalizing linguistic structures under high attention demands. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37. 493–501.
- de Vries, Meinou H., Andre C. R. Barth, Sandra Maiworm, Stefan Knecht, Pienie Zwitterlood, and Agnes Flöel. 2010. Electrical stimulation of Broca's area enhances implicit learning of an artificial grammar. *Journal of Cognitive Neuroscience* 22. 2427–36.
- , Morten H. Christiansen, and Karl-Magnus Petersson. 2011. Learning recursion: Multiple nested and crossed dependencies. *Biolinguistics* 5. 010–35.
- Warker, Jill A., Gary S. Dell, Christine A. Whalen, and Samantha Gereg. 2008. Limits on learning phonotactic constraints from recent production experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 34. 1289–95.
- , ———, ———, and ———. 2009. Speech errors reflect the phonotactic constraints in recently spoken syllables, but not in recently heard syllables. *Cognition* 112. 81–96.
- Wells, J. B., M. H. Christiansen, D. S. Race, D. J. Acheson, and M. C. MacDonald. 2009. Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology* 58. 250–71.
- Wilkinson, Leonora, and David R. Shanks. 2004. Intentional control and implicit sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30. 354–69.
- Wilson, Michael P., and Susan M. Garnsey. 2009. Making simple sentences hard: Verb bias effects in simple direct object sentences. *Journal of Memory and Language* 60. 368–92.
- Wonnacott, Elizabeth, Elissa L. Newport, and Michael K. Tanenhaus. 2008. Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychology* 56. 165–209.